

### A DISCRIMINAÇÃO ALGORÍTMICA COMO PROBLEMA ÉTICO-JURÍDICO DA REGULAÇÃO DOS SISTEMAS DE INTELIGÊNCIA ARTIFICIAL

#### ALGORITHMIC DISCRIMINATION AS AN ETHICAL-JURIDICAL PROBLEM IN THE REGULATION OF ARTIFICIAL INTELLIGENCE SYSTEMS

*Luis Henrique de Menezes Acioly<sup>1</sup>*

*Ricardo Maurício Freire Soares<sup>2</sup>*

**Resumo:** O presente manuscrito parte da constatação de que os esforços imanentes à regulação da Inteligência Artificial se coadunam com a busca da efetividade, considerando-se para tanto as particularidades técnicas que envolvem o desenvolvimento dessa tecnologia. A par disso, pretendeu-se analisar as questões substancialmente relevantes para uma efetiva política regulatória em IA, como sua conceituação, abordagens de aprendizado e potenciais riscos a direitos fundamentais. Metodologicamente, empreendeu-se uma revisão sistemática de literatura. Obteve-se como resultado a conceituação de sistemas de inteligência artificial deve considerar as nuances dos seus atributos formativos, afastando-se outras formas de programação computacional que não sejam pautadas em aprendizado. Constatou-se que o desenvolvimento de sistemas de IA não é uniforme, não sendo aplicável uma estrutura regulatória estática ou unidirecional, sub pena de insuficiência técnica da regulação. Concluiu-se também que o caráter flexível dos princípios éticos contribui para formulação de estruturas regulatórias direcionadas à concretização de direitos fundamentais.

**Palavras-Chave:** Inteligência Artificial. Aprendizado de Máquina. Discriminação Algorítmica. Regulação. Ética na IA.

**Abstract:** This manuscript begins with the observation that inherent efforts in regulating Artificial Intelligence align with the pursuit of effectiveness, considering the technical intricacies involved in the development of this technology. In this context, the aim was to analyze substantially relevant issues for an effective regulatory policy in AI, such as its conceptualization, learning approaches, and potential risks to fundamental rights. Methodologically, a systematic literature review was conducted. The result obtained conceptualizes that artificial intelligence systems should consider the nuances of their formative attributes, distancing themselves from other forms

---

<sup>1</sup>Graduado em Direito pelo Centro Universitário Ruy Barbosa (UniRuy). Pesquisador junto ao Grupo de Estudos em Tecnologia, Informação e Sociedade da UNIFOR – GETIS/CNPq. Vice-Presidente do Laboratório de Inovação e Direitos Digitais da UFBA - LABID<sup>2</sup>. Pesquisador junto ao Grupo de Pesquisa "Conversas Civilísticas" da UFBA. Treinee junto ao Chezzi Advogados. Orcid: <https://orcid.org/0000-0002-1615-6048>. Currículo Lattes: <http://lattes.cnpq.br/0485009486625100>. Contato: [acioly10@gmail.com](mailto:acioly10@gmail.com).

<sup>2</sup> Pós-Doutor em Direito pela Università degli Studi di Roma La Sapienza, pela Università degli Studi di Roma Tor Vergata e pela Università del Salento. Doutor em Direito pela Università del Salento/Universidade de São Paulo. Doutor em Direito Público e Mestre em Direito Privado pela Universidade Federal da Bahia. Professor da Faculdade de Direito da Universidade Federal da Bahia (Graduação, Mestrado e Doutorado). Membro da Comissão Nacional de Defesa do Consumidor do Conselho Federal da Ordem dos Advogados do Brasil. Membro do Instituto dos Advogados Brasileiros e do Instituto dos Advogados da Bahia. Membro da Academia de Letras Jurídicas da Bahia. Membro do Instituto Geográfico e Histórico da Bahia. Orcid: <https://orcid.org/0000-0002-0806-8603>. Currículo Lattes: <https://lattes.cnpq.br/7597880442041621>. Contato: [ricardo.mfsoares01@gmail.com](mailto:ricardo.mfsoares01@gmail.com).

of computer programming not rooted in learning. It was noted that the development of AI systems is not uniform, rendering a static or unidirectional regulatory framework technically insufficient. Additionally, it was concluded that the flexible nature of ethical principles contributes to formulating regulatory structures aimed at realizing fundamental rights.

**Keywords:** Artificial Intelligence. Machine Learning. Algorithmic Discrimination. Regulation. Ethics in AI.

### 1 ASPECTOS INTRODUTÓRIOS

O desenvolvimento das tecnologias de informação e comunicação (TICs) tem levado a um estado de processamento e raciocínio de informações que superam o panorama socioeconômico e cultural de séculos anteriores, conduzindo aos sistemas de inteligência artificial (IA) e a sua correlata aplicação nas atividades centrais da vida social.

Por conta disso, verifica-se uma gama de atores interessados em instituir regras que disciplinem o desenvolvimento das pesquisas e aplicação da inteligência artificial. Tais regras possuem atualmente feições éticas e, em caráter excepcional, prescrições normativas estatais difusas, com previsões de sanções administrativas e econômicas. A correlação de sistemas inteligentes aos diversos setores da sociedade engendra um panorama de difusão e sobreposição normativa, especialmente quanto a questões que envolvam a proteção do mercado e do consumidor ou relativas à administração pública.

Assim, diversos Estado-Nações movimentam seus aparatos normativos para despontarem como reguladores da Inteligência Artificial. No cenário europeu, o *AI Act* tem ocupado o centro das discussões, ainda que se trate de proposta de regulamento em trâmite, por representar o estado da arte do estudo do tema na comunidade europeia, com potencial de influenciar diversos Estados em suas regulações, através do “efeito Bruxelas” (Siegmann; Anderljung, 2022). Por outro lado, o Canadá endereça os seus esforços ao *Artificial Intelligence and Data Act* ou *AIDA*, que, conjuntamente com *Consumer Privacy Protection Act* e o *Personal Information and Data Protection Tribunal Act*, compõe o *Bill-C27*, uma robusta proposta para regulamentar do ambiente digital em seus diversos aspectos<sup>3</sup> a partir de uma abordagem baseada nos riscos inerentes ao fluxo informacional (Scassa, 2023).

Ante o exposto, cumpre, todavia, definir os parâmetros sobre os quais devem incidir o prisma regulatório a fim de suplantar eventuais iniciativas prescritivas rasas e evitar tornar inócua o esforço de estabelecer uma centralização normativa acerca do desenvolvimento de inteligência artificial. A compreensão das particularidades técnicas desse modelo de tecnologia e

---

<sup>3</sup> A normatização robusta de aspectos do ambiente digital e da Inteligência Artificial no Canadá é fruto de um histórico recente de atuação deste país. Segundo a UNESCO, o Canadá foi o primeiro país a lançar uma estratégia nacional de Inteligência Artificial, a “*Pan-Canadian Artificial Intelligence Strategy*”, em 2017 (Ministério das Relações Exteriores, 2022).

idiosincrasias de sua aplicação nos diversos setores da sociedade é um caminho necessário a se trilhar, tendo em vista estabelecer as bases de uma regulação ótima.

Buscando-se estruturar as questões materialmente relevantes para a regulação da inteligência artificial, o presente estudo se debruça na literatura sobre o tema para consignar o panorama de riscos associados ao aprendizado de máquina que deve subsidiar a adoção de práticas responsáveis e a prestação de contas, fazendo correlação com o horizonte de diretrizes éticas para o desenvolvimento desse campo tecnológico.

Para tanto, realizou-se, metodologicamente, a partir do exame crítico da literatura de referência no tema, a presente pesquisa bibliográfica e documental de cunho descritivo, natureza qualitativa e caráter exploratório, em que se preconizou o diálogo entre autores das diferentes áreas das ciências sociais.

