

**XII ENCONTRO INTERNACIONAL DO
CONPEDI BUENOS AIRES –
ARGENTINA**

**DIREITO, GOVERNANÇA E NOVAS TECNOLOGIAS
II**

JONATHAN BARROS VITA

YURI NATHAN DA COSTA LANNES

PAULO CAMPANHA SANTANA

Todos os direitos reservados e protegidos. Nenhuma parte deste anal poderá ser reproduzida ou transmitida sejam quais forem os meios empregados sem prévia autorização dos editores.

Diretoria - CONPEDI

Presidente - Prof. Dr. Orides Mezzaroba - UFSC - Santa Catarina

Diretora Executiva - Profa. Dra. Samyra Haydêe Dal Farra Naspolini - UNIVEM/FMU - São Paulo

Vice-presidente Norte - Prof. Dr. Jean Carlos Dias - Cesupa - Pará

Vice-presidente Centro-Oeste - Prof. Dr. José Querino Tavares Neto - UFG - Goiás

Vice-presidente Sul - Prof. Dr. Leonel Severo Rocha - Unisinos - Rio Grande do Sul

Vice-presidente Sudeste - Profa. Dra. Rosângela Lunardelli Cavallazzi - UFRJ/PUCRio - Rio de Janeiro

Vice-presidente Nordeste - Profa. Dra. Gina Vidal Marcilio Pompeu - UNIFOR - Ceará

Representante Discente: Prof. Dra. Sinara Lacerda Andrade - UNIMAR/FEPODI - São Paulo

Conselho Fiscal:

Prof. Dr. Caio Augusto Souza Lara - ESDHC - Minas Gerais

Prof. Dr. João Marcelo de Lima Assafim - UCAM - Rio de Janeiro

Prof. Dr. José Filomeno de Moraes Filho - Ceará

Prof. Dr. Lucas Gonçalves da Silva - UFS - Sergipe

Prof. Dr. Valter Moura do Carmo - UNIMAR - São Paulo

Secretarias

Relações Institucionais:

Prof. Dra. Daniela Marques De Moraes - UNB - Distrito Federal

Prof. Dr. Horácio Wanderlei Rodrigues - UNIVEM - São Paulo

Prof. Dr. Yuri Nathan da Costa Lannes - Mackenzie - São Paulo

Comunicação:

Prof. Dr. Liton Lanes Pilau Sobrinho - UPF/Univali - Rio Grande do Sul

Profa. Dra. Maria Creusa De Araújo Borges - UFPB - Paraíba

Prof. Dr. Matheus Felipe de Castro - UNOESC - Santa Catarina

Relações Internacionais para o Continente Americano:

Prof. Dr. Heron José de Santana Gordilho - UFBA - Bahia

Prof. Dr. Jerônimo Siqueira Tybusch - UFSM - Rio Grande do Sul

Prof. Dr. Paulo Roberto Barbosa Ramos - UFMA - Maranhão

Relações Internacionais para os demais Continentes:

Prof. Dr. José Barroso Filho - ENAJUM

Prof. Dr. Rubens Beçak - USP - São Paulo

Profa. Dra. Viviane Coêlho de Séllos Knoerr - Unicuritiba - Paraná

Eventos:

Prof. Dr. Antônio Carlos Diniz Murta - Fumec - Minas Gerais

Profa. Dra. Cinthia Obladen de Almendra Freitas - PUC - Paraná

Profa. Dra. Livia Gaigner Bosio Campello - UFMS - Mato Grosso do Sul

Membro Nato - Presidência anterior Prof. Dr. Raymundo Juliano Feitosa - UMICAP - Pernambuco

D597

Direito, Governança e novas tecnologias II [Recurso eletrônico on-line] organização CONPEDI

Coordenadores: Jonathan Barros Vita; Paulo Campanha Santana; Yuri Nathan da Costa Lannes. – Florianópolis: CONPEDI, 2023.

Inclui bibliografia

ISBN: 978-65-5648-831-8

Modo de acesso: www.conpedi.org.br em publicações

Tema: Derecho, Democracia, Desarrollo y Integración

1. Direito – Estudo e ensino (Pós-graduação) – Encontros Nacionais. 2. Direito. 3. Governança e novas tecnologias. XII Encontro Internacional do CONPEDI Buenos Aires – Argentina (2: 2023 : Florianópolis, Brasil).

CDU: 34



XII ENCONTRO INTERNACIONAL DO CONPEDI BUENOS AIRES – ARGENTINA

DIREITO, GOVERNANÇA E NOVAS TECNOLOGIAS II

Apresentação

APRESENTAÇÃO

O Grupo de Direito, Governança e Novas Tecnologias II teve seus trabalhos apresentados no dia 13 de outubro de 2023, com início às 14h, durante o XII ENCONTRO INTERNACIONAL DO CONPEDI, ARGENTINA – BUENOS AIRES, que ocorreu nos dias 12, 13 e 14 de outubro, com o tema: DIREITO, DEMOCRACIA, DESENVOLVIMENTO E INTEGRAÇÃO.

O Grupo de Trabalho teve 16 (dezesseis) apresentações que trataram dos seguintes temas:

A ÉTICA ALGORÍTMICA: O DESAFIO NO AVANÇO DA INTELIGÊNCIA ARTIFICIAL. De Jamile Sabbad Carecho Cavalcante e Paulo José Pereira Carneiro Torres da Silva, o artigo trata da análise do avanço do estabelecimento de padrões éticos para o desenvolvimento e o uso de sistemas de inteligência artificial no mundo, com destaque para o avanço da regulamentação brasileira acerca da temática.

O DESAFIO DO DEVIDO PROCESSO LEGAL NO DESENVOLVIMENTO DE INTELIGÊNCIA ARTIFICIAL NO PODER JUDICIÁRIO BRASILEIRO. De Jamile Sabbad Carecho Cavalcante e Paulo José Pereira Carneiro Torres da Silva, o artigo analisa o desenvolvimento de sistemas inteligentes no Poder Judiciário brasileiro e sua conformidade com o devido processo legal diante da inteligência artificial generativa e da discriminação algorítmica, considerando os desafios na efetivação dos direitos fundamentais.

A INFLUÊNCIA CRESCENTE DA INTELIGÊNCIA ARTIFICIAL NAS RELAÇÕES DE TRABALHO NO BRASIL. De Morgan Stefan Grando, Julia Brezolin e Ipojuca Demétrius Vecchi, o artigo analisa as principais mudanças no mercado do trabalho promovidas pela inteligência artificial (IA), com ênfase no Brasil.

O IMPACTO DA INTELIGÊNCIA ARTIFICIAL NA DISSEMINAÇÃO DO DISCURSO DE ÓDIO ONLINE E MEIOS DE CONTROLE. De Felipe Pinheiro Prestes e Gustavo

Silveira Borges, o artigo trata da proliferação do discurso de ódio nas mídias sociais e a Inteligência Artificial (IA), inclusive a generativa, com análise dos impactos e dos possíveis meios de contenção dessas práticas.

A IMPOSSIBILIDADE DE CONCESSÃO DE PATENTES A SISTEMAS DOTADOS DE INTELIGÊNCIA ARTIFICIAL. De William Andrade, Salete Oro Boff e Joel Marcos Reginato, o artigo discorre, sob a ótica do caso Dabus, a concessão de patentes de invenções para sistemas autônomos dotados de Inteligência Artificial, com base no sistema jurídico brasileiro.

QUAL A SOLUÇÃO PARA A CONCESSÃO, OU NÃO, DE PATENTES A INVENÇÕES CRIADAS POR INTELIGÊNCIA ARTIFICIAL? De Joel Marcos Reginato, Salete Oro Boff e William Andrade, o artigo busca analisar como devem ser tratadas, juridicamente, as invenções provenientes de sistemas de Inteligência Artificial, considerando-se se é devida ou não a concessão de patentes a tais sistemas.

CHAT GPT E O ENSINO JURÍDICO PARA A PREVENÇÃO DE CONFLITOS. De Kátia Cristina Stamberk e Augusto Martinez Perez Filho, o artigo explora a relação entre o ensino jurídico no Brasil, as Diretrizes Curriculares Nacionais e o uso do Chat GPT como uma ferramenta educacional.

PLATAFORMAS DE COMUNICAÇÃO ELETRÔNICA: O DUELO ENTRE PRIVACIDADE E PROTEÇÃO DE DADOS. De Luziane De Figueiredo Simão Leal e Brychtn Ribeiro de Vasconcelos, o artigo aborda o conflito entre proteção de dados e o habitat nas plataformas digitais surgido a partir da difusão e evolução das tecnologias de informação.

