INTELIGÊNCIA ARTIFICIAL E DISCRIMINAÇÃO ALGORÍTMICA: MARCOS REGULATÓRIOS E PARÂMETROS ÉTICOS

Haide Maria Hupffer[1]
Gustavo da Silva Santanna[2]
DOI: 10.29327/5312711.1-3

Resumo: Decisões algorítmicas impulsionadas por técnicas de *Machine Learning* (ML) podem discriminar grupos legalmente protegidos, acirrar preconceitos, perpetuar desigualdades, violar os direitos humanos ou criar novas formas de injustiça. Ferramentas automatizadas usadas para seleção de candidatos à emprego, concessão de empréstimo, liberação de seguro, educação, contratação de plano de saúde e outras formas de decisões automatizadas habilitadas por inteligência Artificial têm se mostrado com vieses discriminatórios e preconceituosos, o que passa a exigir políticas mitigadoras e novas leis. A maioria dos algoritmos de Inteligência Artificial ainda operam como caixas pretas que fornecem pouca ou nenhuma explicação sobre as decisões tomadas. A proteção legal positiva na Constituição Federal de não discriminação é desafiada quando a Inteligência Artificial, e não os humanos, discrimina. O

1 Pós-Doutora e Doutora em Direito pela Unisinos. Pesquisadora no Programa de Pós-Graduação em Qualidade Ambiental e no Curso de Direito da Universidade Feevale. Líder do Grupo de Pesquisa CNPq/Feevale Direito e Desenvolvimento. Este capítulo é parte do resultado da pesquisa produzida no âmbito do seguinte projeto de investigação científica: "Inteligência Artificial e Sociedade de Algoritmos: regulação, riscos discriminatórios, governança e responsabilidades", financiado pela Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul – Fapergs – Processo número 61271.674.22274.26112021, Edital Fapergs 07/2021 – Programa Pesquisador Gaúcho – Pq. ORCID: https://orcid.org/0000-0002-4965-9258. Lattes: http://lattes.cnpq.br/4950629941533824 E-mail: haide@feevale.br.

2 Doutor e Mestre em Direito pela Unisinos. Professor da Universidade Federal do Rio Grande do Sul – UFRGS – e da Atitus Educação. Professor das Especializações em Direito do Estado da Universidade Federal do Rio Grande do Sul – UFRGS, em Direito Digital, Gestão da Inovação e Propriedade Intelectual e Direito Administrativo da Pontifícia Universidade Católica de Minas Gerais – PUC/MINAS. Procurador do Município de Alvorada/RS. Advogado. ORCID: https://orcid.org/0000-0003-1190-3355. Lattes: http://lattes.cnpq.br/3394849903197236. E-mail: gssantanna@hotmail.com.

artigo tem como objetivo principal identificar algumas externalidades negativas e os impactos que algoritmos programados como vieses discriminatórios e preconceituosos podem produzir na sociedade e aos direitos fundamentais, bem como busca-se analisar os principais movimentos regulatórios para lidar com os desafios éticos e legais que envolvem a questão da discriminação algorítmica, responsabilidade pela programação em todas as fases do ciclo de vida, transparência, segurança, explicabilidade, impactos nos direitos fundamentais e nos ecossistemas, reafirmando o direito à dignidade e à não discriminação e os valores da igualdade e da justiça. A pesquisa é exploratória e descritiva, com apoio no método dedutivo e na análise documental e revisão da literatura.

Palavras-chave: Discriminação Algorítmica. Vieses Discriminatórios. Explicabilidade.

1. Introdução

A Inteligência artificial (IA) está revolucionando todas as áreas do conhecimento com mudanças profundas na forma como o ser humano está vivendo e trabalhando. No entanto, como acontece com qualquer tecnologia, existem limitações e desafios a serem considerados para que opere com segurança, responsabilidade, respeito ao ser humano e que se utilize a tecnologia desenvolvida em decisões para o bem e não para o mal. Todas as áreas estão sendo profundamente impactadas e a cada nova inovação, velhas questões continuam sem respostas, dentre as quais citam-se: quem controla um sistema de IA se sua capacidade de aprendizagem leva a decisões que não fo-

ram as intenções projetadas na sua configuração inicial? Se a IA tomar uma decisão discriminatória, de quem é a responsabilidade pelas consequências legais, morais e sociais? Se for delegado às máquinas autonomia plena para tomada de decisão e livre arbítrio, quem garante que sistemas de IA ao adquirirem sua própria consciência não gerem seus próprios vieses?

Dotar a IA com capacidade de tomada de decisão autônoma e capacidade de aprendizado passa a exigir estruturas éticas, legais e de governança. Sistemas de IA precisam de muitos dados e parâmetros para serem treinados com precisão, contudo deve-se ter presente que os dados são alimentados e escolhidos por humano, e possíveis vieses discriminatórios podem ser transferidos gerando decisões racistas e sexistas.

A partir do exposto, o presente estudo objetiva identificar algumas externalidades negativas e os reflexos que algoritmos enviesados podem produzir na sociedade e aos direitos fundamentais, bem como examinar os principais movimentos regulatórios para lidar com as espinhosas questões éticas, sociais e políticas sobre os riscos da discriminação algorítmica para que a IA não perpetue o racismo estrutural enraizado nas decisões humanas.

A pesquisa é de natureza qualitativa e exploratória, adota o método dedutivo e utiliza como procedimento técnico a pesquisa bibliográfica e documental. A pesquisa documental é apoiada nas diretivas da União Europeia, Unesco, OCDE e de alguns países que avançaram na elaboração de diretivas no que concerne a responder aos desafios específicos colocados pela discriminação algorítmica, com aportes éticos e normativos que estabelecem salvaguardas mínimas para a proteção contra a discriminação algorítmica, respeito aos direitos humanos, justiça, transparência, segurança, prestação de contas e respeito aos valores democráticos em todo o ciclo de vida de sistemas de IA.

2. Inteligência Artificial: algumas aplicações

Com a finalidade de prolongar seu ciclo de vida, os algoritmos aprendem a interpretar interativamente as tarefas

a ele submetidas (classificação, regressão, árvore de decisão, modelos estatísticos, processos matemáticos, agrupamento, métodos bayesianos para inferência estatísticas e ANNs). Os algoritmos e os sistemas de IA se adaptam e utilizam a aprendizagem de máquina (*machine learning*) para alcançar objetivos, encontrar insights ocultos e padrões sem serem explicitamente programados para tal. Dito de outro modo, a máquina aprende (*machine learning)* com bases em dados de cálculos/respostas anteriores e melhoram com a experiência, podendo auxiliar na construção de decisões confiáveis e repetíveis. A *machine learning* é uma aplicação da IA que permite o aprendizado e o aperfeiçoamento de sistemas conduzidos por grandes volumes de dados, sendo já utilizados em diferentes áreas, como para auxiliar em prognósticos de doenças, na pontuação de crédito, na detenção de fraude, no reconhecimento facial e reconhecimento de voz, no processamento da linguagem natural (PNL) (Janiesch and Zschench, 2021).

Com base no problema apresentado e nos dados disponíveis, as etapas necessárias para treinar e implantar *machine learning* são diferentes, como observam Lloyd, Mohseni e Rebentrost (2013). Os autores apontam três tipos de *machine learning*: i] supervisionada (a máquina infere uma função a partir de um conjunto de dados que foram treinados para classificar e realizar previsões – o modelo vai se ajustando de forma interativa para avaliar diferentes características até alcançar o resultado desejado); ii] não supervisionada (quando não há supervisão do sistema, mas apenas os dados de entrada são fornecidos – o modelo de aprendizagem não supervisionado ajuda a encontrar tipos de padrões e estruturas ocultas em dados não rotulados, concentrando-se, em particular, no problema de aprendizagem de máquinas em larga escala – *big data* – em que tanto o treinamento como o número de recursos são grandes, o que possibilita uma assertiva maior na segmentação de clientes e uma comunicação mais efetiva com o público-alvo); iii] aprendizagem por reforço (aplicada com grande sucesso em ambientes fechados como em jogos e para sistemas multiagentes) (Lloyd, Mohseni and Rebentrost, 2013).

Startups estão explorando o uso de *machine learning* na indústria farmacêutica para projetar novos medicamentos e com promessas de serem muito mais seguros e eficazes. Os primeiros medicamentos para câncer projetados com o auxílio da Inteligência Artificial (IA) estão agora na fase de rigorosos testes clínicos com voluntários para observar se funcionam e se o tratamento é seguro para, em sequência, solicitarem aos órgãos reguladores a liberação para uso generalizado. Como exemplo, Heaven (2023) cita que desde o ano de 2021, a empresa Exscientia (Reino Unido) com apoio da Universidade de Medicina de Viena (Áustria) e outras empresas farmacêuticas estão testando uma nova tecnologia de *matchmaking*, que desenvolve medicamentos precisos para tratamento de câncer que é individualizado para cada paciente, levando em conta diferenças biológicas sutis entre as pessoas. Coletam amostras do tecido do paciente, dividem a amostra para incluir em uma amostra células normais e na outra amostra de células cancerígenas, dividindo-a novamente em mais cem partes e, em sequência, expõem cada uma das partes a vários coquetéis de drogas. Depois de expostos, utilizam a automação robótica e *machine learning* para observar o que acontece com a exposição de cada um dos coquetéis (Heaven, 2023).