Com efeito, empreendeu-se aqui uma revisão literária, materializada por meio de uma pesquisa bibliográfica, cuja coleta de dados ocorreu através de livros, artigos e trabalhos científicos, repositados em bases de dados eletrônicas – *Scientific Electronic Library Online* (SciELO), *Index Law Journals* e Google Acadêmico –, tendo como descritores: Inteligência Artificial; *Machine Learning*; Deep Learning; Discriminação Algorítmica; Ética.

## 2 PREMISSAS CONCEITUAIS SOBRE INTELIGÊNCIA ARTIFICIAL E PROGRAMAÇÃO ALGORÍTMICA

A definição do que seria um sistema de “inteligência artificial” revela-se objeto de discussão na ciência da computação e na ciência jurídica, que não encontram uma uniformidade (Bartnecket *al.*, 2021). Buscando atribuir características à definição de Inteligência Artificial, Russel e Norvig (2013) sintetizam as abordagens conceituais de: (i) pensamento como o ser humano; (ii) pensamento racional; (iii) ação como ser humano; e (iv) ação racional. As duas primeiras abordagens se relacionam a processos e raciocínio, enquanto as duas últimas buscam na experimentação a conceituação de Inteligência Artificial (Russell; Norvig, 2013). Por sua vez, a adoção de parâmetros em comparação com o ser humano se funda em questões filosóficas e antropocêntricas, ao passo que os parâmetros de racionalidade se baseiam na matemática, envolvendo conhecimento de engenharia da computação e sistemas de informação (Lage, 2022; Bartnecket *al.*, 2021).

Noutro passo, Gabriel (2022) e Lee (2019) fazem referência às abordagens simbólica e conexionista para traçar parâmetros da conceituação de IA. A abordagem simbólica se funda no pensamento lógico, orientado pela matemática, em descrever de forma abstrata os processos que geram comportamento inteligente, enquanto a abordagem conexionista é inspirada na fisiologia do cérebro humano e em seu funcionamento baseado nas redes neurais (Gabriel, 2022). Essa segunda

## DIÁLOGOS POSSÍVEIS

ISSN 2447-9047  
VOLUME 23, Nº 2- JUL/DEZ 2024

Pág: 97 -121

abordagem se relaciona com a definição de IA trazida por Russel (2021), para quem a Inteligência Artificial constitui o estudo das formas como as estruturas computacionais se comportam de maneira inteligente, tal qual o *raciocínio* humano<sup>4</sup>. A vertente conexionista reconhece o processamento de informações de forma inteligente através de camadas de encadeamento lógico e de redes neurais (Gabriel, 2022), dotando o sistema de certa autonomia (Hoffmann-Riem, 2021).

Para o objetivo desta investigação, toma-se como conceito de inteligência artificial, a partir da complementariedade das abordagens simbólica e conexionista (Gabriel, 2022), o modelo automatizado de processamento de informações em camadas de encadeamento lógico, através de treinamento e aprendizagem, cuja função é resolver questões de forma semelhante a algum espectro de inteligência humana<sup>5</sup>. Filia-se à compreensão de Duarte (2021) e de Kaplan (2016), para quem, ainda que se trate de uma inteligência artificial, apenas atributos específicos da cognição humana são passíveis de reprodução por máquinas<sup>6</sup>. Tal especificação tem o condão de afastar do objeto do presente estudo outras formas de tecnologia que, embora representem a evolução da programação algorítmica (Acioly; Oliveira, 2022), não realizam um aprendizado constante com foco na mimetização de aspectos do raciocínio humano, tais quais o *blockchain* e os *smart contracts* (Fonseca, 2021). Como se observa do quanto exposto, para além da resolução de obstáculos de forma racional, é o aprendizado que define os contornos de sistemas de inteligência artificial<sup>7</sup>.

Como esforço inicial para estabelecer critérios de identificação dos sistemas inteligentes, o Teste de Turing tem suas premissas na viabilidade do aprendizado pela máquina, com objetivo de mimetizar o comportamento humano, tornando-se indistinguível (Bartnecket *al.*, 2021). Para

---

<sup>4</sup> Nesse contexto, Lee (2019) usa de um simples comando para retratar a diferença de funcionamento da inteligência artificial quanto às duas abordagens: identificar um gato em uma imagem. A abordagem simbolista, atrelada ao silogismo – *se-então* – raciocina que se identificar duas formas triangulares em cima de uma forma circular, então há um gato na imagem (Lee, 2019). Por sua vez, a abordagem conexionista pressupõe que a máquina seria alimentada com volume considerável, na casa dos milhões, com amostras de imagens rotuladas em “com gato” e “sem gato”, permitindo, assim, que o sistema crie correlações e estabeleça seus critérios para identificação do gato na imagem (Lee, 2019).

<sup>5</sup> A complementariedade das teorias conexionistas e simbólica é um ponto de referência do “AI Act” (União Europeia, 2021), que enumera as seguintes abordagens para definição de sistemas de inteligência artificial: “a) Abordagens de aprendizagem automática, incluindo aprendizagem supervisionada, não supervisionada e por reforço, utilizando uma grande variedade de métodos, designadamente aprendizagem profunda; b) Abordagens baseadas na lógica e no conhecimento, nomeadamente representação do conhecimento, programação (lógica) indutiva, bases de conhecimento, motores de inferência e de dedução, sistemas de raciocínio (simbólico) e sistemas periciais; c) Abordagens estatísticas, estimação de Bayes, métodos de pesquisa e otimização”.

<sup>6</sup> Para o autor (Duarte, 2021) obter um diagnóstico da inteligência de uma máquina com base apenas no Teste de Turing é superestimar a inteligência artificial e subestimar a humana, posto que a inteligência humana é um fenômeno complexo, não compreendido totalmente e que desperta em cada indivíduo de forma individualizada, para além da mera capacidade de resolução de problemas lógico-matemáticos (Kaplan, 2016).

<sup>7</sup> As questões do aprendizado e da autonomia são centrais em algumas conceituações normativas, como aquela realizada no Projeto de Lei brasileiro nº 2338, de 2023: “sistema de inteligência artificial: sistema computacional, com graus diferentes de autonomia, desenhado para inferir como atingir um dado conjunto de objetivos, utilizando abordagens baseadas em aprendizagem de máquina e/ou lógica e representação do conhecimento, por meio de dados de entrada provenientes de máquinas ou humanos, com o objetivo de produzir previsões, recomendações ou decisões que possam influenciar o ambiente virtual ou real”.

## DIÁLOGOS POSSÍVEIS

ISSN 2447-9047  
VOLUME 23, Nº 2- JUL/DEZ 2024

Pág: 97 -121

“passar” nesse teste, o sistema precisa demonstrar os seguintes elementos: (i) processamento de linguagem natural, para viabilidade da comunicação; (ii) representação de conhecimento, armazenando-o com o decorrer da interação; (iii) raciocínio automatizado para usar o conhecimento armazenado de forma lógica; e (iv) aprendizado de máquina para se adaptar a novas circunstâncias e fugir dos padrões de sua programação inicial, demonstrando a capacidade de aprender com o interlocutor<sup>8</sup> (Russel; Norvig, 2013). Este último elemento é o que efetivamente destaca a inteligência do sistema.