SOCIEDADE INFORMACIONAL E A LGPD: A RESPONSABILIDADE CIVIL DO MÉDICO PARTICULAR FRENTE AO DIREITO FUNDAMENTAL DE PROTEÇÃO DE DADOS DO PACIENTE. De Andressa Camoleze Alessi e Deise Marcelino Da Silva, o artigo aborda a responsabilidade civil do médico na relação com o paciente, tendo em conta a Lei Geral de Proteção de Dados (LGPD) e seu impacto sobre a proteção dos dados pessoais.

A PROTEÇÃO DE DADOS COMO DIREITO FUNDAMENTAL E O PAPEL DOS CARTÓRIOS EXTRAJUDICIAIS NA COLETA E TRATAMENTO DOS DADOS: UMA ANÁLISE A PARTIR DO PROVIMENTO 134/2022 CNJ. De Carlos Renato Cunha e Ana Maria Scarduelli Gurgel, o artigo analisa o papel dos cartórios extrajudiciais na coleta e tratamento dos dados dos usuários.

CIBERESPAÇO E AS NOVAS TECNOLOGIAS: A EXCLUSÃO DIGITAL NA SOCIEDADE DA INFORMAÇÃO. De Julia Brezolin , Morgan Stefan Grando e Liton Lanes Pilau Sobrinho, o artigo analisa o desenvolvimento da cibernética e das novas tecnologias ao longo do tempo e os impactos causados pela difusão do computador e da internet na sociedade.

OBSTÁCULOS TECNOLÓGICOS: UM ESTUDO DA EVOLUÇÃO HISTÓRICA DOS DIREITOS HUMANOS E A CONCRETIZAÇÃO EM TEMPOS DE INOVAÇÕES TECNOLÓGICAS. De Clarisse Yamauchi e José Carlos Francisco dos Santos, o artigo analisa os principais obstáculos, em especial os tecnológicos, que dificultaram a efetivação e concretização dos direitos humanos e fundamentais elencados no direito internacional e Constituição Federal de 1988.

O DIREITO DA PERSONALIDADE DIGITAL. De Kátia Cristina Stamberk, Andressa de Souza e Silva e Aline Ouriques Freire Fernandes, o artigo analisa como a LGPD contribui para a proteção dos direitos da personalidade nas mídias sociais.

O IMPACTO DA TECNOLOGIA NA ECONOMIA BRASILEIRA: DESAFIOS LEGAIS E ÉTICOS. De Andressa Camoleze Alessi e Deise Marcelino Da Silva, o artigo analisa os impactos da tecnologia na economia, destacando os desafios legais e éticos.

DATA CENTERS SOB O ENFOQUE DO DIREITO ECNÔMICO AMBIENTAL: NECESSIDADE DE REGULAÇÃO OU AUTORREGULAÇÃO COM VISTAS À SUSTENTABILIDADE. De Carlos Renato Cunha e Ana Maria Scarduelli Gurgel, o artigo busca identificar a necessidade de regulação ou a possibilidade de autorregulação normativa deste segmento.

TECNOLOGIA DA INFORMAÇÃO EM REDE: UM ESTUDO DA ADMINISTRAÇÃO PÚBLICA E SEU PROCESSO DE PRODUTOS E SERVIÇOS. De Clarisse Yamauchi e José Carlos Francisco dos Santos, o artigo apresenta os conceitos da administração pública material e dos princípios constitucionais que norteiam as compras públicas, as inovações de compras públicas e um contexto digital e de uso de internet.

Agradecemos aos colaboradores pelas pesquisas desta obra, desejando a todos uma proveitosa leitura!

Coordenadores:

Professor Dr. Paulo Campanha Santana – Centro Universitário do Distrito Federal (UDF)

Jonathan Barros Vita – Universidade de Marília

Yuri Nathan da Costa Lannes – Faculdade de Direito de Franca

A ÉTICA ALGORÍTMICA: O DESAFIO NO AVANÇO DA INTELIGÊNCIA ARTIFICIAL.

ALGORITHMIC ETHICS: THE CHALLENGE IN THE ADVANCEMENT OF ARTIFICIAL INTELLIGENCE.

Jamile Sabbad Carecho Cavalcante ¹
Paulo José Pereira Carneiro Torres da Silva ²

Resumo

O presente trabalho visa apresentar uma análise do avanço do estabelecimento de padrões éticos para o desenvolvimento e o uso de sistemas de inteligência artificial no mundo, destacando o avanço da regulamentação brasileira acerca da temática. O problema central é a necessidade de uma regulação ética para o avanço responsável do desenvolvimento e implementação desses sistemas ao longo dos próximos anos a fim de assegurar os direitos fundamentais e os novos direitos advindos de uma interpretação adequada ao avanço tecnológico. A preocupação está principalmente em três momentos na elaboração/arquitetura do sistema, no tratamento e inserção dos dados e na interpretação dos resultados pelo ser humano. Para isso, serão analisados artigos científicos nacionais e internacionais sobre a temática da ética algorítmica, além da Carta Europeia de Ética sobre o Uso da Inteligência Artificial em Sistemas Judiciais e seu ambiente, além do Código de Ética da Association for Computing Machinery (Associação para Maquinaria da Computação - ACM) e do Instituto de Engenheiros Eletrônicos e Eletricistas (IEEE) e a Resolução nº 332/2020 do Conselho Nacional de Justiça, bem como a Recomendação sobre Inteligência Artificial da Organização para Cooperação e Desenvolvimento Econômico (OCDE) que o Brasil aderiu na condição de país não-membro em 2019 e a Estratégia Brasileira de Inteligência Artificial (EBIA) e os projetos de lei que tramitam no Congresso Nacional sobre a temática.

Palavras-chave: Inteligência artificial (ia), Ética algorítmica, Discriminação algorítmica, Transparência, Regulamentação

Abstract/Resumen/Résumé

This paper aims to present an analysis of the progress in establishing ethical standards for the development and use of artificial intelligence systems worldwide, highlighting the advancement of Brazilian regulations on the subject. The central problem is the need for ethical regulation to ensure responsible advancement in the development and implementation of these systems in the coming years, in order to safeguard fundamental rights and new rights

¹ Mestranda em Direito Público na UNESA/RJ com ênfase em Direitos Fundamentais e Novos Direitos, bolsista Prosup/CAPES. <jamilesabbad@gmail.com> Pesquisadora da Inteligência Artificial no Poder Judiciário e seus desafios.

² Professor do PPGD da UNESA; Coordenador do Curso de Direito da Estácio de Sá – Copacabana; Doutor em Direito Público pela UNESA/RJ (2021); Mestre em Direito Constitucional pelo PPGDC-UFF (2017)

arising from an appropriate interpretation of technological advancements. The concern primarily lies in three aspects: the system's design/architecture, the handling and integration of data, and the interpretation of results by humans. To this end, national and international scientific articles on the topic of algorithmic ethics will be analyzed, as well as the European Charter of Ethics for the Use of Artificial Intelligence in Judicial Systems and its environment, the Code of Ethics of the Association for Computing Machinery (ACM) and the Institute of Electrical and Electronics Engineers (IEEE), Resolution number 332/2020 of the National Council of Justice, the Recommendation on Artificial Intelligence by the Organization for Economic Cooperation and Development (OECD) that Brazil joined as a non-member country in 2019, the Brazilian Artificial Intelligence Strategy (BAIS), and the bills currently being discussed in the National Congress on this subject.

Keywords/Palabras-claves/Mots-clés: Artificial intelligence (ai), Algorithmic ethics, Algorithmic discrimination, Transparency, Regulation

1. Introdução:

Nos últimos anos vêm aumentando progressivamente o número de decisões importantes que estão sendo delegadas aos sistemas de inteligência artificial no âmbito público e privado, desde a concessão ou não de um crédito pessoal à possibilidade de responder um processo em liberdade com pagamento de fiança, impactando diretamente no exercício de direitos fundamentais na sociedade moderna.

Embora a maioria dos sistemas tenha sido desenvolvido para servirem de apoio a decisões humanas afim de agilizar os processos decisórios, a realidade tem se demonstrado diferente pela crença de maior acertabilidade das decisões quando feitas por máquinas acreditando que a inteligência artificial seria superior à inteligência humana, esquecendo que decisões humanas criaram aquele modelo e abstenções podem resultar em vieses pelo mais variados motivos, vinculam-se cada vez mais aos julgamentos de sistemas de IA mesmo sem supervisão humana.