Com a tecnologia de *matchmaking*, a IA possibilita realizar uma busca exaustiva pelo medicamento certo, sem precisar submeter o paciente ao tratamento convencional que exige vários meses de quimioterapia. Nos experimentos realizados, os pesquisadores perceberam que alguns medicamentos não mataram as células cancerígenas do paciente observado e em outros experimentos prejudicaram células saudáveis. Com o processo de *matchmaking*, o paciente recebeu o medicamento correto contra o seu câncer e dois anos após o tratamento o câncer desapareceu e estava em remissão completa. A indústria farmacêutica está apostando muito, pois além da IA propiciar descobertas de medicamentos de forma muito mais rápida, também pode se tornar mais barata e acessível a todos os pacientes. A aprendizagem de máquina reduz o trabalho meticuloso para o desenvolvimento de novos medicamentos em laboratórios (Heaven, 2023).

A técnica de aprendizado profundo (*deep learning*) possibilita a modelos computacionais compostos por várias camadas de processamento apreenderem com a descoberta de estruturas complexas de grandes conjuntos de dados usando o algoritmo de retropropagação. Este algoritmo é usado para "indicar como uma máquina deve alterar seus parâmetros internos para calcular a representação em cada camada a partir da representação na camada anterior". O *deep learning* propicia avanços no processamento de falas, imagens, imagens, vídeo e áudio, enquanto as redes recorrentes iluminaram os dados sequenciais, como texto e fala" (LeCun, Bengio and Hinton, 2015). A área da medicina utiliza abordagens baseadas em *deep learning*, como *deep neural networks,* que faz parte de uma família mais ampla de técnicas de aprendizado de máquina baseadas em redes neurais artificiais, alcançando resultados impressionantes no processamento de imagens de inúmeras doenças. As abordagens de *deep learning* são inspiradas na capacidade do cérebro humano de abstrair representações de alto nível com estímulos sensoriais de baixo nível (Hossain et al, 2023).

Os exemplos da medicina mostram que estão sendo realizados esforços substanciais para o enriquecimento de aplicações de IA com imagens médicas usando *deep learning* para diagnosticar erros em sistemas de diagnósticos de doenças, prever sintomas precoces de doenças, reconhecer e extrair padrões, melhorar a precisão de diagnósticos para uma variedade de doenças com utilização de algoritmos supervisionados ou não supervisionados. *Machine* e *deep learning* em estudos de dados médicos multidimensionais são utilizados para endossar o processo de decisão usando imagens médicas que apresentam, como toda nova tecnologia, vantagens e desvantagens (Latif et al, 2019).

A combinação entre IA, *machine learning, analytics* e *big data* está revolucionando todas as áreas pela possibilidade de auxiliar gestores e tomadores de decisão pela necessidade de mais previsibilidade sobre os resultados de cada ação tomada. A IA consegue cruzar dados, "identificar relações e gerar *insights* em um nível que seria praticamente impossível para

o cérebro humano". Setores tradicionais, como o agronegócio, estão usando a IA para detectar doenças e pragas, realizar uma estimativa de safra, identificar fenômenos que influenciam nas atividades de plantio, automatizar a contagem de frutas, máquinas agrícolas autônomas e inteligentes, previsão meteorológica personalizada para cada fazenda, distinguir entre frutos verdes de maduros e agricultura de precisão (MIT Technology Review, 2022).

Drones são desenvolvidos para serem usados massivamente em conflitos armados, com capacidade para encontrar seus próprios alvos e atacar infraestruturas sensíveis, o que os tornam apenas mais uma questão de programação de algoritmos, com consequências catastróficas se mal utilizados. A integração da IA generativa nas tecnologias de consumo podem ampliar os efeitos dos atuais sistemas de IA, tanto de forma positiva como negativa, podendo criar textos e imagens

A sensação do momento é o ChatGPT que foi lançado em novembro de 2022 pela OpenAI trouxe a Inteligência Artificial para o debate cotidiano. O que atrai os olhares é sua habilidade de resolver problemas complexos e com mais precisão do que os anteriores, atraindo em apenas um mês a marca de mais de 100 milhões de usuários. É um algoritmo que processa a linguagem humana e gera conteúdos utilizando modelos que foram treinados com GPTs (*Generative Pre-Trained Models*). Outra característica que explica o seu sucesso é o de ser uma interface conversacional mais acessível aos usuários leigos e a enorme base de dados de conhecimento que possui. Os algoritmos são treinados para poder aprender, usando o melhor dos conceitos de aprendizagem de máquina e redes neurais profundas. Por isso, em linguagem mais simples é conhecido como sendo uma mistura de um robô conversacional com algoritmos do tipo generativo aqueles que conseguem gerar conteúdo (Del Rey, 2023).

Neurotecnologias, por serem mais invasivas, geram preocupações éticas e de governança, pois podem adentrar na privacidade e na liberdade cognitiva. São inúmeros os avanços na área de neurotecnologia, como as que permitem o monitoramento da atividade cerebral para análises e pesquisas,

decodificares de humor, tecnologias capazes de ler a mente, bem como identificar culpados na área de segurança. Por isso, a necessidade de criar mecanismos, normativas e fronteiras que estabeleçam limites (MIT Technology Review, 2023a). Outra inovação, conhecida como interface subvocal, possibilita incorporar em óculos inteligentes com uso de IA uma interface para reconhecer comandos não vocalizados apenas com os movimentos dos lábios e da boca, que podem ser executados em conjunto com um celular. No futuro essa tecnologia poderá devolver a vós para os pacientes. Atualmente, ela já é usada para comunicações com outras pessoas em espaços onde falar poderia ser inconveniente ou inapropriado, a exemplo de bibliotecas ou locais barulhentos (Inovação Tecnológica, 2023).

O uso de algoritmos e da Inteligência Artificial para tomada de decisões está presente em uma ampla gama de setores, desde planejamento de tráfego, concessão de plano de saúde, diagnóstico de doenças, filtragem de *spam*, reconhecimento de fala e reconhecimento facial, assertividade para pagamento de empréstimos, se o candidato à vaga será um bom funcionário, gerenciamento de logística, previsão de alvos prováveis para intervenção policial, prevenir crimes ou resolver crimes passados fazendo previsões estatísticas (Borgesius, 2020). Contudo, a tomada de decisão por algoritmos pode colocar em risco os direitos humanos, como o direito à não discriminação, como pontua Borgesius (2020): "embora a tomada de decisão algorítmica possa parecer racional, neutra e imparcial, ela também pode levar à discriminação injusta e ilegal".

Uma decisão algorítmica é definida por Borgesius (2020) como o processo pelo qual o algoritmo produz uma saída (*output*). Dependendo da configuração do sistema de IA, o algoritmo pode decidir de forma totalmente automatizada em processos que antes eram de responsabilidade exclusiva de humanos, como também é utilizado para preparar decisões, ou seja, quando os humanos tomam decisões com base nos algoritmos. Em geral, é difícil de mensurar os limites entre a tomada de decisão humana e a automatizada. Os problemas de processo de tomada de decisão são múltiplos e complexos

e os riscos para indivíduos, grupos e sociedade são semelhantes para decisões totalmente ou parcialmente automatizadas (Council Of Europe, 2018).

Operada por meio de algoritmos, a IA assume uma importante função de realizar análises preditivas, baseadas na análise de dados de informações alimentadas por *big data*, mas muito se discute se esse processo pode interferir na autonomia e na privacidade, se será tendencioso e discriminatório, se poderá ser contaminado com erros e ser excessivamente invasivo. Uma preocupação geral é a falta de transparência desses processos o que passa a exigir medidas precaucionais para mitigar o risco (Zarsky, 2013, p. 1506).

3. As múltiplas faces da discriminação invisível impulsionada por algoritmos em sistemas de inteligência artificial

O uso de algoritmos de Inteligência Artificial, *Machine* e *deep learning* tornaram-se comuns no dia a dia das pessoas, além dos benefícios amplamente divulgados, podem gerar vieses contra minorias, mulheres e outras classes protegidas. Ao longo dos anos inúmeros países aprovaram leis para proteger os consumidores contra a discriminação na concessão de crédito, habitação e emprego, onde reguladores e agências têm a tarefa de fazer cumprir essas leis. Uma das questões mais prementes na adoção de algoritmos é a dificuldade de garantir justiça e transparência na sua utilização. Sistemas de recrutamento e seleção podem estar configurados para dar notas mais baixas para mulheres com pele mais escura, que se graduaram em determinada faculdade e que moram em determinado bairro ou que participaram de movimentos feministas; da mesma forma pela localização geográfica ou geografia da mídia social. Minorias relacionadas à origem racial ou imigrantes podem ser propensas a receberem menos ofertas de marketing relacionadas a moradias (Schmidt and Stephens, 2019, pp. 130-131).