O aprendizado constante figura como elemento que destaca os sistemas de inteligência artificial de demais espécies de programas computacionais, ainda que constituídos por algoritmos (Kaplan, 2016; Hoffmann-Riem, 2021; Acioly; Oliveira, 2022). Toma-se como exemplo o sistema *Deep Blue* que, embora tenha tido a capacidade de vencer o campeão de xadrez Garry Kasparov em uma partida, não realizou um aprendizado das técnicas desse esporte, mas foi constituído apenas para aplicar silogismo nas jogadas do xadrez, não se subsumindo ao conceito de Inteligência Artificial (Duarte, 2021). A aprendizagem autônoma, com descolamento da previsão inicial é o que dá, proporcionalmente, sentido e alcance à inteligência artificial (Peixoto; Silva, 2019; Acioly, 2022; Kaufman, 2022), na medida em que, através de cálculos estatísticos e correlações, se possibilita realizar comparações, deferindo ao sistema a capacidade de modificar o próprio funcionamento a partir das experiências acumuladas (Reis, 2020; Kaufman, 2022).

Neste sentido, é possível instituir categorias de sistemas de IA, conforme a variedade de áreas do conhecimento para as quais foram destinados os esforços de aprendizagem. Russell e Norvig (2013) apresentam as categorias de IA Forte e IA fraca como corolários da discussão sobre a própria existência e conceituação da Inteligência Artificial, distinguindo-as em relação ao grau de independência que o sistema apresenta. Em vista à concretude dessa classificação, Gabriel (2022) propõe que o termo “IA Fraca” seja atribuído aos sistemas em que o foco se restringe a determinada área do conhecimento, o que seria também denominado de Inteligência Artificial Limitada (*Artificial Narrow Intelligence – ANI*). Por outro lado, consiste em “IA Forte”, ou Inteligência Artificial Geral (*Artificial General Intelligence – AGI*), o sistema que possui nível de habilidade capaz de compatibilizar diversas áreas de aprendizado, e consiste em Superinteligência (*Artificial Super Intelligence – ASI*), o sistema capaz de superar as habilidades humanas de processamento de informação (Gabriel, 2022). Essas duas últimas categorias de IA, contudo, resultam de estudo de previsão da próxima geração de sistemas inteligentes (Duarte, 2021; Blackman, 2022), estando a humanidade somente em contato com a IA Limitada ao tempo de formulação do presente estudo.

---

<sup>8</sup> Russel e Norvig (2013) também aludem um “Teste de Turing Total”, em que, além dos elementos assinalados é necessário que o sistema interaja fisicamente com o interlocutor através de objetos, devendo demonstrar visão computacional para perceber os artefatos e robótica para manipulá-los com sucesso.

Cumprir abordar, nesse escopo, que a aprendizagem de máquina não se confunde com os algoritmos, embora mantenham uma relação intrínseca (Peixoto; Silva, 2019; Acioly, 2022). Lage (2022) define algoritmo como o processo a ser seguido na solução de um dado problema, especialmente por um computador. Essa definição, embora se adeque à realidade da programação de *software*, não corresponde plenamente aos processos de aprendizado de máquina. Por sua vez, Rosa (2017) conceitua algoritmo como uma representação abstrata de um dado processo, que considera os elementos cientes do conhecimento para prever respostas. Aplicado a esse escopo, Mendes e Mattiuzzo (2019) e Barocaset *al.* (2014) definem algoritmo como um conjunto de instruções lógicas, organizadas de forma concatenada, que padronizam como algo deve ser feito. Encadeando esses entendimentos, pode-se compreender, para fins do presente estudo, que os algoritmos de um sistema inteligente cuidam de atribuir sentido à aprendizagem, instruindo o processo de abstração da informação a partir de estruturas que espelham o processo pedagógico humano.

Ademais, importa assinalar que a extensão da regulação de inteligência artificial deve considerar as diferenças conceituais que envolvem os atributos formativos desse sistema, especialmente quanto à diferença frente à mera programação algorítmica. O estabelecimento conceitual desses elementos tem o condão de circunscrever o ambiente regulatório aos sistemas de processamento automatizado de informações por camadas de encadeamento lógico e através de mecanismos de treinamentos e algoritmos de aprendizado.

### **3 APRENDIZADO DE MÁQUINA, REDES NEURAIIS ARTIFICIAIS E *BIG DATA ANALYTICS***

O permanente aprendizado por sistemas automatizados desperta questões específicas à programação algorítmica e processamento de informações para gerar resultados práticos. Aprendizado de máquina (*machinelearning*) diz respeito ao termo cunhado por Arthur Samuel, em 1959, para definir o campo de estudo da ciência da computação que confere à máquina a habilidade de aprender sem ser explicitamente programado (Samuel, 1959; Gabriel, 2022). O aprendizado de máquina constitui a forma como a árvore de decisões é estruturada em um dado sistema inteligente com vistas à formação de um resultado (Kaufman, 2022), constituindo critério para sua configuração como inteligência artificial (Gabriel, 2022).

No âmbito da inteligência artificial, os algoritmos desempenham três grupos de funções essenciais (Lage, 2022), a saber: (i) estruturam a sequência que a informação vai seguir no processo de aprendizagem; (ii) correlacionam através de silogismos e condicionantes um *input* a um resultado em uma dada etapa; e (iii) repetem o processo de aprendizagem para atribuir-lhe acurácia (Lage, 2022). Aprendizado não significa somente armazenar informações ou processá-

## DIÁLOGOS POSSÍVEIS

ISSN 2447-9047  
VOLUME 23, Nº 2- JUL/DEZ 2024

Pág: 97 -121

las, mas saber dispor delas de forma estruturada para lhes extrair o resultado de uma questão, construindo novas premissas de raciocínio (Kaplan, 2016). As formas de *machinelearning* podem variar de acordo com a estruturação dos algoritmos de aprendizagem, sendo geralmente categorizado em: aprendizado supervisionado, aprendizado não supervisionado e aprendizado por reforço (Russell; Norvig, 2013; Bartneck *et al.*, 2021; Gabriel, 2022).

Sendo assim, o aprendizado supervisionado se desenvolve a partir de métodos de regressão e classificação, casos em que são fornecidos, além do *input*, um padrão de resultado esperado (*target*), classificando a memória do sistema na idealização do resultado e aplicação desse padrão (Bartneck *et al.*, 2021). O aprendizado não supervisionado (*freelearning*) é desenvolvido quando não há resultado esperado para determinados *inputs*, focando no potencial de agrupamento de dados (*clusters*) para a formação de padrões (Gabriel, 2022), sendo um percussor exploratório para sistemas de aprendizado supervisionado (Bartneck *et al.*, 2021). Ressalta-se que não existem limites nítidos entre o aprendizado supervisionado e não supervisionado, podendo conduzir ao espaço cinzento do aprendizado “semisupervisionado” (Russell; Norvig, 2013), em que os dados de entrada são um conjunto de *inputs* com e sem *targets*, cabendo ao sistema simultaneamente a tarefa de elaborar os “*clusters*” e chegar a determinados padrões (Gabriel, 2022). A seu turno, o aprendizado por reforço (*reinforcementlearning*) se desenvolve a partir de interações – seja por recompensas ou punições – que estimulam e reforçam a aprendizagem em um sistema de “tentativa e erro”, de forma semelhante à psicologia humana e o behaviorismo<sup>9</sup> (Gabriel, 2022).

Um próximo estágio do aprendizado de máquina é o aprendizado profundo (*deep learning*), que se distingue por atribuir ao sistema a capacidade de estabelecer padrões a partir de próprias correlações (Pinto, 2020; Ribeiro, 2021), compreendendo relações, estruturas e arquiteturas sem intervenção humana adicional, de forma que as camadas de encadeamento lógico se entrelaçam dificultando a reconstrução dos caminhos da árvore de decisões que levaram ao resultado do sistema (Russell; Norvig, 2013; Hoffmann-Riem, 2021). O aprendizado profundo pode ser alcançado através de formas não lineares de aprendizado, por meio de uma árvore de decisões cuja sequência foi elaborada por um algoritmo criado pelo próprio sistema, semelhantemente ao que ocorre no cérebro humano (Pinto, 2020; Gabriel, 2022).