Por vezes se esquece que decisões humanas criaram o modelo, decidindo a programação, os dados imputados, implementando o resultado, entre outras etapas, se as decisões humanas podem ser falhas, preconceituosas ou equivocadas, como garantir que decisões algorítmicas fruto de diversas decisões humanas não serão falhas? A opacidade das decisões das máquinas e a quantidade de dados disponíveis a partir do *big data* resulta em uma falsa impressão de que as decisões desses sistemas não são erradas ou danosas apenas pela impossibilidade de refutação em grande parte, o modelo em si é uma caixa preta. De acordo com a matemática Cathy O'Neil (2020, p. 8):

(...) Mesmo assim, muitos desses modelos programavam preconceitos, equívocos e vieses humanos nos sistemas de software que cada vez mais geriam nossas vidas. Como deuses, esses modelos matemáticos eram opacos, seus mecanismos invisíveis a todos exceto os altos sacerdotes de seus domínios: os matemáticos e 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 enriquecia ainda mais os ricos.

Primeiramente, é necessário conceituarmos inteligência artificial: embora seja um conceito discutido pelos teóricos apresentando diversas definições possíveis, para fins deste trabalho iremos considerar o conceito utilizado na Estratégia Brasileira de Inteligência Artificial

- EBIA que é o conceito da Organização para Cooperação e Desenvolvimento Econômico (OCDE)¹ com a seguinte definição para Inteligência Artificial (OCDE, *apud*,

¹ Organismo formado por 37 nações que se dedica à pesquisa, desenvolvimento e aprimoramento de políticas públicas fomentando sua implementação pelos países membros. FIA, 2021, Disponível em: < <https://fia.com.br/blog/ocde/> >

EBIA, 2021, p. 8):

Um sistema de IA é um sistema baseado em máquina que pode, para um determinado conjunto de objetivos definidos pelo homem, fazer previsões, recomendações ou tomar decisões que influenciam ambientes reais ou virtuais. Os sistemas de IA são projetados para operar com vários níveis de autonomia.

Outro conceito relevante para os fins do presente artigo é o de *machine learning*, que vem a ser é uma forma específica de treinamento de inteligências artificiais em que as máquinas aprendem como completar uma determinada tarefa sem serem explicitamente programadas para isso.

Na era do *big data*, os algoritmos têm sido alimentados com uma imensidão de dados sendo matéria prima do processo decisório da máquina permitindo assim a possibilidade do *machine learning*. Segundo Isabela Ferrari, Daniel Becker e Erik Navarro (2018, p.4):

A forma mais simples de *machine learning* é aquela que emprega algoritmos supervisionados, na qual o sistema é alimentado com dados lapidados e previamente escolhidos e por seres humanos. Nesse caso, o conjunto de dados rotulados e a saída desejada são carregados no sistema. Enquanto é treinado, o modelo ajusta as suas variáveis para mapear as entradas para a saída correspondente. Um exemplo são os algoritmos utilizados pelos bancos para aprovar a concessão de empréstimos. Nesse caso, os dados analisados serão referentes ao histórico de crédito do cliente, e as informações utilizadas para treinar o sistema são dados já rotulados como positivos ou negativos para a concessão de crédito.

Por fim é preciso considerar a o conceito de viés algorítmico, uma categoria ainda é bastante discutida pela doutrina e que, para fins desse trabalho, será considerado como um resultado enviesado causado por um sistema que discrimina injustamente indivíduos ou grupos ou gera resultados fundados em premissas equivocadas (intencionalmente ou não), podendo ser caracterizado pela própria realidade de distorção de dados, algoritmos construídos de forma pouco cuidadosa ou mesmo por uma seleção de dados enviesada como sendo representativa de uma realidade social pré-existente.

Na doutrina² as principais causas apontadas para o viés algorítmico são a estrutura social pré-existente, os dados enviesados, a construção da ferramenta ou a interpretação dos resultados, enfim, seja qual for a causa de enviesamento, se a sua existência não se discute como iremos observar nos mais variados campos de atuação de sistemas de inteligência artificial, o viés de cognição por si só pode gerar um prejuízo, ainda mais aliada à falta de transparência ou motivação daquela decisão, este prejuízo é majorado. Desta forma, é

² SIMÕES-GOMES, Letícia; ROBERTO, Enrico; MENDONÇA, Jônatas. Viés algorítmico—um balanço provisório. Estudos de Sociologia, v. 25, n. 48, 2020. Disponível em <https://periodicos.fclar.unesp.br/estudos/article/view/13402> Acesso em 10 de maio de 2022.

preciso que seja possível a intervenção humana em todo sistema de IA para preservação dos direitos humanos, da lei e dos princípios democráticos que nos são tão caros.

A ética é definida como a teoria, o conhecimento ou a ciência do comportamento moral que busca explicar, compreender, justificar e criticar a moral ou as morais de uma sociedade³. Portanto, segundo a professora Dora Kaufman (2016, p. 9) ética significa um tipo de comportamento adquirido ou conquistado, mas não instintivo. Dessa forma, se a Inteligência Artificial representa uma nova inteligência que coexiste com a inteligência humana, devemos ter uma nova ética que envolva conceitos éticos tradicionais com a elaboração de novos.

A preocupação está em estabelecer princípios éticos que orientem três momentos importantes da inteligência artificial, primeiramente, na programação do sistema inteligente, após na escolha e inserção dos dados e, por último, na possibilidade de revisão ou interpretação da decisão algorítmica por um ser humano.

Segundo o professor Paulo Caliendo⁴ a elaboração de regras éticas para inteligência artificial se divide em três momentos regulatórios. No primeiro, com a elaboração das três leis da robótica, elaboradas por Isaac Asimov em 1942, são elas: (1) um robô não pode ferir um humano ou permitir que um humano sofra algum mal; (2) os robôs devem obedecer às ordens dos humanos, exceto nos casos em que tais ordens entrem em conflito com a primeira lei; (3) um robô deve proteger sua própria existência, desde que não entre em conflito com as leis anteriores.

Esse modelo dedutivista em que se estabelece leis gerais a serem aplicadas em cada caso apresenta claras contradições, como por exemplo, como o robô vai proteger um humano sem causar mal a ele, se ele estiver com intenção de causar mal a um terceiro ser humano? Não há uma resposta dentro dessas três leis.

Na segunda tentativa busca-se estabelecer regras do piso para o topo, com base no imperativo categórico⁵ de Kant ou no consequencialismo⁶ para criação de um modelo de ética artificial, entretanto, não é possível a verificação de resultados pela alta

³ Comissão de Ética Pública da FURG. *Ética Pública*. Rio Grande: Universidade Federal do Rio Grande, 2022. Disponível em: <https://eticapublica.furg.br/moral-e-etica?id=26> acesso em 28 de novembro de 2022.

⁴ CORDIOLI, Leandro. *Inteligência Artificial e Direito*. Prof. Dr. Paulo Caliendo. Youtube, 25 de janeiro de 2021. Disponível em: <https://www.youtube.com/watch?v=xp-KDkyqIDM>. Acesso em 12 de setembro de 2022.

⁵ “Introduzido na *Fundamentação da Metafísica da Moral* de Kant, de 1785, é uma forma de avaliar as motivações para a ação. É mais conhecido em sua formulação original: “Aja apenas de acordo com aquela máxima pela qual você pode, ao mesmo tempo, querer que ela se torne uma lei universal.” Disponível em https://pt.wikipedia.org/wiki/Imperativo_categórico. Acesso em 28 de novembro de 2022.

⁶ “Consequencialismo é um termo filosófico criado por Elizabeth Anscombe em “Modern Moral Philosophy”, 1958, para defender a tese de que um agente é responsável tanto pelas consequências intencionais de um ato, como pelas não intencionais quando previstas e não evitadas. Portanto, as consequências deveriam ser levadas em consideração quando se faz juízos sobre o correto e incorreto.” Disponível em <https://pt.wikipedia.org/wiki/Consequencialismo>. Acesso em 28 de novembro de 2022

capacidade de processo que inviabilizaria o próprio funcionamento da IA e tão pouco a elaboração de regras universais e gerais que possam ser aplicadas a todos os casos até o momento.

Com isso, restou buscar na ética das virtudes⁷ de Aristóteles a possibilidade de construção de modelos computacionais que aprendem com base no resultado, observa as maneiras de realizar determinada ação e aprendem a melhor decisão para aquela situação, é o aprendizado de máquina pelo reforço e de modo evolucionário. O objetivo não está nas premissas e nem nos resultados, mas no uso virtuoso dos meios para atingir os fins, essa é a base do uso ético de sistemas inteligentes na atualidade.