Há uma crescente preocupação e debates sobre os riscos e benefícios que essas inovações tecnológicas represen-

tam, por serem considerados pouco transparentes e reprodutoras de desigualdades e racismo. Se mal programadas e usadas são capazes de perpetuar injustiças tendenciosas e resultar em decisões imprecisas, discriminatórias, bem como violar os direitos humanos fundamentais. Nos Estados Unidos, a Casa Branca, o *Federal Reserve Board*, o *Consumer Financial Protection Bureau* (CFPB), o *Government Accountability Office* e a *Federal Trade Commission* divulgam relatórios e frequentemente solicitam informações sobre pautas relacionadas à justiça algorítmica quanto ao uso da Inteligência Artificial, *Machine* e *deep learning* nos mercados de emprego, crédito e habitação, e no uso de dados de consumidores, funcionários e mercado (Schmidt and Stephens, 2019, pp. 130-131).

Quando o sistema é desenvolvido com viés discriminatório ou quando aprendeu com decisões humanas discriminatórias e tendenciosas, as decisões algorítmicas resultam em mais discriminação e em grave ofensa aos direitos humanos. Como exemplo, registra-se a utilização de decisões de algoritmos para policiamento com base em estatísticas criminais alimentadas com viés discriminatório. Se o sistema policial prestar mais atenção em um bairro onde se concentram muitos imigrantes, ele passará a registrar mais crimes daquele bairro do que em outro lugar. Esse registro não ocorre porque realmente há um volume maior criminalidade naquele local, mas pelo fato de deslocar mais policiais e, consequentemente, fazer o registro de todas as abordagens. O policiamento preditivo tem o potencial de causar um ciclo de feedback extremamente perigoso, pois os algoritmos podem representar decisões humanas discriminatórias, o que resulta em novas decisões preconceituosas que ferem os direitos humanos. (Borgesius, 2020).

Schmidt e Stephens (2019, p. 138) levantam o seguinte questionamento: o que significa um algoritmo ser justo? Os autores reportam que os algoritmos de *machine learning* treinados e os dados nos quais eles se baseiam para tomar decisões podem não ser isentos de viés, visto que são capazes de refletir os resultados de um processo de decisão tendencioso (humano ou não). Da mesma forma, um modelo bem intencionado

é capaz de criar um algoritmo que eterniza o viés existente achado nos dados utilizados para treinar o algoritmo. Igualmente, Schmidt e Stephens (2019, p. 138) chamam a atenção para o fato de que "algoritmos imprecisos ou mal construídos podem também captar correlações espúrias em dados que inadvertidamente levam a resultados relativamente piores para diferentes grupos demográficos". Por isso, o alerta dos autores sobre a importância de a sociedade reconhecer que é possível a ocorrência de vieses algorítmicos, que dados recém disponibilizados, mesmo que estejam associados à confiança, não estão livres de conterem vieses discriminatórios. Por essa razão, torna-se premente um grande esforço para desenvolver novos métodos e implementar abordagens que possam garantir que algoritmos utilizados em processos de tomada de decisão sejam tão justos quanto possível (Schmidt and Stephens, 2019, p. 144).

No âmbito da e-Administração, o uso da Inteligência Artificial pode incorporar as mazelas de uma Administração Pública Patrimonialista, ou seja, pessoalizada, e, consequentemente, discriminatória. Se sabe que muitas das dificuldades enfrentadas hoje em termos de gestão pública advêm das vicissitudes de um modelo de gestão que deveria já ter sido ultrapassado. Diante disso, não podem essas distorções ser incorporadas aos sistemas robotizados. Um exemplo desse problema pode ser extraído do COMPAS (*Correctional Offender Management Profiling for Alternative Sanctions*), um sistema de IA utilizado em estágios sequenciais da justiça criminal nos Estados Unidos em vários Estados (Wisconsin, Arizona, Delaware, Colorado, Louisiana, Kentucky, Washington, Virgínia), que inclui prejulgamento e correções comunitárias, liberdade condicional, avaliação de risco de reincidência de réus, auxiliando magistrados na dosimetria de penas e na concessão de liberdade condicional. O COMPAS considera uma série de variáveis sobre o réu (*input*) e é alimentado pelo próprio réu quando é efetuada uma prisão, com informações sobre histórico de violência, histórico criminal, abuso de substâncias, local de moradia, histórico escolar, profissão, ambiente social, lazer, se algum familiar já foi preso ou está preso, envolvimen-

to com pessoas que pertencem a organizações criminosos, problemas financeiros. Esses dados inseridos pelo réu são utilizados para contabilizar a porcentagem de reincidência. As respostas são inseridas no software COMPAS para gerar várias pontuações, inclusive análise preditiva de "Risco de reincidência" e "Risco de reincidência violenta" (Larson et al, 2016).

Larson et al (2016) analisaram mais de 10.000 réus criminais do Condado de Broward, Flórida, e compararam as taxas de reincidência previstas pela ferramenta COMPAS com as taxas reais de reincidência dos réus em um período de dois anos. Na análise realizada, os pesquisadores desvelaram que os réus negros, por este sistema, eram muito mais propensos do que os réus brancos a serem julgados incorretamente como estando em maior risco de reincidência, enquanto os réus brancos tinham um escore mais baixo de risco. Na análise realizada estava correta a previsão de reincidência do infrator em 61% das vezes, mas estava correta apenas em suas previsões de reincidência violenta 20%. Em relação a prever quem poderia reincidir em crime, o algoritmo previu "corretamente a reincidência para réus negros e brancos aproximadamente na mesma taxa (59% para réus brancos e 63% para réus negros), mas cometeu erros de maneiras muito diferentes. Ele classifica erroneamente os réus brancos e negros de maneira diferente quando examinados ao longo de um período de acompanhamento de dois anos". Nas análises os pesquisadores também descobriram que "réus negros que não reincidiram em um período de dois anos tinham quase duas vezes mais chances de serem erroneamente classificados como de maior risco em comparação com seus colegas brancos (45% contra 23%)". Restou claro aos pesquisadores que réus negros eram erroneamente classificados como possíveis reincidentes em crimes violentes (Larson et al, 2016). É um exemplo emblemático de um sistema de IA desenvolvido com algoritmos tendenciosos com pontuações de risco que eram racialmente injustas, o que gera discriminação algorítmica e reproduz preconceitos enraizados na sociedade.

Na área de recrutamento e seleção empresas estão preocupadas com possíveis viés inconsciente projetados na

construção de algoritmos em sistemas de IA com preconceitos que estão enraizados no cotidiano. A *startup* brasileira Jobecam fundada em 2019 está desenvolvendo uma plataforma que possibilita que o candidato ao emprego e o recrutador se concentrem apenas nas habilidades e competências exigidas para a vaga. À vista disso, a *startup* alterou sua plataforma "que permite filtrar currículos e realizar entrevistas, pré-gravadas ou ao vivo, de forma totalmente anônima". O objetivo não é esconder o candidato ao emprego, mas sim dar ênfase às reais habilidades e competências do candidato. Outra forma utilizada para minimizar a discriminação é não inserir o ano da formação universitária ou de curso profissionalizante para "evitar os cálculos e driblar o etarismo, disponibiliza apenas o período de cada um". Outra opção é possibilitar para a contratante ocultar os nomes da universidade na qual o candidato se graduou e o nome das empresas em que já trabalhou, para minimizar discriminação (Cardial, 2023).

Críticas de que os algoritmos não são neutros e que podem representar novas formas de preconceitos ou exacerbar as já existentes dependerá em grande medida das escolhas que são feitas na concepção de sistemas de IA. Muitas das escolhas são moldadas pelo ambiente ético e legal em que a empresa opera. Contudo, não é tão simples capturar os dados ameaçados por algoritmos tendenciosos que violam a legislação (Kim, 2017, p. 865). Empregadores tendenciosos escondem sua intenção discriminatória em algoritmos, podendo até intencionalmente desejarem que determinado resultado discriminatório ocorra e, assim, usam modelos que mascaram decisões discriminatórias por trás da fachada neutra de análise de dados. No geral, a vítima tem dificuldade de provar ter sido discriminada. Por isso, é preciso ficar atento em relação as escolhas na codificação da informação pelos programados de sistemas de IA, "erros nos dados, confiança em amostras não representativas, ou a seleção de variáveis para exclusão ou inclusão pode produzir um modelo impreciso de maneira sistemática", para evitar que esses erros sistemáticos reduzam ainda mais as oportunidades de grupos desfavorecidos (Kim, 2017, p. 884).

Uma outra forma de gerar discriminação algorítmica é quando um erro no registro de um indivíduo (redes sociais, dados coletados de sites públicos) pode sugerir que ele seja inadimplente em um empréstimo ou que possa ter um antecedente criminal, quando na verdade essa informação que está na rede não é verdadeira. "Informações imprecisas não aumentam inerentemente preocupações de igualdade, pois os erros podem ser distribuídos aleatoriamente, infectando os registros de membros de grupos privilegiados, bem como de grupos protegidos" (Kim, 2017, pp. 885-886).