Com efeito, as redes neurais, ou neurônios artificiais, subsidiam o processo de *deep learning*, constituindo ferramentas de concretização de uma frente multicamada da inteligência artificial (Russell; Norvig, 2013). A estrutura em camadas de encadeamento lógico se torna

---

<sup>9</sup> O behaviorismo é uma escola de pensamento psicológico que se debruça em compreender as nuances do comportamento humano, suas causas e influências (Skinner, 2009). A teoria behaviorista vislumbra nos reforçadores positivos e negativos um caminho para o condicionamento operante do comportamento humano (Skinner, 2009).

## DIÁLOGOS POSSÍVEIS

ISSN 2447-9047  
VOLUME 23, Nº 2- JUL/DEZ 2024

Pág: 97 -121

imbricada o suficiente elaborar novas formas de interação dos “neurônios”, para além da programação original (Kaplan, 2016). Gabriel (2022) explicita que o modelo de redes neurais baseadas em matemática e algoritmos teve origem em 1943, por trabalho do neurocientista Warren McCulloch e do matemático Walter Pitts, que concentraram sua aplicação em processos biológicos no cérebro humano<sup>10</sup> e em computadores<sup>11</sup> (McCulloch; Pitts, 1943). Essa segunda aplicação das redes neurais artificiais se desenvolveu no âmbito da inteligência artificial, sendo bem sucedida em diversas atividades, inclusive no desenvolvimento de *neurocomputadores* (Gabriel, 2022). Kaplan (2016) sintetiza as redes neurais artificiais como a programação computacional fundada em princípios organizacionais atribuíveis a uma rede neural real.

Por sua vez, a aplicação das redes neurais artificiais e das demais formas de aprendizado de máquina são compatibilizadas com o processo de mineração de dados<sup>12</sup> – ou *Data Mining* –, tendo em vista adaptar a técnica a um resultado que potencialize a análise de estatísticas e reconhecimento de padrões com grande quantidade de dados (Russell; Norvig, 2013). A análise de grande quantidade de dados brutos, através da qual se extrai conhecimento vendável (Gabriel, 2022), tem viabilizado que empresas combinem dados primitivos a fim de criarem inferências e previsões (Mendes, 2014). O desenvolvimento de sistemas de IA voltados à análise de bancos de dados – *Big Data Analytics* – tem fornecido subsídios técnicos e comerciais para o estabelecimento de vários *Big Data*<sup>13</sup> ao redor do globo (Mayer-Schonberger; Cukier, 2013), potencializando uma economia baseada em dados (Bioni, 2021).

O fenômeno do *Big Data* acarreta profundas transformações para a economia informacional, na medida em que permite a parametrização de informações para sua análise em

---

<sup>10</sup> Kaplan (2016) explica que, ainda que se conheça sobre a estrutura grosseira do cérebro humano, como as regiões que estão envolvidas em uma dada atividade, se ignora a sua estrutura intermediária, como os neurônios se interconectam para realizar tarefas. A construção de redes neurais artificiais, nesse contexto, viabiliza a simulação de comportamento dos neurônios e suas interconexões para estudo dos seus resultados, auxiliando na compreensão do cérebro humano (Kaplan, 2016).

<sup>11</sup> O modelo de redes neurais artificiais, idealizados por McCulloch e Pitts (1943) era simples comparado às formas atuais, porém continua sendo visto como uma das principais referências em relação ao tema (Kovács, 2006). O neurônio “booleano” de McCulloch era um dispositivo *binário* que tinha diversas entradas e uma saída, cujo resultado poderia representar um pulso ou não pulso, calculando-se a soma ponderada das entradas com os fatores de ponderação positivos, nos casos excitatórios, e negativos, nos casos inibitórios (Kovács, 2006). Esse cálculo de predicados – ou fatores de ponderação – seria na visão de McCulloch e Pitts (1943) seria o conceito de inteligência, cumprindo à rede de neurônios o papel de implementar esses cálculos a partir de funções booleanas, restando, assim cientificamente comprovada a ligação entre inteligência e atividade nervosa (Kovács, 2006). O conceito de *booleano*, na ciência da computação, diz respeito ao dado primitivo que possui apenas dois valores possíveis: verdadeiro ou falso (Kovács, 2006).

<sup>12</sup> Duarte (2021) aponta que o processo de mineração de dados compõe uma das cinco etapas de um processo mais amplo, denominado Descoberta de Conhecimento em Banco de Dados (“*Knowledge Discovery in Databases*” – KDD), que visa suprir a limitação humana no tratamento massivo de informações.

<sup>13</sup> As características que identificam o fenômeno dos *Big Data* são, segundo Hoffmann-Riem (2021): (i) o alto volume, isto é o acesso a grandes quantidades de dados; (ii) a alta variedade de informações armazenadas; (iii) a alta velocidade do seu processamento; (iv) a garantia de sua veracidade, isto é qualidade dos dados armazenados e processados; e (v) a possibilidade de fundamentar novos modelos de negócios e gerar valor agregado.



tempo real, obtendo valor de uma massiva massa de dados, demandando por uma ampliação global da capacidade de armazenamento e processamento de dados (Gabriel, 2021).

De outro lado, a substituição da causalidade por correlações nos processos realizados por sistemas de inteligência artificial (Mendes; Mattiuzzo, 2019) torna a análise de massivos bancos de dados uma atividade econômica lucrativa, porém, carregadas de um risco social discriminatório (Barocas; Selbst, 2016), na medida em que fornece ideais de comportamento humano e cria previsões de condutas e preferências (Gillespie, 2018; Hoffmann-Riem, 2021). Os riscos sociais decorrentes do uso de inteligência artificial serão abordados em tópico próprio desse estudo, sendo, porém, necessário consignar aqui que o uso comercial de sistemas inteligentes para gerir *Big Data* tem se direcionado à tomada de decisões automatizadas, a partir de uma avaliação preditiva, denominados ciclos de antecipação, com uma gama de aplicações práticas (Gillespie, 2018).

Sobremais, convém registrar que o aprendizado de máquina, enquanto elemento distintivo dos sistemas de Inteligência Artificial, lhes defere um amplo espectro de utilizações, bem como descortina uma gama de discussões concernente ao processo cognitivo humano. Deve-se ter como pressuposto à regulação de Inteligência Artificial a consideração que o desenvolvimento de tais sistemas não é uniforme, mas detém particularidades quanto à sua estrutura e à sua aplicação, especialmente no contexto de *Big Data Analytics*, de forma que não se deve guardar uma correlação estática com o sistema normativo.

#### **4 OS RISCOS DE UTILIZAÇÃO DA INTELIGÊNCIA ARTIFICIAL: O DEBATE ÉTICO-JURÍDICO SOBRE A DISCRIMINAÇÃO ALGORÍTMICA**

O processo de mimetização do aprendizado humano não conduz linear e necessariamente a resultados eticamente válidos. Os modelos de aprendizado e as redes neurais podem, através de camadas de raciocínio lógico, produzir inferências e previsões pessoais que não condigam com a realidade (Barocas; Selbst, 2016; Gillespie, 2018), ou, ainda que se coadunem com um resultado estatisticamente correto, não observam o ideal de justiça ínsito ao contexto histórico-cultural da sociedade em que se insere (Mendes; Mattiuzzo, 2019).

A crescente utilização de sistemas inteligentes na sociedade contemporânea para a tomada das mais variadas decisões tem conduzido a um estado de simbiose tecnológica (Zampier, 2021) que levanta questões sobre o uso justo e potencial discriminatório da IA no seio social. A acurácia dos sistemas de IA depende da qualidade do processo de aprendizagem que lhe foi programada ganhando determinabilidade na previsão de caracteres pessoais a partir da substituição de probabilidades por correlações, mediante uma cadeia complexa de eventos probabilísticos (Acioly, 2022).