Neste trabalho utilizamos o método de pesquisa dedutivo apoiado em revisão bibliográfica examinando artigos nacionais e internacionais que explorem a temática da ética algorítmica, definições, causas, diagnósticos e perspectivas apresentadas para o fenômeno, analisando a Carta Europeia de Ética sobre o Uso da Inteligência Artificial em Sistemas Judiciais e seu ambiente, além do Código de Ética da Association for Computing Machinery (Associação para Maquinaria da Computação - ACM) e do Instituto de Engenheiros Eletrônicos e Eletricistas (IEEE), bem como a Recomendação sobre Inteligência Artificial da Organização para Cooperação e Desenvolvimento Econômico (OCDE) que o Brasil aderiu na condição de país não-membro em 2019 e a Estratégia Brasileira de Inteligência Artificial (EBIA) e os projetos de lei que tramitam no Congresso Nacional sobre a temática.

2. O risco do uso do algoritmo em tomada de decisões privadas.

Para que se compreenda a necessidade de regulação ética ao decorrer do avanço do uso e desenvolvimento de sistemas inteligentes é importante sabermos de alguns casos em que mecanismos de inteligências artificiais estão sendo discriminatórios.

No que se refere aos algoritmos de reconhecimento facial, tecnologia amplamente investida no Brasil e já utilizada por alguns estados como o Rio de Janeiro, uma pesquisa realizada pelo *National Institute of Standards and Technology* (Nist)⁸ em dezembro de 2019 por meio do programa de teste de fornecedor de reconhecimento de rosto avaliou 189 algoritmos de software de 99 desenvolvedores e resultou que a maioria da indústria possibilita a ocorrência de falsas acusações com cem vezes mais chances às

⁷ “A virtude, para Aristóteles, é uma prática e não um dado da natureza de cada um, tampouco o mero conhecimento do que é virtuoso, como para Platão (427-347 a.C.). Para ser praticada constantemente, a virtude precisa se tornar um hábito.” Disponível em <https://novaescola.org.br/conteudo/1390/aristoteles-o-defensor-da-instrucao-para-a-virtude>. Acesso em 28 de novembro de 2022.

⁸ *NIST Study Evaluates Effects of Race, Age, Sex on Face Recognition Software*. Disponível em <https://www.nist.gov/news-events/news/2019/12/nist-study-evaluates-effects-race-age-sex-face-recognition-software>. Acesso em 8 de junho de 2022.

peessoas negras, o falso positivo em uma pesquisa de um paramuitos que é utilizado por governos para auxilio na segurança pública.

Nos Estados Unidos após a morte de George Floyd e a onda de protestos no ano de 2021 empresas de tecnologia se posicionaram contra o uso policial da tecnologia⁹, a IBM enviou uma carta ao Congresso Americano se opondo a desenvolver ferramentas de reconhecimento facial que viole direitos humanos, a Amazon informou que irá proibir o uso pela polícia por um ano de seu algoritmo de reconhecimento facial e pediu rigidez nas normas éticas dessas tecnologias e a Microsoft concordou com as demais informando que só voltará a vender a tecnologia para polícia após a regulamentação federal.

No âmbito da saúde, foi constatado que o algoritmo de pontuação de risco fornecido pela *Optum* usado por hospitais para identificar pacientes com doenças crônicas tem viés racial, em estudo realizado pela revista *Science* que trouxe a informação de que o algoritmo reduz pela metade o número de pacientes negros que precisam de cuidados extras, a quantidade de pacientes negros recebendo ajuda adicional no momento da pesquisa era de 17,7%, se não houvesse essa disparidade seria de 46,5%. Isso ocorre pois o algoritmo usa os custos de saúde como um *proxy*, considerando gastos históricos, para triagem e diagnóstico, como menos dinheiro é gasto por pacientes negros historicamente, embora tenham o mesmo nível de necessidade, o algoritmo conclui falsamente que pacientes negros são mais saudáveis do que pacientes brancos igualmente doentes.

Segundo matéria publicada pela *MIT Technology Review* (MIT, 2021), há alguns anos o LinkedIn descobriu que os algoritmos de recomendação que usa para cruzar candidatos e oportunidades de vaga estava sendo tendencioso, mesmo excluindo nome, idade, sexo e raça de uma pessoa, há outras características que contribuem a parcialidade, os candidatos eram classificados de acordo com a probabilidade de se candidatarem a uma posição ou de responderem a um recrutador, resultando em um encaminhamento maior de homens do que de mulheres para os cargos disponíveis, pela agressividade com que o sexo masculino busca novas oportunidades, os homens são mais propensos a vagas que exigem experiência além de suas qualificações, enquanto mulheres se restringem as vagas de acordo com as suas qualificações.

Ainda na matéria veiculada, o *LinkedIn* afirma que criou um algoritmo separado para neutralizar a parcialidade do programa, garantindo que o sistema incluia uma distribuição de gênero uniforme dos usuários antes de encaminhar as correspondências aos recrutadores. Necessário refletir se máquinas podem corrigir máquinas, ou se é

⁹ *Software de inteligência artificial é banido na cidade de George Floyd*. Disponível em <https://epocanegocios.globo.com/Tecnologia/noticia/2021/02/software-de-reconhecimento-facial-e-banido-na-cidade-de-george-floyd.html>. Acesso em 7 de junho de 2022.

necessária supervisão humana das decisões para evitar outros vieses que esse segundo programa revisor pode ter.

Outras plataformas de vagas de empregos apresentam possíveis vieses, como, por exemplo, a *CareerBuilder*, que diz está focada em ensinar os empregadores a eliminar a parcialidade dos anúncios de vagas a partir da coleta de dados, ou a *ZipRecruiter*, que não revela critérios de seleção e diz não levar em consideração características de identificação, mas sim outras 64 informações (sem que no entanto discrimine quais são estes critérios ou como eles são correlacionados), o que, como vimos tanto no caso destacado pelo *LinkedIn*, como no caso do *Compas*, não impede a discriminação.

Conforme trazido pela professora Cathy O’Neil (2020, p. 13) os sistemas matemáticos demandam um feedback, ou seja, saber quando o resultado não foi o esperado, quando “saíram dos trilhos”, a partir dos erros que os estatísticos treinam seus modelos e os fazem cada vez mais inteligentes. Entretanto, pode ocorrer um ciclo tóxico de retroalimentação quando o sistema não é treinado para que entenda os erros como decisões ruins e escolha novos caminhos, fazendo-o crer que estava certo e deve continuar a tomar essas decisões precipitadas como ideais. Quanto à discriminação algorítmica de gênero, a professora Ana Frazão acredita que:

O mercado de tecnologia e desenvolvimento é dominado por agentes do gênero masculino que desenvolvem linguagens e programas com base na suavivência e visão de mundo sem muita transparência a automatização, inteligência artificial e o tratamento de dados em massa podem se tornar poderosas ferramentas para excluir as mulheres do mercado de trabalho ou perpetuarem preconceitos existentes na sociedade. (JOTA, 2021)

O lançamento do *Apple Card*, cartão de crédito da Apple, foi investigado pelo Departamento de Serviços Financeiros do Estado de Nova York por sexismo (BBC, 2019) após a divulgação de posts no Twitter que demonstravam a diferenciação de limites de crédito para homens e mulheres, o influente programador dinamarquês David Heinemeier Hansson, uma das pessoas a iniciar essa discussão na rede social, postou que embora ele e a mulher fizessem declaração conjunta de Imposto de Renda, viverem no mesmo imóvel e serem casados há muito tempo, o algoritmo da Apple deu um limite de crédito 20 vezes maior a ele do que à sua esposa.

Um ano e meio após o início das investigações, que foi lançada com base na Lei da Oportunidade de Crédito Igualitária de 1974 que impede empresas e serviços financeiros de discriminar pessoas com base em raça, cor, religião, nacionalidade, gênero, status marital ou idade no momento de conceder crédito, o órgão regulador não encontrou nenhum indício nas decisões de crédito do algoritmo que evidenciasse possíveis vieses discriminatórios, para chegar a essa conclusão foram analisados dados de mais de 400 mil pessoas que solicitaram o cartão.

Existe uma tendência a desconsiderar o gênero feminino na área financeira de

empréstimos bancários como capaz de fazer frente à determinados pagamentos e dívidas justamente porque há uma consolidação cultural de que a mulher não é provedora, não sendo atoa a necessidade de uma lei como essa em Nova York. Segundo Leo Kelion, editor de tecnologia da BBC, situação em que a discriminação social seria embutida na própria automação, se levado em conta séries históricas sem uma compensação dos dados e revisão humana do processo de *machine learning* é possível que ocorra essa discriminação de modo vedado, uma vez que as mulheres há tão pouco tempo são independentes profissionais e financeiramente.