Se o algoritmo fizer uma previsão e tomar uma decisão com base em erros de dados da pessoa, pode privar injustamente o indivíduo de oportunidades de emprego, acesso à crédito, acesso a plano de saúde e seguro, por exemplo.

Essa "discriminação invisível" ou "discriminação por associação" pode pôr em xeque a ideia de os algoritmos serem possuidores de "neutralidade" (Navarro, 2017, pp. 48-51). Isso porque as condutas estabelecidas pelos robôs que utilizam inteligência artificial o são com base em padrões de comportamentos gerados pelos próprios seres humanos e que a podem levar à discriminação com base na raça, etnia, idade, localização geográfica, renda, etc. O Estado-Administração, a contrário senso, de base desses dados, deve, sim, direcionar suas políticas públicas a fim de otimizar seus serviços e custos, de maneira a transformar a "discriminação" em "igualdade social" (Navarro, 2017, pp. 48-51).

Na Alemanha, essa aceitação, que envolve diretamente elementos da personalidade, denomina-se: direito à "autodeterminação informativa" ou "autopresentação" que consiste no direito de o próprio indivíduo decidir acerca da divulgação e utilização de seus dados pessoais, possibilitando, inclusive, que o indivíduo "se insurja contra as falsas, não autorizadas, degradantes ou deturpadas representações de sua pessoa, bem como o protege das observações secretas e indesejadas de sua personalidade" (Menke, 2015, p. 210). Diferencia-se da autodeterminação que trata do direito de o indivíduo determinar/definir sua identidade, desde a origem biológica até a orientação sexual. Também se distingue da "autopreservação"

que consiste na garantia de o indivíduo recolher-se para si, de ficar só, abrangendo o sigilo a diários pessoais, boletins médicos e materiais biológicos (Menke, 2015, p. 210). Essa preocupação se deve, principalmente, em razão da utilização que pode ser dada aos dados dos indivíduos, uma vez que, estes podem ser manipulados por instituições (públicas ou privadas) sem que o indivíduo saiba disso (Menke, 2015, pp. 210-2011). É exatamente no combate à manipulação dos dados e à discriminação que pode ser gerada em razão das informações, que o respeito à privacidade se impõe.

Stefano Rodotà (2008, p. 35), inclusive, chega a sugerir que, se fosse possível redefinir uma classificação acerca da privacidade, o grau máximo de opacidade seria dado àquelas que pudessem gerar práticas discriminatórias, sendo de grau "máximo de transparência aquelas que, referindo-se à esfera econômica dos sujeitos, concorrem para embasar decisões de relevância coletiva." Logo, a privacidade não possui caráter absoluto (Batalla, 2010, p. 767) e sua limitação deve estar condicionada a determinadas circunstâncias, como a exigência legal de seu abrandamento, para evitar eventuais "intromissões ilegítimas" (Limberger, 2007, pp. 127-130). Da mesma maneira os dados coletados deveriam ser mínimos, "*no sólo para preservar el derecho a la protección de datos, sino también para darles mayor efectividad*" (Batalla, 2010, p. 735).

Algoritmos são considerados os elementos-chave que sustentam serviços e infraestruturas cruciais na sociedade da informação. Ao serem projetados para proporcionar acesso amplo e igualitário à sociedade, cada vez mais a humanidade depende de algoritmos para tomar decisões importantes. Contudo, algoritmos não são eticamente neutros e podem acirrar ou perpetuar diferentes riscos de discriminação, como já registrado, e ameaçar direitos de grupos legalmente protegidos. São questões complexas que precisam ser enfrentadas pelo Direito e que exigem um grande esforço para desbloquear a "caixa preta" algorítmica para entender como esses sistemas tomam decisões (Xenidis and Senden, 2020, pp. 1-2). Programadores que desenvolvem sistemas podem também não saber como ele se comportará quando utilizado na prática e

quando for alimentado com determinados dados. De igual forma, segredos de Estado ou segredos comerciais podem dificultar a obtenção de informações sobre sistemas algorítmicos (Borgesius, 2020).

O foco principal dos desenvolvedores é garantir que os algoritmos executem as tarefas para os quais foram projetados. Compreender o tipo de "pensamento que orienta os desenvolvedores é essencial para entender o surgimento de vieses em algoritmos e na tomada de decisões algorítmicas", como pontuam Tsamados et al (2022). Os autores observam que o pensamento que predomina no campo do desenvolvimento de algoritmos é o "formalismo algorítmico", ou seja, os desenvolvedores aderem as regras e formas prescritas e se esforçam para serem neutros, mas esquecem que essa abordagem ignora a complexidade da sociedade e do mundo real, o que amplia o risco de consolidar as condições sociais existentes, como desigualdades estruturais que prejudicam etnias. Vieses algorítmicos podem ocorrer em qualquer estágio do processo, ou seja, desde o design, desenvolvimento, implantação e tomada de decisão. Tsamados et al (2022) alertam que podem ocorrer ciclos viciosos quando algoritmos realizam avaliações equivocadas de um determinado grupo no com base na mera correlação. Resultados injustos e com vieses ampliam a discriminação.

4. O desafio de regulamentar a Inteligência Artificial para garantir a promoção da igualdade e a não discriminação

Há consenso sobre a necessidade de avançar em busca da justiça algorítmica, em especial, para mitigar os riscos de discriminação direta (tratamento desigual) e indireta (impacto desigual) frente as decisões algorítmicas. Contudo, não há consenso sobre como definir, medir e estabelecer padrões de justiça algorítmica. Uma forma de mitigar os riscos é investir em pesquisas para esclarecer a natureza dos riscos éticos com orientação sólida e transparente para a governança do design dos algoritmos e do uso da tecnologia. (Tsamados et al, 2022).

Não é a tecnologia específica (computação em nuvem, celulares, computadores, plataformas online e assim por diante) que causa problemas éticos e discriminatórios, mas sim o que o hardware faz com software e com os dados em todo o ciclo de vida, desde o desenvolvimento de algoritmos, coleta de dados, compartilhamento, armazenamento, análise e uso dos dados. Logo, a macroética está se voltando para as diferentes dimensões morais da utilização de cada vez mais dados (pessoais sensíveis), questões de privacidade, anonimato, transparência, confiança, responsabilidade e com a crescente dependência de algoritmos para analisar os dados, a fim de moldar escolhas e tomar decisões. Outra questão presente nas discussões é a redução gradual do envolvimento humano nas decisões de algoritmos ou mesmo na supervisão de muitos processos automatizados, o que passa a exigir que sejam colocadas questões prementes de equidade, responsabilidade, respeito pelos direitos humanos e não discriminação (Floridi and Taddeo, 2016).

Floridi e Taddeo (2016) pontuam que os desafios éticos colocados pela ciência dos dados estão delineados em três eixos de pesquisa: a ética dos dados, a ética dos algoritmos e a ética das práticas. Para os autores, a ética dos dados diz respeito a coleta e análise de grandes conjuntos de dados (desde o uso de *big data* até pesquisas biomédicas, das áreas de ciências sociais, publicidade, filantropia de dados, criação de perfis), a possível reidentificação de indivíduos com a utilização de tecnologias de mineração de dados e com a reutilização de grandes conjuntos de dados, o que pode resultar em riscos para a chamada "privacidade de grupo" e sérios problemas éticos de discriminação de grupos (por idade, sexismo, preconceito, etnia) até violência contra grupos. No que concerne a ética dos algoritmos, especialmente no caso de ML (*machine learning*), os desafios éticos incluem "responsabilidade moral, responsabilidade do designer e dos cientistas de dados em relação as consequências imprevistas e indesejadas (por exemplo, discriminação ou promoção de conteúdo antissocial), bem como oportunidades perdidas". Por sua vez, a ética das práticas está centrada no consentimento do titular de dados, na privacidade

do usuário e no uso secundário dos dados. Nessa linha, a ética das práticas preocupa-se com a "responsabilidade de pessoas e organizações encarregadas de processos, estratégias e políticas de dados, incluindo cientistas de dados, com o objetivo de definir uma estrutura ética para moldar códigos profissionais sobre inovação, desenvolvimento e uso responsáveis". As práticas éticas devem garantir que os algoritmos e sistemas de IA possam promover "tanto o progresso da ciência de dados quanto a proteção dos direitos de indivíduos e grupos" (Floridi and Taddeo, 2016).