## DIÁLOGOS POSSÍVEIS

ISSN 2447-9047  
VOLUME 23, Nº 2- JUL/DEZ 2024

Pág: 97 -121

Sem pretensão de exaurir a temática sobre o reconhecimento e a conceituação de *discriminação* (Schippers, 2018; Duarte, 2021), o presente estudo parte da premissa que as discriminações se materializam a partir da construção social de práticas historicamente contingentes, pautadas na subjetivação de grupos de pessoas por fatores como a cor da pele, origem étnico-racial, gênero, orientação sexual, ou qualquer forma de estigmatização (Duarte, 2021).

Por sua vez, há discriminação ainda quando as características individuais de uma pessoa culminam por serem desconsideradas para que esta somente seja associada como parte de um grupo homogêneo (Mendes; Mattiuzzo, 2019). Nesse contexto, o conceito de discriminação se afasta dos conceitos de preconceito, estereótipo ou ideologia, na medida em que, embora possa se fundar em alguns desses elementos, está efetivamente associado a um comportamento (Duarte, 2021).

A verificação de discriminação em sistemas inteligentes não pressupõe necessariamente a existência de subjetividades no programador ou aplicador da IA, mas pode decorrer da própria dinâmica do modelo de aprendizagem implementado (Baroca; Selbst, 2016). Dessa forma, ainda que inexista consciência, vontade ou culpa, um resultado proferido por um sistema inteligente pode representar uma discriminação.

Deve-se, contudo, ressaltar que o presente estudo não tem o condão de se filiar à teoria da irresponsabilidade distribuída (Magrani, 2019), em que a atribuição de responsabilidade ao ser inanimado – o sistema inteligente, cujo processamento se distribui em etapas – afastaria eventual responsabilização dos agentes humanos envolvidos. De igual maneira, cuida-se de se afastar da premissa de neutralidade tecnológica<sup>14</sup> (Gillespie, 2018), tendo em vista que a interação entre ser humano e artefato com potencial de alterar a realidade performa o conceito de sistemas sociotécnicos, carregados de uma intrínseca moralidade própria, que deve ser objeto de avaliação ética (Magrani, 2019; Selbst *et al.*, 2019).

Neste sentido, alguns eventos têm levantado a discussão sobre o potencial discriminatório decorrente do aprendizado de máquina. Magrani (2019) cita o caso do sistema inteligente da *Microsoft*, denominado *Tay*, que, dotado de uma capacidade de aprendizado profundo, ao ser inserido na rede social *Twitter*, passou a interagir através de conversações com usuários e aprender com as experiências acumuladas. Com menos de vinte e quatro horas de funcionamento, a *Tay* passou a reproduzir conteúdo em textos que, além de violar os Termos de

---

<sup>14</sup> O argumento da neutralidade tecnológica tem conduzido a discursos que banalizam a discriminação por algoritmos, num processo denominado de *matwashing*, mascarando a discussão sobre opacidade, justiça e responsabilidade na condução de sistemas inteligentes (Costa; Acioly, 2022).

## DIÁLOGOS POSSÍVEIS

ISSN 2447-9047  
VOLUME 23, Nº 2- JUL/DEZ 2024

Pág: 97 -121

Uso da plataforma, tinham conteúdo racista, sexistas e antisemita (Magrani, 2019; Kaufman, 2022), tornando-se vetor de reprodução explícita de discursos de ódio<sup>15</sup> (Vale, 2016).

A discriminação baseada no tratamento automatizado de dados pessoais deflui do processo de tomada de decisão por um sistema inteligente (Schippers, 2018), pautado em camadas lógicas de aprendizado que, como resultado, influi injustamente na esfera individual ou coletiva do ser humano (Kleinberget *al.*, 2018). Schippers (2018) sintetiza em dois os momentos de origem de discriminação na estrutura do sistema inteligente: na criação do modelo e *design* do algoritmo, estruturando os critérios que subsidiarão a tomada de decisão; e na formação da base de dados que irá ser utilizada no processo de treinamento e na composição dos alvos. Kleinberget *al.* (2018) registram, ainda, que a própria escolha do formato de aprendizagem pode ensejar um momento originário da discriminação no sistema inteligente.

Indo-se além, Barocas e Selbst (2016) analisam sistematicamente a inserção de potenciais discriminatórios na estruturação de um sistema inteligente pautado na técnica de mineração de dados, a partir dos seguintes elementos: (i) definição da variável-alvo (*target*) e dos rótulos de classe; (ii) treinamento do sistema com dados; (iii) seleção de recursos, ou critérios; (iv) *proxies*; e (v) mascaramento. A seleção da variável-alvo pode representar um contexto discriminatório, na medida em que o *target* escolhido pode, ao estabelecer um caminho de ordenação dos algoritmos, representar valores do programador, conscientes ou não, que resultem em um injusto resultado (Barocas; Selbst, 2016). Sobre esse contexto, Kleinberget *al.* (2018) apontam como exemplo o uso de sistema de IA para seleção de analistas orçamentários iniciantes que tem por alvo candidatos de universidades de maior prestígio, o que, nesse caso, reproduz a estrutura discriminatória que impede jovens negros e de baixa renda de terem acesso a oportunidades de estudo e trabalho.

Com efeito, esta mesma questão conduz à rotulagem de dados que, caso não seja realizada pelo programador – ora denominado *data miner*, no específico contexto da mineração de dados (Baroca; Selbst, 2016) – ter-se-ão dados não classificados, desconsiderando as especificidades de cada indivíduo (Simão, 2022). O fomento de dados pessoais nos sistemas inteligentes (*dataset*) com a finalidade de treinamento e os critérios estabelecidos para a construção do aprendizado, principalmente em sistemas de aprendizado não supervisionado (Kleinberget *al.*, 2018), semelhantemente conduz a um panorama discriminatório, tendo em vista

---

<sup>15</sup> Ainda à guisa de exemplificação, Bettega (2021) aponta casos de discriminação algorítmica de gênero: (1) em 2018, a Amazon desistiu de utilizar um sistema de IA de recrutamento de colaboradores quando detectado que esse havia sido maculado com um viés discriminatório contra mulheres, privilegiando-se a contratação de homens, pois havia sido alimentado com dados do estado recente da empresa (há até 10 anos), de predomínio da figura masculina; (2) após lançamento de cartão de crédito, a Apple foi posta em investigação, pois seu sistema de IA relacionava a questão de gênero ao limite oferecido, privilegiando-se homens com maiores linhas de crédito; (3) em 2019, o Google admitiu uma falha em seu algoritmo de pesquisa que associava a busca por profissionais mulheres ou por lésbicas a conteúdo pornográfico.

## DIÁLOGOS POSSÍVEIS

ISSN 2447-9047  
VOLUME 23, Nº 2- JUL/DEZ 2024  
Pág: 97 -121

não necessariamente o alvo, mas o caminho de aprendizado traçado para se chegar a um resultado (Schippers, 2018).