A questão da opacidade preocupa quando tratamos de vieses, segundo a professora Ana Frazão (JOTA, 2021) é necessário que seja dada inteligibilidade as decisões algorítmicas, se a abertura total do código não for uma opção, se não estaremos diante de um arbítrio que gera discriminações abusivas. Ou seja, é preciso mais transparência nas decisões algorítmicas afim de que sejam auditáveis, a transparência consiste no acesso ao processo decisório da máquina e na sua compreensão.

3. A regulação ética algorítmica em diversos setores ao redor do mundo:

A Comissão Europeia para Eficácia da Justiça (CEPEJ) aprovou em dezembro de 2018 a Carta Europeia de Ética sobre o Uso da Inteligência Artificial em Sistemas Judiciais e seu ambiente (COE, 2018) estruturada em cinco princípios: (1) princípio do respeito dos direitos fundamentais: garantir que a concepção e a implementação de instrumentos e serviços de inteligência artificial sejam compatíveis com os direitos fundamentais; (2) princípio da não discriminação: prevenir especificamente o desenvolvimento ou a intensificação de qualquer discriminação entre indivíduos ou grupos de indivíduos; (3) princípio da qualidade e da segurança: no que diz respeito ao tratamento de decisões e dados judiciais, utilizar fontes certificadas e dados intangíveis com modelos concebidos de forma multidisciplinar, em ambiente tecnológico seguro; (4) princípio da transparência, imparcialidade e equidade: tomar os métodos de tratamento de dados acessíveis e compreensíveis, autorizar auditorias externas e (5) princípio “sob controle do usuário”: impedir uma abordagem prescritiva e garantir que os usuários sejam atores informados e controlem suas escolhas.

É preciso dar maior destaque ao princípio da não discriminação, segundo a carta, os intervenientes públicos e privados devem garantir que os métodos não reproduzem ou agravam essa discriminação e que não conduzam a análises ou utilizações determinísticas, dada a capacidade destes métodos de tratamento para revelar a discriminação existente, através do agrupamento ou da classificação de dados relativos a indivíduos ou grupos de indivíduos.

Para isso, observando com atenção a fase de desenvolvimento e de implantação, especialmente quando o tratamento se baseia, direta ou indiretamente, em dados sensíveis, que pode ser etnia, dados genéticos e biométricos, antecedentes socioeconômicos, opiniões políticas, convicções religiosas ou filosóficas, dados relativos à saúde ou à orientação sexual. Quando identificada a discriminação, devem ser consideradas medidas corretivas para limitar ou, se possível, neutralizar esses riscos, bem como a sensibilização das partes interessadas. No entanto, a utilização da aprendizagem automática e de análises científicas multidisciplinares para combater esta discriminação deve ser incentivada.

Quanto ao princípio da transparência, imparcialidade e equidade deve ser encontrado um equilíbrio entre a propriedade intelectual de certos métodos de tratamento e a necessidade de transparência, ou seja, de acesso ao processo de elaboração, de imparcialidade que é ausência de parcialidade e de equidade e integridade intelectual que significa ter como prioridade os interesses da justiça quando são utilizados instrumentos que podem ter consequências jurídicas ou afetar significativamente a vida das pessoas. Deve ficar claro que estas medidas se aplicam a toda a cadeia de elaboração e funcionamento, uma vez que o processo de seleção e a qualidade e organização dos dados influenciam diretamente a fase de aprendizagem.

Para alcançar esse equilíbrio, segundo Carta Europeia de Ética sobre o Uso da Inteligência Artificial em Sistemas Judiciais e seu ambiente (COE, 2018), a primeira opção é a transparência técnica total, por exemplo, código-fonte aberto e documentação, por mais que por vezes seja limitada pela proteção dos segredos comerciais. Ou o sistema pode ser explicado em linguagem clara e familiar para descrever como os resultados são produzidos comunicando, por exemplo, a natureza dos serviços oferecidos, as ferramentas que foram desenvolvidas, o desempenho e os riscos de erro. Além das autoridades ou peritos independentes poderem ser encarregados de certificar e auditar métodos de processamento ou de prestar aconselhamento prévio. As autoridades públicas poderiam conceder a certificação, que seria regularmente revista.

Atualmente se observa essa preocupação em elaborações éticas que se adequem ao uso e desenvolvimento de sistemas inteligentes ao redor do mundo e em diversos setores, as organizações profissionais estão particularmente preocupadas com o desenvolvimento de sistemas éticos. A *Association for Computing Machinery* (Associação para Maquinaria da Computação - ACM)¹⁰ apresentou recentemente um

¹⁰ A Association for Computing Machinery foi fundada em 1947 como a primeira sociedade científica e educacional dedicada à computação e conta com mais de 100 mil membros em todo o mundo atualmente, sendo sediada em Nova Iorque.

documento com o Código de Ética e conduta profissional atualizado (ACM, 2018). O código da ACM identifica os princípios gerais de ética orientadores de uma conduta profissional, organizando-os em quatro grupos, nomeadamente: princípios gerais de ética, responsabilidades profissionais, princípios de profissionais de liderança e conformidade com o código.

Quanto aos princípios gerais de ética, a primeira área de atuação, um dos princípios de partida nesta área de atuação é consciência que os profissionais da área de Sistemas de Informação e Tecnologias da Informação devem estar contribuindo para o bem-estar social, uma vez que todos são partes intervenientes e interessadas nos sistemas inteligentes. Nesse sentido, os profissionais devem desenvolver os sistemas com a preocupação de melhorar a qualidade de vida das pessoas, atuando para o benefício da sociedade e evitando danos e impactos negativos, especialmente aqueles que sejam significantes e desajustados, como ameaças físicas e mentais, e a destruição não justificadas de informação.

A base da relação organizacional é construída nos pilares de honestidade e confiança, portanto, o profissional deve ser transparente e fornecer todas as informações relativas às capacidades, limitações e potenciais problemas de um sistema, assim como ser honesto sobre as suas qualificações e limitações para a concretização de uma determinada tarefa. Deverá também ser justo e não tomar ações de discriminação, defender e aplicar valores de igualdade, tolerância, respeito pelos outros.

Um profissional deverá ser orientado e guiado por um conjunto de princípios que definam as suas responsabilidades profissionais, sendo essa a segunda área de atuação e que é composta pelos seguintes princípios: buscar elevados padrões de qualidade, a competência profissional começa com os conhecimentos técnicos, e consciente do contexto social onde o seu trabalho vai ser implementado.

Para além de *soft skills* necessárias o profissional deverá ser capaz de reconhecer e identificar os desafios éticos associados ao seu trabalho; conhecer e respeitar as regras existentes pertinentes para o trabalho profissional, aceitar e providenciar uma revisão profissional do trabalho a realizar ou realizado e realizar apenas tarefas da sua competência; promover a conscientização pública e compreensão dos sistemas inteligentes relativas à tecnologia e às suas consequências. Cabe ao profissional da área de informática, dado estar numa posição de confiança e conhecimento fornecer objetivamente, avaliações e informações aos vários intervenientes no processo, devendo ter um maior cuidado quando identificar potenciais riscos associados ao *machine learning* devendo mitigá-los.

A terceira área de atuação identificada é a Liderança Profissional e tem como ponto de partida o bem comum da sociedade. O líder ou gestor deverá também ser um

articulador, um elemento ativo e incentivar a equipe que lidera à agir com base na responsabilidade social e desencorajar tendências que as ponham em causa. Nesse sentido deve encorajar atitudes orientadas à transparência, qualidade e bem-estar da sociedade. Por último, cabe ao líder articular, aplicar e basear-se em políticas e processos que reflitam a aplicação do código.

A última área de atuação identificada é a Conformidade com o Código de Ética. Os sistemas de informação no futuro dependem essencialmente do componente técnico da ética, na sua elaboração, desenvolvimento e utilização. Relativamente aos profissionais que desenham e desenvolvem esses sistemas devem fazê-lo com base nos princípios do código contribuindo sempre para o melhor, devendo agir para resolver questões éticas quando identificadas.

Além da existência do código e de princípios que norteiem cada área de atuação do profissional da computação é necessário ter um instrumento que permita aferir em que medida este código está sendo aplicado e se a conduta profissional está sendo cumprida. Para isso foi criado o *CARE: Consider, Analyze, Review, Evaluate*, um processo que permite avaliar a conformidade, identificar e estabelecer ações a serem tomadas em consideração que estejam alinhadas com o código de conduta profissional da ACM.