Um dos principais desafios na atualidade apontados por Morley et al (2020) relaciona-se ao que é nominado vantagem dupla de *machine learning* éticas, ou seja, como avançar para que as oportunidades sejam capitalizadas se ainda não estão claros os danos que alimentam todo o ciclo de vida dos algoritmos. Como construções sociotécnicas poderosas, os algoritmos de *machine learning* levantam inúmeras preocupações em relação aos desenvolvedores e organizações que os projetam quanto sobre seu código que pode ser a maior promessa como a maior ameaça à humanidade. Os autores argumentam que sem transparência e pela complexidade e dificuldade de interpretar os códigos será difícil controlar, monitorar e corrigir sistemas algorítmicos. Contudo, há um apelo coletivo de cientistas sociais, filósofos, eticistas, formuladores de políticas, tecnólogos e sociedade civil para que sejam desenvolvidos mecanismos de governança apropriados para a sociedade. Dito de outro modo, além de capitalizar as oportunidades é necessário ter certeza de que os valores éticos, princípios e os direitos humanos estão sendo respeitados e que a tomada de decisão seja justas e eticamente responsável (Morley et al, 2020).

A busca pela justiça algorítmica, responsabilidade e transparência, são temáticas cada vez mais presentes nas discussões sobre as implicações éticas dos algoritmos. A cada dia surgem novos problemas éticos e formas de abordá-los. Falar sobre viés algorítmico tornou-se um tema central para lidar com essa nova realidade digital e "suas consequências para o Estado Democrático de Direito, à ordem jurídica estabelecida e

aos princípios gerais do direito". A União Europeia avançou no que concerne a responder aos desafios específicos colocados pela discriminação algorítmica, com aportes normativos que estabelecem salvaguardas mínimas para a proteção contra a discriminação nos Estados-Membros. O direito de não ser discriminado é um princípio geral e um direito fundamental no direito da União Europeia (Xenidis and Senden, 2020, pp. 1-2).

O art. 22 do *General Data Protection Regulation* (GPDR) proíbe certas decisões automatizas incluindo a definição de perfis ao dispor que 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". Contudo, a proibição não se aplica para os seguintes casos: i] quando a decisão "for necessária para a celebração ou a execução de um contrato entre o titular dos dados e um responsável pelo tratamento"; ii] quando o titular dos dados deu consentimento à decisão automatizada; iii] quando for "autorizado por lei da União Europeia 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" (União Europeia, 2016).

Em situações em que o controlador puder confiar no consentimento informado do titular ou na exceção contratual uma regra diferente é acionada (Borgesius, 2020). Essa regra está prevista também no art. 22 da GPDR que deixa claro que o responsável pelo tratamento de dados tem o dever de aplicar "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" (União Europeia, 2016).

Como exemplo, Borgesius (2020) cita que um banco pode possibilitar que uma decisão automatizada seja reconsiderada ao negar automaticamente um empréstimo por meio de seu site. Nessa situação, o cliente pode ligar para o banco

solicitando que um funcionário reconsidere a decisão. Em relação a transparência para decisões automatizadas, a GDPR nos artigos 13, inciso 2 (f) e 14, inciso 2 (f) dispões que o responsável pelo tratamento deve informar ao titular dos dados sobre a "existência de tomadas de decisão automatizadas, incluindo a definição de perfis" e neste caso o responsável deve dar informações claras, compreensíveis e "significativas sobre a lógica envolvida, bem como o significado e as consequências previstas de tal processamento para o titular dos dados" (Borgesius, 2020).

Diretivas éticas e documentos foram desenvolvidos por Estados, empresas, grupos de pesquisadores, universidades e organizações nos últimos anos com recomendações para adoção de princípios éticos para a IA. Registra-se que o primeiro padrão intergovernamental sobre IA foi publicado pela *Organisation for Economic Cooperation and Development* (OECD), que anunciou no dia 22 de maio de 2019 que seus trinta e seis países membros, juntamente com outros seis países parceiros (Argentina, Brasil, Colômbia, Costa Rica, Peru e Romênia), concordaram em estabelecer padrões éticos para a IA. O documento recomendatório contempla diretrizes e princípios da OECD (2019) que visam promover a inovação para que a IA seja confiável, assim como garantir que os algoritmos sejam projetados para serem justos, seguros, robustos e que respeitem os direitos humanos e os valores democráticos em todas as etapas do ciclo de vida. A recomendação se concentra em torno de cinco princípios: i] Crescimento inclusivo, desenvolvimento sustentável e bem-estar; ii] valores centrados no ser humano e na justiça; iii] transparência e explicabilidade; iv] robustez e segurança; v] prestação de contas (OECD, 2019).

Também no ano de 2019, Cingapura lançou a primeira edição do "*Model AI Governance Framework (Model Framework)*" que fornece orientação detalhada para empresas do setor privado com abordagens éticas e de governança importantes para implantar soluções de IA e, assim, promover a compreensão e confiança do público. Como princípios orientadores também elegem que as decisões tomadas pela IA devem ser explicáveis, transparentes, justas e centradas no ser humano. O fra-

mework desenvolvido está alicerçado em cinco práticas que as empresas devem observar: i] criar estruturas de governança interna com funções e responsabilidades claras para monitorar e gerenciar riscos e treinar as equipes; ii] determinar o nível de envolvimento humano na tomada de decisão com grau apropriado para minimizar o risco de danos aos indivíduos; iii] atentar para o gerenciamento de operações para que sejam minimizados vieses nos dados e nos algoritmos, para atender esse requisito é importante medidas de redução de risco como a explicabilidade, robustez e possibilitar ajustes regulares; iv] interação e comunicação com uma rede ampla de partes interessadas, ou seja, as políticas de IA devem ser divulgadas para os usuários em linguagem compreensível, como também deve ser permitido que os usuários forneçam feedback; v] robustez e reprodutibilidade. Registra-se que no ano de 2018 foi criado o Conselho Consultivo sobre o Uso Ético de IA e Dados, cujos membros são nomeados pelo Ministro das Comunicações e da Informação. Em 4 de dezembro de 2020, Cingapura lançou o primeiro guia para auxiliar "organizações e funcionários a entenderem como as funções de trabalho existentes podem ser redesenhadas para aproveitar o potencial da IA, de modo que o valor de seu trabalho seja aumentado" e os riscos minimizados (Singapore, 2022).

Outro documento importante foi publicado pela União Europeia no ano de 2020 com o título *The ethics of artificial intelligence: Issues and initiatives*. É um extenso documento que sistematiza os principais dilemas éticos e as questões morais associadas à implantação da IA, levantando questões sobre robôs assumindo papéis de humanos nos relacionamentos, manipulação intencional do mercado, problemas legais relacionadas às decisões tomadas por IAs, responsabilidade por incitar atividades criminosas, assédio, tortura, crimes sexuais, fraude, responsabilidade por acidentes com veículos autônomos, direitos humanos e bem-estar, danos emocionais, risco existencial, dano ambiental e sustentabilidade, dano social e justiça social, dentre outros temas. O documento também apresenta diversas iniciativas nacionais e internacionais sobre ética para a IA nos seguintes temas: direitos

humanos, bem-estar das pessoas, impactos da IA na experiência emocional e na manipulação emocional e intencional dos usuários, cultural, perpetuação de injustiças e possibilidade de ampliar a discriminação contra grupos específicos da sociedade, recomendações de estruturas de governança, necessidade de prestação de contas, ser auditável, sobre o significado de dar autonomia de decisão aos algoritmos, padrões e órgãos reguladores para supervisionar o uso da IA, iniciativas que tratam de como ampliar a inclusão social e a diversidade, como minimizar as diferenças entre países ricos e países pobres, preocupações com danos ambientais, necessidade de maior engajamento público e educação no que diz respeito aos danos da IA, sistemas de armas autônomas, riscos associados à utilização da IA na saúde e na educação, dentre outras iniciativas que mostram a necessidade de regular a IA (EU, 2020).

Um problema delicado em relação a discriminação algorítmica é a questão da responsabilidade. Xenidis e Senden (2020, p. 26-27) questionam: Quem deveria ser responsabilizado: o designer e o programador de algoritmos discriminatórios ou o provedor e empresa que utilizam os algoritmos projetados? As vítimas de discriminação devem recorrer a qual tribunal? Os tribunais e os advogados estão preparados para tratar reivindicações individuais e coletivas de denúncias? Para os autores, a discriminação algorítmica é um tema que ainda não é do conhecimento da maioria dos cidadãos, advogados, judiciário e dos legisladores. Para além da regulação devem ser realizadas campanhas e divulgação de pesquisas que ajudariam a expor as diferentes manifestações de discriminação algorítmica. Exemplos de debates públicos sobre o tema é o da European Union Agency for Fundamental Rights (FRA) que desde 2018 vem estudando o impacto da IA e dos algoritmos nos direitos fundamentais com lançamento do documento *#BigData: Discrimination in data-supported decision-making* e o da Presidência finlandesa do Conselho da Europa que organizou no ano de 2019 uma conferência sobre o tema o que mostra uma crescente consciência política. (Xenidis and Senden, 2020, pp. 26-27).