A influência da composição da base de dados no treinamento do sistema inteligente e seu potencial discriminatório é também visualizada na ótica antropocêntrica da abordagem da IA, de forma que Escrich e Reis (2020) ressaltam o caráter de representatividade do ser humano na contaminação da IA, de forma que o *dataset* escolhido para o treinamento irá representar a inteligência social que se busca implicar nesse sistema. A remoção de informações potencialmente discriminatórias, como raça e gênero, nas variáveis de entrada, no escopo da rotulagem ou ainda no *dataset*, não garante que o aprendizado não se tornará discriminatório (Schippers, 2018; Simão, 2022), tendo em vista que o processo de aprendizagem pode gerar inferências a partir da conjugação de dados intermediários – os chamados *proxies* – levando a resultados carregados de premissas enviesadas (Barocas; Selbst, 2016; Blackman, 2022). Duarte (2021, p. 68) recorre à expressão “*o CEP tem cor*” para exemplificar a correlação dos *proxies* a dados com potencial discriminatório a partir de informações tidas como neutras, o que incorre em uma codificação redundante<sup>16</sup>. Ademais, o mascaramento é a forma de manejo consciente do sistema de IA, para encobrir as reais intenções do *data miner* (Barocas; Selbst, 2016), a partir do discurso de neutralidade tecnológica e a consciência da possibilidade de vieses ínsitos ao aprendizado do sistema inteligente (Schippers, 2018).

Sob o prisma sociojurídico dos riscos discriminatórios associados aos processos de aprendizado de máquina, Mendes e Mattiuzzo (2019) elencam quatro formas de produção de resultados injustos no seio social, quais sejam: (i) a partir de discriminação oriunda de falha no processamento do algoritmo, seja no estabelecimento do *dataset*, seja na mecânica do aprendizado ou da árvore de decisões; (ii) a partir da discriminação por conta do uso de dados sensíveis<sup>17</sup> na formação do *dataset* ou na própria variável de entrada, levando em conta questões íntimas capazes de infirmar preconceitos, conscientes ou não, do programador; (iii) a partir da discriminação que resulte em limitação de direitos para os afetados – tal qual a prática comercial de *scores* de crédito (Acioly, 2022) –, caso em que o sistema inteligente carece de legitimidade social para tanto; e (iv) a partir da discriminação por generalização, isto é, quando o sistema atribui, diretamente ou através de *proxies*, características de um grupo a um certo indivíduo, desconsiderando suas

---

<sup>16</sup> Em linhas superficiais, “o CEP tem cor” (Duarte, 2021, p. 68), traz à luz a questão da marginalização da comunidade negra, que ocupa espaços urbanos periféricos em decorrência da impregnação do racismo estrutural nas camadas urbanísticas da sociedade (Rezende, Andrade, 2022). Portanto, ainda que o dado racial não seja diretamente tratado pelo sistema de IA, o manejo do Código de Endereçamento Postal pode revelar a informação sobre o espaço urbano que o indivíduo ocupa, ensejando o tratamento indireto da raça do indivíduo e, consequentemente, gerando resultados discriminatórios.

<sup>17</sup> Tomando-se como parâmetro a conceituação legal, aduzida no art. 5º, inciso II, da Lei Geral de Proteção de Dados Pessoais, dado sensível é o “dado pessoal sobre origem racial ou étnica, convicção religiosa, opinião política, filiação a sindicato ou a organização de caráter religioso, filosófico ou político, dado referente à saúde ou à vida sexual, dado genético ou biométrico”.

particularidades, seja por omissão do processo de rotulagem (Barocas; Selbst, 2016), seja por especificação insuficiente dos alvos e escolha de critérios de construção de aprendizado.

Decerto, tal formatação socio-legal de tipos de discriminação algorítmica, associada ao momento originário (Kleinberget *al.*, 2018) e a sistemática de estruturação do potencial discriminatório (Baroca; Selbst, 2016) tem o condão de contextualizar generalisticamente a discussão sobre a discriminação por tratamento automatizado de dados pessoais (Schippers, 2018), ensejando uma gama de respostas político-sociais e acadêmicas no escopo da antidiscriminação (Duarte, 2021; Green, 2021). Os programadores e aplicadores de IA detêm um poder político-social de construção normativa da sociedade, através da formatação de algoritmos de aprendizado (Green, 2021), impondo-se o reconhecimento do dever de estes profissionais aplicarem justiça social na elaboração dos algoritmos de aprendizado.

Sendo assim, torna-se importante salientar, como pressuposto da regulação da inteligência artificial, o risco imanente ao potencial de discriminação no âmbito do aprendizado de máquina deve afastar o paradigma da neutralidade tecnológica, bem como a irresponsabilidade distribuída, devendo a regulação ter por efeito o direcionamento do programador e aplicador da inteligência artificial no papel político-social de construção normativa da sociedade pautado em axiomas éticos e antidiscriminatórios.

### **5 DIRETRIZES ÉTICAS PARA O IMPLEMENTO DE UMA REGULAÇÃO ANTIDISCRIMINATÓRIA DOS SISTEMAS DE INTELIGÊNCIA ARTIFICIAL**

A crescente utilização de sistemas de inteligência artificial tem servido à formulação de políticas econômicas e socioculturais que impactam diretamente no contexto global. A utilização de redes neurais artificiais em conjunto com a crescente exponencial de *big data* têm conduzido a uma conjuntura que torna a inteligência artificial um mecanismo de tomada de decisões sobre aspectos centrais da vida humana (Diakopoulos, 2015; Lee, 2019). Pode-se tomar como exemplo a demanda crescente de sistemas inteligentes para a precificação de produtos no comércio a partir da geolocalização do consumidor e sua utilização na formulação de *scores* de crédito para instituições financeiras (Acioly, 2022), bem como o seu emprego para seleção de candidatos a cargos profissionais (Barocas; Selbst, 2016) ou acadêmicos (Kleinberget *al.*, 2018), ou, ainda, em sistemas de reconhecimento facial para fins de segurança pública (Magalhães; Gomes, 2021).

Deveras, este repertório de utilizações aliada ao potencial discriminatório inerente ao aprendizado de máquina tem revelado particularidades na abordagem socio-legal da regulação de inteligência artificial, a partir de um esforço difuso para o estabelecimento de diretrizes éticas. Blackman (2022) endereça três principais desafios éticos à governança ética em inteligência artificial: (i) privacidade, isto é a forma como o aprendizado de máquina observa as normas de

## DIÁLOGOS POSSÍVEIS

ISSN 2447-9047  
VOLUME 23, Nº 2- JUL/DEZ 2024

Pág: 97 -121

privacidade digital e uso coerente de dados pessoais; (ii) explicabilidade; (iii) mitigação de vieses discriminatórios. O imperativo de preservação de ditames éticos ganha relevo a partir da contextualização da discussão em ordenamentos jurídicos que põe em seu cerne a figura do ser humano, o que conduz ao movimento da IA centrada no ser humano – *humancentered design* – (Mulholland; Gomes, 2021), estruturando abordagens sociotécnicas para a sua inclusão no processo de desenvolvimento de IA (Earthy *et al.*, 2001).

Muito embora o movimento de *humancentered design* não seja objeto de aprofundamento no presente estudo, é importante consignar que a perspectiva de interação entre ser humano – considerando-o em suas necessidades, motivações, emoções, comportamento e perspectivas – e sistema computacional a partir de abordagens de racionalização e *design* humanístico (Auernhammer, 2020) tem influenciado a elaboração de princípios éticos para desenvolvimento de IA por diversas organizações (Mulholland; Frajhof, 2021). A ética, nesse contexto, se diferencia de valores morais por ter uma feição objetiva, estabelecendo princípios socialmente aceitos como fundantes das ações dos indivíduos num dado contexto sociocultural (Blackburn, 2002). A ética aplicada aos sistemas de IA, a seu turno, busca construir princípios que orientam o desenvolvimento do maquinário inteligente, bem como a sua capacidade de chegar a resultados que respeitem dados valores do ambiente sociocultural contemporâneo (Bartnecket *al.*, 2021).