Segundo a ACM (2018), o Código foi elaborado para inspirar e orientar a conduta ética de todos os profissionais de computação, incluindo profissionais atuais e aspirantes, instrutores, alunos, influenciadores e qualquer pessoa que use a tecnologia de computação de forma impactante. O Código como um todo se preocupa com a forma como os princípios éticos fundamentais se aplicam à conduta de um profissional de computação, ele não é um algoritmo para resolver problemas éticos, mas serve como base para a tomada de decisão ética. Ao pensar sobre um problema específico, um profissional de computação pode descobrir que vários princípios devem ser levados em consideração e que princípios diferentes terão relevância adequada ao problema, entendendo que o bem da sociedade é a consideração primordial. Toda a sociedade se beneficia quando o processo de tomada de decisão ética é responsável e transparente para todas as partes interessadas, com base nisso discussões abertas sobre questões éticas promovem essa responsabilidade e transparência.

O Instituto de Engenheiros Elétricos e Eletrônicos¹¹ também tem seu Código

¹¹ O IEE é a maior organização profissional técnica do mundo dedicada ao avanço da tecnologia em benefício da humanidade. Fundado em 1884 nos Estados Unidos é responsável pela definição de padrões mundiais para dispositivos elétrico e eletrônicos. Dirigida por um conselho de diretores e um comitê executivo, compõe-se de dez regiões, 36 sociedades técnicas, 4 conselhos técnicos, cerca de 1.200 divisões e 300 seções. Reúne mais de 300 mil associados, entre engenheiros, cientistas, pesquisadores e demais profissionais, em 150 países.

de Ética (IEEE, 2020) baseado em dez princípios gerais; (1) a primazia pela segurança, saúde e bem-estados indivíduos, visando o desenvolvimento de práticas de design ético e sustentabilidade; (2) devem ser evitados os conflitos de interesse sempre que possível; (3) honestidade e realismo do trabalho a ser desenvolvido pelos profissionais; (4) rejeição de suborno; (5) importa-se aumentar a literacia digital e tecnológica da sociedade em geral; (6) necessidade de atualização e desenvolvimento de conhecimento e competências nas áreas tecnológicas; (7) aceitar e propor críticas honestas dos outros, bem como a de dar crédito aos contributos dos outros; (8) tratamento cordial e a não discriminação baseada na raça, religião, género, deficiência, idade, nacionalidade, orientação sexual, identidade de género, ou expressão de género; (9) evitar prejudicar os outros, a sua propriedade ou reputação; (10) interajuda entre colegas no cumprimento do código de ética da IEEE.

O IEE lançou a discussão pública um documento com um conjunto de linhas diretoras sobre o alinhamento práticas de design ético e de sustentabilidade dos sistemas inteligentes, com os seguintes princípios bases: incorporar os mais altos ideais de valorização e benefício humano como um superconjunto dos direitos humanos; priorizar os benefícios para a humanidade e o meio ambiente do uso de sistemas inteligentes/autónomos. Note-se que, estes não devem estar em substituição um do outro, um depende do outro. Priorizar o ser humano e o bem-estar não significa degradar o meio ambiente; mitigar os riscos e os impactos negativos, incluindo o uso indevido, à medida que os sistemas inteligentes evoluem como sistemas sociotécnicos.

Em particular, assegurando que os sistemas inteligentes sejam responsáveis e transparentes. Foram definidos também cinco princípios gerais orientadores na conceção, desenvolvimento, implementação e monitorização destes sistemas inteligentes: (1) salvaguarda dos direitos humanos pelos sistemas; (2) prioridade ao bem-estar com resultado de um sistema desenhado de acordo com as regras de referência; (3) responsabilidade e corresponsabilidade (*accountability & responsibility*) bem definidas, na parte de legislação, no desenho e no desenvolvimento dos sistemas, na definição de ecossistemas de *stakeholders* múltiplos e em sistemas de registo que permitam identificar a evolução do desenvolvimento dos sistemas; (4) transparência para os utilizadores perceberem o que os programas estão a fazer e porquê. A transparência é também importante para a certificação dos sistemas, bem como para verificação dos sistemas inteligentes em caso de acidente e em caso de deliberação e atribuição de responsabilidades legais; (5) consciência nos usos dos sistemas autónomos e inteligentes, nomeadamente a consciência de que estes sistemas podem ser mal

utilizados com fins para os quais não foram desenvolvidos, sublinhado a necessidade de serem ministradas ações de sensibilização sobre a ética e o uso desses sistemas.

4. Apontamentos preliminares sobre a regulamentação da Inteligência Artificial no Brasil e os padrões éticos estabelecidos

O presente trabalho representa um *work in progress* do marco regulatório ético das IA no Brasil, por este motivo, não pretende oferecer uma análise conclusiva sobre o tema. Inobstante, à guisa de observações preliminares se faz necessário delinear o cenário regulatório atual bem como realizar algumas observações sobre o estado da arte.

O Brasil aderiu em 2019 na condição de país não-membro à Recomendação sobre Inteligência Artificial (IA) da Organização para Cooperação e Desenvolvimento Econômico (OCDE, 2019), a recomendação visa fomentar a inovação e a confiança na IA, promovendo a gestão responsável da IA confiável, garantindo o respeito pelos direitos humanos e valores democráticos. A Recomendação identifica cinco princípios-chave para a preservação responsável de IA sendo eles, o crescimento inclusivo, o desenvolvimento sustentável e bem-estar; valores centrados no ser humano e equidade; transparência e explicabilidade; robustez, segurança e proteção; e prestação de contas/responsabilização.

Essa recomendação observa a transparência como uma necessidade a ser observada no desenvolvimento de mecanismos de IA para que se promova uma compreensão geral dos sistemas e para permitir que aqueles afetados por estes sistemas possam compreender o resultado e desafiá-lo com base em informações simples e de fácil compreensão de acordo com os fatores que serviram de base para a previsão, recomendação ou decisão.

Desde o ano de 2021 já é possível observar reflexos da Recomendação sobre Inteligência Artificial da OCDE no Brasil, por meio da aprovação pelo Governo Federal da Estratégia Brasileira de Inteligência Artificial (EBIA, 2021) que assume o papel de nortear as ações do Estado brasileiro em prol do desenvolvimento das ações, em suas várias vertentes, que estimulem a pesquisa, inovação e desenvolvimento de soluções em Inteligência Artificial, bem como, seu uso consciente, ético e em prol de um futuro melhor. A fim de uma gestão responsável dos sistemas de IA, a EBIA traz as diretrizes traçadas pela OCDE e asseguradas agora pelo Brasil.

A EBIA tem como objetivos a contribuição para a elaboração de princípios éticos para o desenvolvimento e uso de IA responsáveis, a promoção de investimentos sustentados em pesquisa e desenvolvimento em IA, a remoção de barreiras à inovação em IA, a capacitação e formação de profissionais para o ecossistema da IA, o estímulo

a inovação e o desenvolvimento da IA brasileira em ambiente internacional e a promoção de um ambiente cooperativo entre os entes públicos e privados, a indústria e os centros de pesquisas para o desenvolvimento da Inteligência Artificial.

Para tanto, a Estratégia estabelece nove eixos temáticos, caracterizados como os pilares do documento; apresenta um diagnóstico da situação atual da IA no mundo e no Brasil; destaca os desafios a serem enfrentados; oferece uma visão de futuro; e apresenta um conjunto de ações estratégicas que nos aproximam dessa visão.

A Resolução nº 332 de 2020 do Conselho Nacional de Justiça dispõe sobre a ética, a transparência e a governança na produção e no uso da Inteligência Artificial no Poder Judiciário e dá outras providências. Destaco os dois capítulos que mais se relacionam com o tema aqui tratado, o capítulo III da não discriminação e o IV da publicidade e transparência.

Quanto à não discriminação é observado que as decisões judiciais apoiadas em sistemas de IA devem preservar a igualdade, a não discriminação, a pluralidade e a solidariedade, auxiliando no julgamento justo, com criação de condições que visem eliminar ou minimizar a opressão, a marginalização do ser humano e os erros de julgamento decorrentes de preconceitos. Em seus parágrafos traz acerca da necessidade de homologação do sistema a fim de identificar preconceitos ou generalização que possam ter influenciado o desenvolvimento do sistema, se verificado o óbvio discriminatório deve ser corrigido e se for impossível a correção deverá ser retirada de utilização.