O documento *#BigData: Discrimination in data-supported decision-making* publicado em 2018 pela FRA, foca a discussão especificamente nos direitos humanos e como a discriminação algorítmica afetou direitos fundamentais, buscando assim contribuir para a compreensão dos desafios encontrados neste campo. No final do documento listam exemplos de como é possível avançar em direção a salvaguardar os direitos fundamentais no desenvolvimento e uso de algoritmos, como: i] ser o mais transparente possível e possibilitar abrir os sistema para ver como os algoritmos foram construídos; ii] avaliar periodicamente o impactos nos direitos fundamentais para identificar potenciais preconceitos e abusos na aplicação e nas decisões de algoritmos; iii] verificar a qualidade dos dados coletados e usados para construir algoritmos; iv] certificar-se de que o algoritmo construído pode ser explicado, ou seja, conhecer a lógica por trás dos cálculos que alimentam a tomada de decisão, principalmente, conhecer os dados que foram usados para criar o algoritmo (EU, 2018).

A Conferência Geral da Organização das Nações Unidas para a Educação, Ciência e Cultura (Unesco), reunida em Paris, aprovou no dia 21 de novembro de 2021 a Recomendação sobre a Ética da Inteligência Artificial para orientar as "sociedades para que lidem de forma responsável com os impactos conhecidos e desconhecidos das tecnologias de IA sobre seres humanos, sociedades, meio ambiente e ecossistemas, oferecendo-lhes uma base para aceitar ou rejeitar essas tecnologias". A preocupação ética, a dignidade humana, o bem estar da humanidade, a promoção da responsabilidade empresarial e a prevenção de danos são temas centrais que pautam a Recomendação. É dirigida aos Estados-membros, grupos, instituições públicas, empresas, indivíduos, responsáveis pela formulação de legislação, para garantir a incorporação da ética em todas as etapas o ciclo de vida de sistemas de IA. (Unesco, 2022, pp. 10-14).

A Recomendação sobre a Ética da Inteligência Artificial da Unesco (2022, p. 18) está fundamentada em um conjunto de valores e princípios que objetivam motivar a formulação de normas jurídicas e políticas. O conjunto de valores estão as-

sentados em torno de quatro eixos: i] respeito, proteção e promoção dos direitos humanos, das liberdades fundamentais e da dignidade humana; ii] prosperidade ambiental e ecossistêmica; iii] garantir diversidade e inclusão; iv] viver em sociedades pacíficas, justas e interconectadas. Para que os valores sejam concretizados com mais facilidade em declarações, normativas e ações políticas, a Recomendação é fundamentada nos seguintes princípios: i] proporcionalidade e não causar dano; ii] segurança e proteção; iii] justiça e não discriminação; iv] sustentabilidade; v] direito à privacidade e proteção de dados; vi] supervisão humana e determinação; vii] transparência e explicabilidade; viii] responsabilidade e prestação de contas; ix] conscientização e alfabetização; x] governança e colaboração adaptáveis e com múltiplas partes interessadas (Unesco, 2022, p. 18-23).

A União Europeia está realizando um esforço regulatório para instituir um marco legal para a IA e que poderá servir de modelo para outros países. Em 19 de fevereiro de 2020 publicou o *Livro Branco* sobre IA que traz a abordagem europeia voltada para promover a IA com excelência e abordar os riscos associados a utilização desta tecnologia. Em abril de 2021 foi formalizada uma proposta de Lei para a IA pelo Parlamento Europeu e do Conselho alinhando-a com os valores e direitos fundamentais da UE para o desenvolvimento de um ecossistema de confiança, centrada no risco, e com imposição de obrigações para toda a cadeia de valor da IA. Realçam no documento que a UE deve estar na vanguarda mundial no que concerne a desenvolver uma IA que seja segura, ética e de confiança, a serviço das pessoas, com o objetivo de aumentar o bem-estar dos seres humanos e que possa garantir segurança jurídica para facilitar investimentos e a inovação (European Commission, 2021).

A proposta está sendo examinada pelos legisladores europeus com previsão de aprovação para o ano de 2023. Considerada uma normativa extremamente ambiciosa, segue uma abordagem baseada no risco com sanções caso sejam infringidas as regras: i] risco inaceitável; ii] risco elevado; iii] risco baixo ou mínimo. Riscos inaceitáveis, como sistemas de IA

concebidos para distorcer o comportamento humanos e que podem provocar danos físicos ou psicológicos devem ser proibidos. Da mesma forma são considerados de risco inaceitável e, portanto, proibidos sistemas de IA que podem criar resultados discriminatórios e violar o direito à dignidade e à não discriminação e os valores da igualdade e da justiça. Também devem ser proibidos sistemas de identificação biométrica à distância "em tempo real" e quando estritamente necessários estarão sujeitos a autorização expressa "de uma autoridade judiciária ou de uma autoridade administrativa independente de um Estado-Membro". Em relação ao risco elevado, sistemas de IA "só podem ser colocados no mercado da União ou colocados em serviço se cumprirem determinados requisitos obrigatórios" que estão enumerados no documento. Para cada grau de risco estão previstas regras ou a possibilidade de criação de normativas que deverão ser executadas, acompanhadas e fiscalizadas por um sistema de governança do Conselho Europeu de Inteligência Artificial e cada "Estado-Membro deverá designar, ou criar, uma autoridade notificadora responsável por estabelecer e executar os procedimentos necessários para a avaliação, a designação e a notificação de organismos de avaliação e fiscalização da conformidade" (European Commission, 2021).

Também há previsão para restringir o uso de reconhecimento facial em locais públicos, tanto por policiais quanto por empresas privadas, para evitar a vigilância em massa. Assim, a proposta objetiva estabelecer regras sobre a confiabilidade, transparência, *accountability*, controles rigorosos. São inúmeros pontos que estão em discussão, como obrigar fornecer a auditores externos ou reguladores o acesso ao código-fonte e algoritmos para fazer cumprir a lei, quais tipos de IA são classificados como de "alto risco", preocupação se uma legislação mais restritiva pode desacelerar a inovação e se a forma como está redigido o Projeto de Lei será suficiente para proteger as pessoas de danos graves (MIT Technology Review, 2022).

No ano de 2023 foi publicada a "Declaração Europeia sobre os direitos e princípios digitais para a década digital" em que reafirma que a visão da UE se centra nas pessoas, respei-

to pelos direitos fundamentais, o Estado de Direito e a democracia, a acessibilidade, a inclusão, a igualdade, a resiliência, a segurança, melhoria da qualidade de vida e a disponibilidade de serviços. A Declaração está dividida em 12 capítulos, assim sistematizados: i] Dar prioridade às pessoas no processo de transformação digital; ii] Solidariedade e inclusão; iii] Conectividade; iv] Educação, formação e competências digitais; v] Condições de trabalho justas e equitativas; vi] Serviços públicos digitais em linha; vii] Liberdade de escolha para interações com algoritmos e sistemas de inteligência artificial e um ambiente justo; viii] Participação no espaço público digital; ix] ambiente digital protegido e seguro; x] Privacidade e controlo individual dos dados; xi] Proteção e capacitação das crianças e dos jovens no ambiente digital; xii] sustentabilidade. A Declaração objetiva "servir de referência para atividades no contexto de organizações internacionais, como a concretização da Agenda 2030 para o Desenvolvimento Sustentável, bem como para a abordagem multilateral à governação da Internet" (União Europeia, 2023).

No Brasil, foi aprovado na Câmara dos Deputados o PL n. 21/2020 de autoria do Deputado Eduardo Bismarck (PDT-CE) em 29 de setembro de 2021 que objetiva estabelecer fundamentos, princípios e diretrizes para o desenvolvimento e a aplicação da inteligência artificial. O PL n. 21/2020 foi encaminhado ao Senado pelo Presidente da Câmara dos Deputados no dia 30 de setembro de 2021. No Senado, foi criada uma comissão de juristas para analisar o Projeto de Lei, que foi instalada em março de 2022. A comissão, composta por 18 juristas, após longo processo dialógico com a sociedade que contemplava audiências públicas, criação grupos temáticos (compreensão, conceitos, impactos, classificação de IA, direitos, deveres, transparência, governança, prestação de contas e fiscalização), seminários, reuniões com a participação de especialistas e representantes brasileiros e internacionais para discutir a necessidade de regular a IA e examinar experiências de regulação em outros países, entregou ao Presidente do Senado Federal em dezembro de 2022 um extenso relatório (Brasil, 2022).

O Ministro Ricardo Villas Bôas Cueva, do Superior Tribunal de Justiça, foi o portador do relatório final da Comissão de Juristas ao Presidente do Senado, com a apresentação de um substitutivo aos Projetos de Leis (PLs) n. 5.05/2019, 21/2020 e 872/2021, que pode ser o embrião da regulamentação no Brasil. O Ministro do STJ observa que o relatório representa um amplo e profundo trabalho desenvolvido pelos juristas refletido nas mais de 900 páginas, representando o que a sociedade espera da regulação da IA no Brasil. O objetivo de um Marco Legal para IA é "estabelecer direitos para proteção do elo mais vulnerável em questão, a pessoa natural que já é diariamente impactada por sistemas de inteligência artificial", bem como dispor sobre ferramentas de governança, arranjos institucionais de fiscalização e supervisão, propiciar segurança jurídica para inovação e o desenvolvimento econômico-tecnológico (Brasil, 2022).