Decerto, constata-se uma diversidade de documentos elaborados por instituições privadas e órgãos multissetoriais e internacionais, que se fundam na necessidade de implementar vetores deontológicos que fomentem o desenvolvimento auditável e confiável da IA (Mulholland; Frajhof, 2021). À guisa de exemplo, Mulholland e Gomes (2021) sintetizam a existência de documentos idealizados por instituições privadas transnacionais como a *Microsoft* e a *IBM*, bem como produzidos por entidades supranacionais, como a Conferência de Asilomar<sup>18</sup> e a *AcessNow*. A seu turno, Duarte (2021) aponta a idealização principiológica de entidades como a *Fairness, Accountability and Transparency in Machine Learning Organization* (“*FAT-ML*”) e a *Association for Computing Machinery* (“*ACM*”), que estruturam suas diretrizes éticas no entorno da responsabilidade e precaução.

Para fins do presente estudo, toma-se como critério de análise a fundamentação explícita no movimento de *humancentered design*, denotando caráter humanístico à formulação de modais deontológicos, e a correlação com marcos normativos de concretização de direitos fundamentais (Magrani, 2019), como a Declaração Universal dos Direitos Humanos de 1948. O presente estudo põe em foco a *Recommendation of the Council on Artificial Intelligence*, da Organização para a

---

<sup>18</sup> Mulholland e Gomes (2021) analisam os princípios de Asilomar a partir dos vetores de: (i) segurança e precaução a danos; (ii) transparência sobre incidentes de segurança; (iii) responsabilidade pelas implicações morais do uso da IA; (iv) proteção e controle de dados pessoais; e (v) atendimento a valores humanos.

## DIÁLOGOS POSSÍVEIS

ISSN 2447-9047  
VOLUME 23, Nº 2- JUL/DEZ 2024

Pág: 97 -121

Cooperação e Desenvolvimento Econômico (OCDE, 2019) e as Orientações Éticas para uma IA de Confiança, da Comissão Europeia (Comissão Europeia, 2019).

A Organização para a Cooperação e Desenvolvimento Econômico (OCDE) promoveu uma Recomendação a seu conselho, em vista a estabelecer uma diretriz de horizontalização da regulação de IA para seus Estados-membros, expressando a necessidade de segurança entre as relações econômicas no seu entorno (Yeung, 2020). Foram estabelecidos os princípios de (OCDE, 2019): (i) crescimento inclusivo, desenvolvimento sustentável e bem-estar, isto é, os agentes de IA devem realizar uma administração responsável para a produção de resultados benéficos para as pessoas e para o planeta; (ii) justiça e valores centrados no ser humano, isto é, os agentes de IA devem respeitar o estado de direito, os direitos humanos e os valores democráticos durante todo o ciclo de vida da IA; (iii) transparência e explicabilidade, isto é, os agentes de IA devem fornecer informações significativas, adequadas ao contexto e consistentes com o estado da arte do tema; (iv) robustez, segurança e precaução, em vista a promover o gerenciamento de riscos e fortalecer a segurança integral do sistema; e (v) *accountability*, que assume feição de responsabilidade, virtude de responsabilizar-se pelo bom funcionamento do sistema de IA (OCDE, 2019).

Por outro lado, a Comissão Europeia instituiu um Grupo Independente de Peritos de Alto Nível sobre a Inteligência Artificial – “GPAN IA” – em vista a fornecer orientações aos países que compõem a comunidade europeia no que se refere à promoção de uma IA de confiança<sup>19</sup>, indicando meios para operacionalizar certas diretrizes em sistemas sociotécnicos (Esteves, 2022). Embora haja uma série de lacunas práticas (Veale, 2020), a construção de sistemas sociotécnicos perpassa pelo endereçamento de discussões sobre o desenvolvimento de IA ao escrutínio público (Magrani, 2019). Seguindo-se esse pressuposto, o Grupo Independente de Peritos de Alto Nível sobre a Inteligência Artificial erigiu os seguintes ditames éticos para o desenvolvimento de uma IA confiável: (i) respeito à autonomia humana; (ii) prevenção de danos; (iii) equidade; (iv) explicabilidade (Comissão Europeia, 2019). O respeito à autonomia humana é um vetor principiológico com dupla funcionalidade. Na medida em que orienta a precaução contra coação ou condicionamento humano ao sistema inteligente, impõe que a construção da IA seja realizada para complementar as competências cognitivas e socioculturais humanas (Esteves, 2022).

O GPAN IA, ao descrever o princípio da prevenção, correlacionou-o à proteção da pessoa humana no tocante à sua dignidade e integridade física e mental, contra utilizações malélicas, referenciando, ainda, o risco de agravamento de assimetria de poder e assimetrias informacionais (Comissão Europeia, 2019). A equidade, a seu turno, representa, em sua dimensão substantiva, o

---

<sup>19</sup> A construção de uma IA de confiança, na forma do Guia elaborado pelo GPAN IA, se fundamenta em três componentes que devem ser conjugados simultaneamente (Comissão Europeia, 2019): o respeito à legislação e regulamento aplicáveis; a observância das diretrizes éticas; a solidez do ponto de vista técnico e social, tendo em vista que os sistemas de IA podem causar danos não intencionais (Esteves, 2022).

## DIÁLOGOS POSSÍVEIS

ISSN 2447-9047  
VOLUME 23, Nº 2- JUL/DEZ 2024

Pág: 97 -121

compromisso de garantir a distribuição justa dos benefícios e custos da IA (Comissão Europeia, 2019). Em sua dimensão processual, a equidade representa a garantia de contestação em face de decisões automatizadas, identificando-se a entidade responsável pela decisão e impondo explicabilidade aos processos decisórios (Comissão Europeia, 2019). O princípio da explicabilidade significa que os processos devam ser transparentes e as capacidades e finalidades dos sistemas inteligentes devam ser explicáveis na medida do possível, considerando que a necessidade de explicabilidade depende do contexto e da gravidade das consequências de um resultado incorreto (Comissão Europeia, 2019).

Em comum aos panoramas éticos postos em foco na presente pesquisa, verifica-se a explicabilidade dos sistemas inteligentes como supedâneo ao controle social dos algoritmos e escrutínio público dos riscos discriminatórios decorrentes do aprendizado de máquina (Mittelstadt *et al.*, 2016). A característica de inescrutabilidade das árvores de decisões que compõem as redes neurais artificiais, e conseqüentemente, a sua imprevisibilidade (Frazão; Goettenauer, 2021), tem conduzido à prevalência de uma certa opacidade na governança social da IA. Pasquale (2015) denomina os algoritmos de aprendizado de *black-boxes*, posto que envoltos em um segredo técnico e comercial que impede qualquer esforço para seu controle ou correção<sup>20</sup>.

Como assinalado no item acima, o mascaramento de vieses a partir do argumento de neutralidade tecnológica (Barocas; Selbst, 2016) tem deslocado um centro de poder para os desenvolvedores do sistema de IA e grandes corporações do mercado tecnológico (Pasquale, 2015). O desconhecimento estratégico que circunscreve a opacidade das redes neurais artificiais tem sido utilizado politicamente para evitar responsabilidades dos agentes de IA (Bucher, 2018), postergando ou mitigando discussões sobre resultados discriminatórios.

Sem embargo de consignar a discussão que envolve a semântica dos termos explicabilidade, interpretabilidade e compreensibilidade<sup>21</sup> (Rudin, 2019), para fins do presente estudo, importa consignar que assimetria informacional oriunda da opacidade (Frazão; Goettenauer, 2021) constitui objeto de enfrentamento pelos vetores éticos de desenvolvimento da IA.

Dando cores ao problema da opacidade, Duarte (2021) assevera que a mera disponibilidade de informações, ainda que do código-fonte, não tem o condão de suprir a lacuna deixada pela inescrutabilidade das redes neurais artificiais, posto que a própria lógica dos

---

<sup>20</sup> Pasquale (2015) utiliza o conceito de “*onewaymirror*” para fazer alusão à figura dos espelhos unidirecionais utilizados em estabelecimentos policiais americanos, que permite a vigilância do ambiente por quem está atrás do espelho e reflete a imagem para aqueles que estão internos à sala. A realização dessa analogia serve à demonstração do estado de supervigilância ocasionado pela opacidade dos algoritmos presentes no cotidiano social, permitindo a coleta massiva de informações pessoais dos cidadãos por grandes *players* do mercado e por governos (Pasquale, 2015).