Quanto a publicidade e transparência traz em que consiste para a resolução em incisos, sendo a divulgação responsável, considerando a sensibilidade própria dos dados judiciais, a indicação dos objetivos e resultados pretendidos pelo uso do modelo de Inteligência Artificial; a documentação dos riscos identificados e indicação dos instrumentos de segurança da informação e controle para seu enfrentamento; a possibilidade de identificação do motivo em caso de dano causado pela ferramenta de Inteligência Artificial; apresentação dos mecanismos de auditoria e certificação de boas práticas; fornecimento de explicação satisfatória e passível de auditoria por autoridade humana quanto a qualquer proposta de decisão apresentada pelo modelo de Inteligência Artificial, especialmente quando essa for de natureza judicial.

A questão é que hoje não há uma preocupação das limitações éticas durante todo o processo de criação e desenvolvimento de um sistema de inteligência artificial, deve se preocupar desde a fase inicial com a ética, havendo uma correção de vieses que o processo de *machine learning* pode resultar. Entretanto, é importante deixar claro que, por mais que avancemos em diretrizes éticas para o uso e desenvolvimento de

mecanismos de inteligência artificial, dificilmente um julgamento algoritmo pode ser perfeito e independente de supervisão humana.

O Poder Legislativo tem avançado nas discussões sobre a elaboração de um marco teórico para inteligência artificial no Brasil. No ano de 2022 foi instaurada no Senado Federal uma Comissão Temporária Interna de Juristas que para subsidiar a elaboração da minuta substitutiva aos Projetos de Lei 5.051/2019, 21/2020 e 872/2021, que visam determinar princípios, regras, diretrizes e fundamentos para regular o desenvolvimento e a aplicação da Inteligência Artificial (IA) no País, o presidente é o Ministro do Superior Tribunal de Justiça, Ricardo Villas Bôas Cueva, o anteprojeto foi apresentado no dia 6 de dezembro pela Comissão e transformado em projeto de lei recentemente no dia 3 de maio de 2023. Os projetos bases se preocupam com algumas questões interessantes para fins de uma regulamentação ética.

O PL no 5051 de 2019 é o segundo de iniciativa do Senador Styvenson Valentim do PODEMOS/RN e estabelece os princípios para o uso da Inteligência Artificial no Brasil, deforma a regulamentar seu uso, traz em seus fundamentos a transparência, a confiabilidade e a possibilidade de auditoria dos sistemas. É um projeto mais simples se comparado ao PL no 872 de 2021 de autoria do Senador Veneziano Vital do Rêgo do MDB/PB que dispõe sobre o uso da inteligência artificial observando os marcos éticos e diretrizes que fundamentam o desenvolvimento e o uso da inteligência artificial no Brasil.

O texto original do PL 872/2021 é mais generalista baseado nas diretrizes internacionais como as recomendações da OCDE e a Declaração de Montreal, buscando principalmente internalizar os marcos éticos que fundamentam o uso da inteligência artificial, foram apresentadas 17 emendas modificativas que ampliam consideravelmente o escopo da proposta original.

O artigo 4º do PL 872 de 2021 elenca as diretrizes necessárias para minimizar os riscos e preocupações que a tecnologia pode trazer, com isso o Senador Rogério Carvalho propôs uma emenda que desse a seguinte redação ao inciso VI: “prover decisões rastreáveis, que promovam a transparência e explicabilidade do modelo utilizado e sem viés discriminatório ou preconceituoso”, ou seja, a transparência afim de evitar o viés algorítmico como ressaltamos na presente pesquisa.

O Senador ressalta em sua justificativa que há uma importância da transparência por possibilitar ao interessado conhecer das entradas e saídas do modelo, base de dados do treinamento de *machine learning* e o tipo de algoritmo utilizado, quanto à explicabilidade, devem ser fornecidas informações de quais variáveis pesaram na tomada de decisões, relatórios de impacto algorítmico para analisar se o algoritmo é eficaz ao seu propósito ou se ocasiona impactos negativos a algum direito fundamental.

O Senador Styvenson Valentim propôs a inclusão de um artigo que limitasse o

uso dos sistemas decisórios baseados em inteligência artificial a auxiliar a tomada de decisão humana, devendo haver sempre uma forma de supervisão humana de acordo com o tipo, gravidade e implicação da decisão submetida a IA e traz a responsabilidade civil do supervisor da IA pelos danos decorrentes da utilização dos sistemas.

A propositura desse artigo que limita o uso dos sistemas de inteligência artificial pode ser aplicada ao uso da IA no Poder Judiciário, impossibilitando decisões autônomas sem qualquer supervisão humana, por haver situações não previstas pelo algorítmico. Entretanto, é necessário adotar uma abordagem baseada no grau de risco de cada sistema, promovendo a criação de padrões e códigos de conduta, há casos em que a supervisão humana possa vir a ser considerada necessária, mas há outros em que impossibilitaria o funcionamento dos sistemas.

O PL no 21 de 2020 estabelece fundamentos, princípios e diretrizes para o desenvolvimento e aplicação da inteligência artificial no Brasil; e dá outras providências. O projeto prevê os fundamentos no artigo 4º entre eles o respeito à ética, aos direitos humanos e aos valores democráticos, à igualdade, à pluralidade, à não-discriminação, o respeito aos direitos e garantias fundamentais do cidadão, à segurança, à privacidade e a proteção de dados pessoais. No artigo 5º do projeto estão os princípios, dois são relevantes para o presente estudo: a busca pela neutralidade e a transparência.

Quanto à busca pela neutralidade, o projeto elucida que “é recomendável que os agentes que atuam na cadeia de desenvolvimento e operação de sistemas de inteligência artificial busquem identificar e mitigar vieses contrários ao disposto na legislação vigente”. É uma importante contribuição do projeto de lei, entretanto, ainda possui um conteúdo muito genérico, uma vez que não elucida os meios pelos quais os agentes irão atuar para identificar e mitigar esses vieses algorítmicos.

O princípio da transparência encontra-se no inciso V, destacando de maneira detalhada a importância da transparência no uso de IA, a partir da divulgação do seu modo de funcionamento de acordo com cada caso retratado, entretanto, importa-se ressaltar os limites de divulgação para assegurar o segredo comercial e outras hipóteses de uso de sistemas não retratados ou como divulgar os critérios em caso de inteligência artificial que utilize o método de *machine learning* que apresente opacidade.

No dia 6 de dezembro de 2022 foi realizada a entrega do relatório final da CJSUSBIA (BRASIL, 2022) que conta com 908 páginas entre elas está a exposição de motivos e a minuta do Substitutivo, além da pesquisa completa feita pelos especialistas durante a elaboração do anteprojeto de regulação com as debates em audiências públicas, no seminário internacional, as contribuições escritas e os estudos das autoridades regulatórias de inteligência artificial nos países integrantes da OCDE.

A minuta substitutiva visa estabelecer normas gerais de caráter nacional para o desenvolvimento, implementação e uso responsável de sistemas de inteligência artificial (IA) no Brasil, com o objetivo de proteger os direitos fundamentais e garantir a implementação de sistemas seguros e confiáveis, em benefício da pessoa humana, do regime democrático e do desenvolvimento científico e tecnológico.

O anteprojeto tem 45 artigos e aborda os fundamentos da lei, os princípios, traz importantes definições, além de seis direitos para as pessoas afetadas por sistemas de IA, como o direito à informação prévia; à explicabilidade da decisão, recomendação ou previsão tomada por IA; à contestar essas decisões ou previsões que impactem de maneira significativa interesses do afetado; à determinação e participação humana nas decisões de sistemas inteligentes; à não discriminação e à correção de vieses discriminatórios diretos, indiretos, ilegais ou abusivos; e à privacidade e à proteção de dados pessoais. Deixando claro ainda que os agentes de IA deverão informar de forma clara e inteligível como exercer esses direitos e explicando de maneira detalhada cada direito nas seções seguintes.

Na exposição de motivos a comissão parte do objetivo normativo de conciliar uma abordagem baseada em riscos com uma modelagem regulatória baseada em direitos, o que foi claramente alcançado na minuta. Procurando, ainda, prever instrumentos de governança para prestação de contas e premiação dos agentes econômicos que agirem de boa-fé no gerenciamento dos riscos na elaboração e implementação desses sistemas. Também há uma categorização dos riscos no capítulo III em risco excessivo e alto risco, no seguinte são apresentadas medidas de governança para sistemas de IA de maneira geral e para sistemas de alto risco de maneira diferente dos demais, além da avaliação de impacto algorítmico.

A minuta também aborda o regime de responsabilidade civil obrigando o fornecedor ou operador de sistemas de IA que cause dano patrimonial, moral, individual ou coletivo a repará-lo integralmente, independente do grau de autonomia do sistema. Sendo o sistema de alto risco ou de risco excessivo, o fornecedor ou operador responderão objetivamente pelos danos causados, na medida da sua participação no dano, quando diferente for, os sistemas não forem de alto risco a culpa do agente causador do dano será presumida, aplicando-se a inversão do ônus da prova em favor da vítima. O anteprojeto também aborda excludentes de responsabilidade do agente de IA.