O art. 2º do substitutivo apresentado pela Comissão de Juristas dispõe que a IA no Brasil deve estar assentada nos seguintes fundamentos: centralidade da pessoa humana, no livre desenvolvimento da personalidade, respeito aos direitos humanos e aos valores democráticos, igualdade, a não discriminação, pluralidade e os respeito aos direitos trabalhistas, privacidade, proteção de dados e autodeterminação informativa; proteção ao meio ambiente e o desenvolvimento sustentável, inovação e desenvolvimento tecnológico, princípios da livre iniciativa, livre concorrência e defesa do consumidor, promoção da pesquisa, acesso à informação e à educação (Brasil, 2022).

No mês de maio de 2023, o Senador Rodrigo Pacheco apresentou um novo Projeto de Lei, tendo como base as conclusões da Comissão de Juristas. São definidos fundamentos e princípios gerais para o desenvolvimento e a utilização de sistemas de IA, com um capítulo específico para proteger os direitos das pessoas afetadas por sistemas de IA, categorias de riscos de IA, regras de responsabilização civil, governança, transparência, mitigação de vieses, avaliação do impacto algorítmico, proteção contra a discriminação, fiscalização, di-

reitos autorais, propriedade intelectual e fomento a inovação (Pacheco, 2023).

O aplicativo ChatGPT-4 pela OpenAI, que chamou a atenção pelo rápido crescimento em número de usuários, fez com que as concorrentes realizassem esforços para não perderem mercado, acendendo um alerta na comunidade internacional para regulamentar a IA pelos riscos à privacidade, segurança pública, educação, riscos à informação, impactos na segurança nacional, antes que as inovações se transformem em ameaças catastróficas sem possibilidade de retorno.

Uma das principais pesquisadora que analisa o viés algorítmico é Meredith Broussard, professora associada na Universidade de Nova York (EUA), que observa que os computadores são excelentes para resolverem problemas pela matemática, mas não são bons para resolver problemas sociais, contudo, como a nova ordem é usar sistemas algoritmos alimentados por *big data* já estão presentes para solucionar questões sociais, com o que a pesquisadora não concorda. A ciência de dados apresenta muitas limitações para lidar com questões sociais e a auditoria algorítmica é um campo novo, mas será fundamental para examinar parcialidades e possíveis tendenciosidades por comportarem vieses codificadores e arraigados, reduzindo a responsabilidade e inferindo informações. Pesquisadores das áreas da ciência da computação, ciências sociais, direito e humanidades criaram a Conferência FAccT como um espaço interdisciplinar para discutir como tornar os algoritmos mais justos e equitativos, justiça algorítmica, transparência, responsabilidade e se as decisões devem ser terceirizadas para sistemas de *machine learning* (MIT Technology Review, 2023b).

Nenhuma análise preditiva poderá indicar com segurança que não haverá risco discriminatório de sistemas de IA. Por isso, desenvolvedores e governos devem trabalhar juntos para garantir que a IA seja usada de maneira ética e justa. Sendo uma área de crescimento exponencial, a IA traz consigo inúmeros desafios éticos e morais como amplamente discutido no presente estudo, podendo ser programada com vieses preconceituosos, reproduzir e acirrar desigualda-

des já impregnadas na sociedade ou criar novas desigualdades, tomar decisões injustas mesmo não tendo a intenção e não tendo consciência de vieses discriminatórios. Razão pela qual, é urgente que os Estados e organismos internacionais tomem a iniciativa de exigir um arcabouço legal para garantir a todo cidadão transparência e responsabilidade das decisões tomadas por algoritmos. As normativas devem estar alicerçadas nas premissas éticas discutidas neste estudo devendo ser exigido que a decisão algorítmica seja transparente e não discriminatória.

5. Conclusão

A tomada de decisões humanas está a cada dia mais sendo substituída por algoritmos, o que pode exacerbar ainda mais a desigualdade e ampliar preconceitos, manipulação, discriminação em relação ao sexo, raça, origem étnica ou social, idade, orientação sexual, cor da pele, características genéticas, religião ou crença, idioma, opinião política, censura, violações relacionadas à privacidade, igualdade de tratamento, autonomia e ao livre desenvolvimento da personalidade. Com impactos cada vez mais amplos e com danos já observados em indivíduos e na sociedade, preocupações éticas e de governança de algoritmos passam a exigir iniciativas legislativas para regular a IA.

É inegável que o desenvolvimento tecnológico traz grandes vantagens para a sociedade. Controle de tráfico, reconhecimento facial, desenvolvimento de novos medicamentos mais assertivos e com menos efeitos colaterais, cirurgias robotizadas, diagnósticos de doenças com a identificação de sintomas precoces, tarefas domésticas, acesso e compartilhamento de informação, em fim, são incontáveis as áreas alcançadas pelas novas tecnologias. Não obstante, todo esse avanço, novos riscos também são gerados como a discriminação algorítmica que atinge diretamente a igualdade, a distribuição robotizada de *fake news* que fragilizam as democracias, a reprodução de decisões discriminatórias que eternizam uma segregação que

a muito já deveriam ter sido erradicadas, demonstram a necessidade de uma intervenção estatal.

A necessidade de regulação na área de IA não surge com a premissa de restringir seu acesso e desenvolvimento. *Ex surge,* sim, sob a perspectiva de proteger o cidadão do uso indevido dessa ferramenta. Em um mundo onde as fronteiras estão cada vez mais porosas, a normatização perpassa sob uma análise global, ou seja, as regulações estatais devem estar, de certa forma, em sincronia, sob pena de embates jurídicos esvaziarem qualquer forma de proteção. Daí vem a necessidade de estudar e debater outras formas, ou tentativas, de regulação, como a proposta de Lei para a IA pelo Parlamento Europeu e do Conselho ou da Recomendação sobre a Ética da Inteligência Artificial da Unesco.

Sob essas lentes é possível afirmar que o vértice de todas as propostas de regulação da IA passam pela centralidade no cidadão. A partir daí, outros objetivos passam a ser buscados através de uma normatização estatal como a prioridade de acesso das pessoas no processo de transformação digital, a solidariedade e inclusão, a educação, formação e competências digitais, condições de trabalho justas e equitativas, os serviços públicos digitais em linha, a liberdade de escolha para interações com algoritmos e sistemas de inteligência artificial e um ambiente justo, a participação no espaço público digital, um ambiente digital protegido e seguro, a privacidade e controle individual dos dados, a sustentabilidade, dentre outros.

Em definitivo, a regulação da IA visa impedir que o desenvolvimento tecnológico tenha um caráter unicamente econômico, mantendo, assim, seu viés na centralidade humana. Tem por objetivo a manutenção do caráter inovador de uma sociedade informatizada, sem perder o foco com as preocupações éticas, sociais e políticas.

Referências

Batalla, A. R. (2010) 'Intimidad y administración electrónica', in Hueso, L. C., Torrijos, J. V. (coord.) *Administración electrónica: la ley 11/2007, de 22 de junio, de acceso electrónico de los ciudadanos a los Servicios Públicos y los retos jurídicos del e-gobierno en España*, Valencia, Tirant do Blanch, pp. 729-748.

Borgesius, F. J. Z. (2020) 'Strengthening legal protection against discrimination by algorithms and artificial intelligence', *The International Journal of Human Rights,* 24(10), pp. 1572-1593 [online]. Available at: https://www.tandfonline.com/doi/full/10.1080/13642987.2020.1743976?src=recsys (Accessed: 01 August 2023).

Brasil (2022) Senado Federal**,** *Relatório final comissão de juristas responsável por subsidiar elaboração de substitutivo sobre Inteligência Artificial no Brasil* [online]. Available at: https://www.stj.jus.br/sites/portalp/SiteAssets/documentos/noticias/Relato%cc%81rio%20final%20CJSUBIA.pdf (Accessed: 01 August 2023).

Cardial, I. (2023) 'Com recrutamento às cegas, Jobecam esconde para visibilizar'. *Re*/*set* [online]. Available at: https://capitalreset.uol.com.br/diversidade/com-recrutamento-as-cegas-jobecam-esconde-para-visibilizar/ (Accessed: 01 August 2023).

Council of Europe (2018) *Study on the Human Rights Dimensions of Automated Data Processing Techniques (in particular Algorithms) and Possible Regulatory Implications.* Prepared by the Committee of Experts on Internet Intermediaries (MSI-NET). [online]. Available at: https://rm.coe.int/algorithms-and-human-rights-en-rev/16807956b5 (Accessed: 01 August 2023).

Del Rey, A. (2023) *O que está por trás do ChatCPT, e por que ele é tão impressionante. Associação Internacional de Inteligência Artificial (I2AI).* [online]. Available at: https://www.i2ai.org/content/blog/2023/3/o-que-esta-por-tras-do-chatgpt-e-por-que-ele-e-tao/#:~:text=O%20futuro%20continua%20promissor%20para,100%20milh%C3%B5es%20de%20usu%C3%A1rios%20ativos. (Accessed: 01 August 2023).

European Commission (EC) (2021) *Proposal for a regulation of the european parliament and of the council laying down harmonised rules on artificial intelligence and amending certain union legislative acts (artificial intelligence act)*. European Commission, Brussels, Belgium.

European Union (EU) (2020) European Parliamentary Research Service, *The ethics of artificial intelligence: Issues and initiatives* [online]. Available at: https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)634452_EN.pdf (Accessed: 01 August 2023).

European Union (EU) (2018) European Union Agency for Fundamental Rights (FRA), *#BigData: Discrimination in data-supported decision making* [online]. Available at: http://fra.europa.eu/sites/default/files/fra_uploads/fra-2018-focus-big-data_en.pdf (Accessed: 01 August 2023).

Floridi, L. and Taddeo, M. (2016) 'What is data ethics? Subject Areas: Author for correspondence', *Philos Trans Ser A*, 374 (2083) [online]. DOI: https://doi.org/10.1098/rsta.2016.0360 (Accessed: 01 August 2023).

Heaven, W. D. (2023) 'A IA está inventando drogas que ninguém jamais viu', *MIT Technology Review* [online]. Available at: https://www.technologyreview.com/2023/02/15/1067904/ai-automation-drug-development/?truid=&utm_source=the_algorithm&utm_medium=email&utm_campaign=the_algorithm.unpaid.engagement&utm_content=02-20-2023 (Accessed: 01 August 2023).

Heikkilä, M. (2022) 'A quick guide to the most important AI law you've never heard of. The European Union is planning new legislation aimed at curbing the worst harms associated with artificial intelligence', *MIT Technology Review* [online]. Available at: https://www.technologyreview.com/2022/05/13/1052223/guide-ai-act-europe/?utm_campaign=site_visitor.unpaid.engagement&utm_medium=tr_social&utm_source=Twitter (Accessed: 01 August 2023).

Hossain, D., Scott, S. H., Cluff, T. and Dukelow, S. P. (2023) 'The use of machine learning and deep learning techniques to assess proprioceptive impairments of the upper limb after stroke', *Journal of NeuroEngineering and Rehabilitation*. 20(15) [online]. Available at: https://jneuroengrehab.biomedcentral.com/articles/10.1186/s12984-023-01140-9 (Accessed: 01 August 2023).

Inovação Tecnológica (2023) *Óculos equipados com IA entendem fala silenciosa* [online]. Available at: https://www.inovacaotecnologica.com.br/noticias/noticia.php?artigo=oculos-equipados-ia-entendem-fala-silenciosa&id=010150230410&ebol=sim#.ZDWZaXbMJPZ (Accessed: 01 August 2023).

Janiesch, C., Zschench, P. and Heinrich, K. (2021) 'Machine learning and deep learning'. *arXiv.org*, Springer, [2331-8422]. Available at: https://arxiv.org/ftp/arxiv/papers/2104/2104.05314.pdf (Accessed: 01 August 2023).

Kim, P. T. (2017) 'Data-driven discrimination at work', *William & Mary Law Review,* 53 (3), pp. 857-936 [online]. Available at: https://scholarship.law.wm.edu/cgi/viewcontent.cgi?article=3680&context=wmlr (Accessed: 01 August 2023).

Larson, J., Mattu, S., Kirchner, L. and Angwin, J. (2016) 'We Analyzed the Compas Recidivism Algorithm', *ProPublica* [online]. Available at: https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm. (Accessed: 01 August 2023).

Latif, J. et al (2019) 'Medical imaging using machine learning and deep learning algorithms: a review', *International conference on computing, mathematics and engineering technologies (iCoMET)* IEEE, pp. 1-5.

LeCun, Y., Bengio, Y. and Hinton, G. (2015) 'Aprendizado profundo', *Natureza,* 521 :436-444.

Limberger, T. (2007) *O direito à intimidade na era da informática: a necessidade de proteção dos dados pessoais*, Porto Alegre: Livraria do Advogado.

Lloyd, S., Mohseni, M. and Rebentrost, P. (2013) 'Quantum algorithms for supervised and unsupervised machine learning', *arXiv preprint arXiv:1307.0411*.

Melo, A. K. A., Souza, G. C., Vasco, A. C. and Reis, B. S. (2022) *Regulação da Inteligência Artificial: Benchmarking de países selecionados*. Brasília: EvEx e Enap.

Menke, F. (2015) 'A proteção de dados e o novo direito fundamental à garantia da confidencialidade e da integridade dos sistemas técnicos-informacionais no direito alemão' in Mendes, G. F., Sarlet, I. W. and Coelho, A. Z. P. (coord.), *Direito, inovação e tecnologia,* São Paulo: Saraiva, p. 205-230.

MIT Technology Review (2022) *Emerging Technologies Agro –– 2022***,** Special Edition, [online]. Available at: https://rd.mittechreview.com.br/special-edition-tecnologiasno agro-cubo (Accessed: 01 August 2023).

MIT Technology Review. (2023a) *As tecnologias capazes de ler nossas mentes estão chegando.* O que fazer? [online]. Available at: https://mittechreview.com.br/as-tecnologias-capazes-de-ler-nossas-mentes-estao-chegando-o-que-fazer/ (Accessed: 01 August 2023).

MIT Technology Review. (2023b) *Conheça a especialista em Inteligência Artificial que acredita que deveríamos reduzir o uso da tecnologia* [online]. Available at: https://mittechreview.com.br/conheca-a-especialista-em-inteligencia-artificial-que-acredita-que-deveriamos-reduzir-o-uso-da-tecnologia/ (Accessed: 01 August 2023).

Morley, J. et al (2020) 'From what to how: an initial review of publicly available AI ethics tools, methods and research to translate principles into practices', *Science and engineering ethics,* 26(4), pp. 2141-2168 [online]. Available at: https://link.springer.com/article/10.1007/s11948-019-00165-5#Tab1 (Accessed: 01 August 2023).

Navarro, S. N. (2017) 'Derecho e inteligencia artificial desde el diseño. Aproximaciones', in Navarro, S. N. (dir.), *Inteligencia artificial, tecnología, derecho.* Valencia: Tirant to Blach, p. 23-72.

Organisation For Economic Cooperation And Development (OECD). (2019) Committee on Digital Economy Policy (CDEP). *Recommendation of the Council on Artificial Intelligence 2019* [online]. Available at: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 (Accessed: 01 August 2023).

Organização das Nações Unidas para a Educação, a Ciência e a Cultura (Unesco) (2022) *Recomendação sobre a Ética da Inteligência artificial.* Paris: Unesco [online]. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000381137_por (Accessed: 01 August 2023).

Pacheco, R. (2023) *Projeto de Lei que dispõe sobre a Inteligência Artificial*. Brasília, Senado Federal.

Rodotà, S. (2008) *A vida na sociedade da vigilância: a privacidade hoje*. Translated by Danilo Doneda and Luciana Cabral Doneda. Rio de Janeiro: Renovar.

Schmidt, N. and Stephens, B. (2019) 'An Introduction to Artificial Intelligence and Solutions to the Problems of Algorithmic Discrimination', *Quarterly Report*, 73(2), pp. 130-144. [online]. Available at: https://arxiv.org/ftp/arxiv/papers/1911/1911.05755.pdf (Accessed: 01 August 2023).

Singapore (2022) 'Personal Data Protection Commission Singapore', *Singapore's Approach to AI Governance,* 2022 [online]. Available at: https://www.pdpc.gov.sg/help-and-resources/2020/01/model-ai-governance-framework (Accessed: 01 August 2023).

Tsamados, A., Aggarwal, N., Cowls, J. Morley, J., Roberts, H., Taddeo, M. and Floridi, L. (2022) 'The ethics of algorithms: key problems and solutions', *AI & SOCIETY*, 37, pp. 215–230. Available at: https://link.springer.com/article/10.1007/s00146-021-01154-8 (Accessed: 01 August 2023).

União Europeia. (2016) *Regulamento (UE) 2016/679 (General Data Protection Regulation) do Parlamento Europeu e do Conselho de 27 de abril de 2016*. [online]. Available at: https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=CELEX:32016R0679& from=EN (Accessed: 01 August 2023).

União Europeia. (2023). *Declaração Europeia sobre os direitos e princípios digitais para a década digital (2023/C 23/01).* [online]. Available at: https://eur-lex.europa.eu/legal-content/PT/TXT/HTML/?uri=CELEX:32023C0123(01)&qid=1681768365786&from=PT (Accessed: 01 August 2023).

Xenidis, R. and Senden, L. (2020) 'EU non-discrimination law in the era of artificial intelligence: Mapping the challenges of algorithmic discrimination', General Principles of EU law and the EU Digital Order, *Kluwer Law International,* pp. 151-182. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3529524 (Accessed: 01 August 2023).

Zarsky, T. (2013) 'Transparent predictions', *University of Illinois Law Review*, 2013(4), pp. 1503-1570. Available at: https://www.illinoislawreview.org/wp-content/ilr-content/articles/2013/4/Zarsky.pdf (Accessed: 01 August 2023).