No dia 3 de maio o texto foi apresentado formalmente ao Senado Federal, sendo agora o Projeto de Lei nº 2338 para votação. Enquanto isso foram apresentados, na Câmara dos Deputados dois novos projetos de lei, são eles, o PL nº 753/2023 que regulamenta os sistemas de Inteligência Artificial, e dá outras providências e o PL nº 1153/2023 que dispõe sobre normas gerais para a pesquisa, o desenvolvimento e a aplicação da inteligência artificial – IA, e seu uso consciente e ético no âmbito da União,

dos Estados, do Distrito Federal e dos Municípios.

Diante de todo o exposto com os casos analisados nesta pesquisa resta ponderar acerca dos riscos da inexistência ou da demora da elaboração de um marco ético da temática. Os direitos fundamentais dos cidadãos estão sendo expostos a novos riscos até pouco tempo não experimentados que surgem com o uso de sistemas inteligentes em tomadas de decisões privadas e públicas. Isso ocorre, pois, a não regulamentação permite uma obscuridade de como eles têm sido desenvolvidos e utilizados na prática mundo a fora. Pouco se fala a respeito do limite de uso destes sistemas, dos dados utilizados no aprendizado de máquina e da preocupação de cada empresa quanto à transparência, à não discriminação e a acurácia dos mesmos, por exemplo.

Nesse momento em que nos encontramos, onde não há uma regulamentação brasileira sobre o uso e desenvolvimento de sistemas de inteligência artificial, muito embora o Senado Federal esteja trabalhando arduamente para muito em breve termos, a EBIA e a resolução nº 332/2020 e tantas outras elaborados pelo CNJ nos últimos anos estão norteando os primórdios dessa utilização e devem ser respeitadas e bem utilizadas.

5. Referências:

ACM. ACM Code of Ethics and Professional Conduct. ACM, 2018. Disponível em: <https://www.acm.org/code-of-ethics>. Acesso em 12 de setembro de 2022.

A IA para busca de candidatos a vagas do LinkedIn era tendenciosa. A Solução da empresa? Mais IA. MIT Technology Review, 2021. Disponível em: <https://mittechreview.com.br/a-ia-para-busca-de-candidatos-a-vagas-do-linkedin-era-tendenciosa-a-solucao-da-empresa-mais-ia/> Acesso em 14 de junho de 2022.

BRASIL. CÂMARA DOS DEPUTADOS. Projeto de Lei nº 21 de 2020. Estabelece princípios, direitos e deveres para o uso de inteligência artificial no Brasil, e dá outras providências. Brasília, 2020. Disponível em: <https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=2236340>. Acesso em 25 de junho de 2022.

BRASIL. CÂMARA DOS DEPUTADOS. Projeto de Lei nº 1153 de 2023. Dispõe sobre normas gerais para a pesquisa, o desenvolvimento e a aplicação da inteligência artificial. Brasília, 2023. Disponível em: <https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=2351386> Acesso em 20 de março de 2023.

BRASIL. CÂMARA DOS DEPUTADOS. Projeto de Lei nº 753 de 2023. Regulamenta os sistemas de Inteligência Artificial, e dá outras providências. Brasília, 2023. Disponível em: <https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=2349654>. Acesso em 20 de março de 2023.

BRASIL. SENADO FEDERAL. Projeto de Lei nº 5.051 de 2019. Estabelece os princípios para o uso da Inteligência Artificial no Brasil. Brasília, 2019. Disponível em:

<https://www25.senado.leg.br/web/atividade/materias/-/materia/138790> Acesso em 1 de julho de 2022.

BRASIL. SENADO FEDERAL. Projeto de Lei nº 872 de 2021. Dispõe sobre o Uso da Inteligência Artificial. Brasília, 2021. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/147434>. Acesso em 30 de junho de 2022.

BRASIL. SENADO FEDERAL. Relatório Final da Comissão de Juristas responsável por subsidiar elaboração de substitutivo sobre Inteligência Artificial no Brasil. 2022. Disponível em: <https://www.stj.jus.br/sites/portalp/SiteAssets/documentos/noticias/Relatório%20final%20CJSUBIA.pdf> Acesso em 20 de dezembro de 2022.

BRASIL. SENADO FEDERAL. Projeto de Lei nº 2338 de 2023. Dispõe sobre o Uso da Inteligência Artificial. Brasília, 2023. Disponível em: https://legis.senado.leg.br/sdleg-getter/documento?dm=9347622&ts=1691011102888&disposition=inline&_gl=1*yzoh4k*_ga*MTQ2MTczMTgyNy4xNjkwODMwMjAx*_ga_CW3ZH25XMK*MTY5MTY5MDQ4OS4yLjEuMTY5MTY5MDY3MC4wLjAuMA. Acesso em 30 de maio de 2023.

CEPEJ. The CEPEJ European Ethical Charter on the use of artificial intelligence (AI) in judicial systems and their environment. COE, 2018. Disponível em: <https://www.coe.int/en/web/cepej/cepej-european-ethical-charter-on-the-use-of-artificial-intelligence-ai-in-judicial-systems-and-their-environment>. Acesso em: 2 junho 2020.

DIREITO DIGITAL: Discriminação algorítmica de gênero. [Locução de]: Ana Frazão e Caitlin Mulholland. Brasil: Spotify, 27 de setembro de 2021. Acesso em 13 de junho de 2022.

DUARTE, ALAN. A antidiscriminação no contexto da inteligência artificial: possibilidades de governança mediante a normatização de algoritmos. 1ª edição. Fortaleza: Editora Mucuripe, 2021.

GOV.BR, Ministério da Ciência, Tecnologia e Inovações. Estratégia Brasileira de Inteligência Artificial. EBIA, 2021. Disponível em: https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/arquivosinteligenciaartificial/ebia-diagramacao_4-979_2021.pdf. Acesso em 6 de novembro de 2021.

IEEE. IEEE Code of Ethics, 2020. Disponível em: <https://www.ieee.org/about/corporate/governance/p7-8.html>. Acesso em 16 de setembro de 2022.

FERRARI, Isabela; BECKER, Daniel; WOLKART, Erik Navarro. Arbitrium exmachina: panorama, riscos e a necessidade de regulação das decisões informadas por algoritmos. Revista dos Tribunais, v. 995, 2018. Disponível em: <http://governance40.com/wp-content/uploads/2018/11/ARBITRIUM-EX-MACHINA-PANORAMA-RISCOS-E-A-NECESSIDADE.pdf>. Acesso em 10 de maio de 2022.

FRAZÃO, ANA. Discriminação algorítmica: Compreendendo o que são os julgamentos algorítmicos e os eu alcance na atualidade. JOTA, 2021. Disponível em: <https://www.jota.info/opiniao-e-analise/colunas/constituicao-empresa-e-mercado/discriminacao-algoritmica-16062021> Acesso em 15 de junho de 2022.

KAUFMAN, Dora. Inteligência artificial: questões éticas a serem enfrentadas. IX Simpósio Nacional ABCiber, PUC São Paulo, 2016. Disponível em: <https://abciber.org.br/anaisletronicos/wp->

[content/uploads/2016/trabalhos/inteligencia_artificial_questoes_eticas_a_serem_enfrentadas_dora_kaufman.pdf](#). Acesso em 14 de set de 2022.

ODERMEYER, Ziad; POWERS, Brian; VOGELI, Christine; MULLAINATHAN, Sendhil. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, vol 336, issue 6464, pp. 447-453, 2019. Disponível em <https://www.science.org/doi/10.1126/science.aax2342> Acesso em 8 de junho de 2022.

OECD. Recommendation of the Council on Artificial Intelligence. OECD, 2019. Disponível em: < <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL0449#:~:text=The%20OECD%20Council%20adopted%20the,on%2022%2D2%20May%202019.&text=The%20OECD%20Recommendation%20on%20AI,governments%20in%20their%20implementation%20efforts> > Acesso em 18 de abril de 2021.

O'NEIL CATHY. Algoritmos de destruição em massa: Como o big data aumenta a desigualdade e ameaça a democracia; tradução Rafael Abraham. Santo André, SP: Editora Rua do Sabão, 1ª edição, 2020. Título original: weapons of math destruction: howbig data increases inequality and threatens democracy.

Por que o serviço de cartão de crédito da Apple está sendo acusado de sexismo. BBC, 2019. Disponível em: <https://www.bbc.com/portuguese/geral-50355369>. Acesso em 15 de setembro de 2022.