<sup>21</sup> Nesse contexto, Rudin (2019) destaca que explicabilidade é a possibilidade de o sistema ser inteligível através de explicações adicionais, não nativas ao próprio código, enquanto a interpretabilidade é a compreensibilidade intrínseca, inerente, a despeito de qualquer explicação adicional.



algoritmos de aprendizado, de natureza adaptativa e autônoma, é um entrave à efetiva governança social.

Por sua vez, a explicabilidade se materializa a partir da construção do sistema de forma que seja possível a efetiva prestação de contas por meio de documentos compreensíveis (Kroll *et al.*, 2017). O sentido e alcance do princípio da explicabilidade é dado pela necessidade de compreensibilidade e acessibilidade na resolução de problemas no caso concreto, mantendo-se uma estrutura de confiança entre o usuário e o agente de IA (Mittelstadt *et al.*, 2016).

Destarte, muito além de retórica argumentativa (Blackman, 2022), a explicabilidade pode ser endereçada como um princípio que viabiliza a concretização de seus pares, o que dirige à sua materialização no caso concreto.

### CONSIDERAÇÕES FINAIS

Ao longo do presente estudo, foi possível entender que a noção conceitual de inteligência artificial é difusa, não encontrando consenso doutrinário. Para o objetivo de subsidiar uma política regulatória eficiente, a conceituação de sistemas de inteligência artificial deve considerar, todavia, as nuances dos seus atributos formativos, inclusive para distanciá-la da mera programação algorítmica sofisticada.

Para tanto, o presente estudo propõe a conceituação da sistemas de inteligência artificial como o modelo automatizado de processamento de informações em camadas de encadeamento lógico, através de treinamento e aprendizagem, cuja função é resolver questões de forma semelhante a algum espectro de inteligência humana. Circunscreve-se o ambiente regulatório aos sistemas que se encaixem nessa conceituação, afastando obrigações de desenvolvedores e implementadores de outras formas de sistemas de informação.

Esse distanciamento se concretiza quando se observa que o aprendizado de máquina (*machinelearning*) comporta um amplo espectro de abordagens, que em diversos casos são manejadas conjuntamente para aprimoramento do processo cognitivo do sistema de Inteligência Artificial. A par dessa perspectiva, o presente estudo põe em foco a constatação de que o desenvolvimento de sistemas de IA não é uniforme, não sendo aplicável uma estrutura estática ou unidirecional, sub pena de insuficiência técnica da regulação.

Neste sentido, esta amplitude de abordagens tem conduzido a um conjunto plúrimo de riscos a direitos fundamentais, especialmente no tocante ao potencial de discriminação por tratamento automatizado de dados pessoais. Para tanto, o presente estudo consigna a necessidade de afastamento do paradigma da neutralidade tecnológica e da irresponsabilidade distribuída, para compreender que a abordagem do aprendizado e questões técnicas podem resultar em um processo

discriminatório, o que conduz à necessidade de imposição de padrões éticos e responsabilidades aos desenvolvedores e implementadores de sistemas de IA.

Esses vetores éticos se encontram distribuídos em normas supranacionais, de entidades privadas e associações acadêmicas, refletindo um esforço difuso na construção de parametrização do desenvolvimento de sistemas de IA.

O estabelecimento de vetores éticos para o desenvolvimento de IA constitui ponto de partida, não de chegada, para o enfrentamento de riscos técnicos e sociais, especialmente de discriminação algorítmica, no que concerne aos desafios de formulação de uma prescrição normativa endereçada ao aprendizado de máquina e às redes neurais artificiais.

Deste modo, o caráter flexível desses modais deontológicos tem por funcionalidade servir como supedâneo à formulação de estruturas regulatórias para a concretização de direitos fundamentais do ser humano na direção do *humancentered design*.

### REFERÊNCIAS

ACIOLY, Luis Henrique de Menezes. Reflexões sobre a Utilização da Inteligência Artificial e Algoritmos nas Relações Consumeristas à Luz da Lei Geral de Proteção de Dados. **Revista Brasileira de Direito Comercial: Concorrência, Empresarial e do Consumidor**. Porto Alegre: Lex Magister. v. 48, ago./set. 2022, p. 165-187.

ACIOLY; Luis Henrique de Menezes; OLIVEIRA, Micael Araújo de Souza. Espectro de Aplicação dos Smart Contracts à luz do Direito Contratual Brasileiro: Breves Considerações no contexto da Boa-fé objetiva. **Conversas Civilísticas**, Salvador, v. 2, n. 2, 2022, p. 98-121. Disponível em: <https://periodicos.ufba.br/index.php/conversascivilisticas/article/view/51935>. Acesso em: 2 jul. 2023.

AUERNHAMMER, Jan. Human-centered AI: The role of Human-centered Design Research in the development of AI. In: BOESS, S., CHEUNG, M. and CAIN, R. (eds.). **Synergy - DRS International Conference 2020**, p. 11-14, ago. 2020. Disponível em: <https://doi.org/10.21606/drs.2020.282>. Acesso em: 15 jul. 2023.

BAROCAS, Solon; ROSENBLAT, Alex; BOYD, Danah; GANGADHARAN, Seeta Peña; YU, Corrine. Data & Civil Rights: Technology Primer. **Data & Civil Rights Conference**, October 2014, Disponível em: <http://dx.doi.org/10.2139/ssrn.2536579>. Acesso em: 02 jul. 2023.

BAROCAS, Solon; SELBST, Andrew. Big Data's Disparate Impact. **California Law Review**, v. 104, 2016, p. 671-732. Disponível em: <http://dx.doi.org/10.2139/ssrn.2477899>. Acesso em: 28 jun. 2023.

BARTNECK, Christoph; LUTGE, Christoph; WAGNER, Alan; WELSH, Sean. **Na Introduction to Ethics in Robotics and AI**. Cham: Springer Switzerland, 2021.

BETTEGA, Elisa. O que fazer sobre o viés algorítmico baseado em gênero?. In: BARBOSA, Bia; TRESKA, Laura; LAUSCHNER, Tanara (orgs.). **Governança da Internet e Gênero: tendências e desafios**. [s.l.]: CGI.Br., 2021. p. 125-133.

## DIÁLOGOS POSSÍVEIS

ISSN 2447-9047  
VOLUME 23, Nº 2- JUL/DEZ 2024

Pág: 97 -121

BIONI, Bruno Ricardo. **Proteção de dados pessoais**: a função e os limites do consentimento. 3. ed. rev., ampl. e atual. Rio de Janeiro: Forense, 2021.

BLACKBURN, Simon. **Being Good**: a short introduction to ethics. Oxford: Oxford University Press, 2002.

BLACKMAN, Reid. **Ethical Machines**: Your Concise Guide to Totally Unbiased, Transparent, and Respectful AI. Cambridge: Harvard Business Review Press, 2022.

BUCHER, Taina. **If... then**: algorithmic power and politics. New York: Oxford University Press, 2018.

COMISSÃO EUROPEIA. **Orientações éticas para uma IA de confiança**. Publications Office, 2019. Disponível em: <https://data.europa.eu/doi/10.2759/2686>. Acesso em: 15 jul. 2023.

COSTA, Jéssica Hind Ribeiro; ACIOLY, Luis Henrique de Menezes Acioly. Discriminação Racial e Inteligência Artificial: Caminhos Possíveis à Plena Cidadania Virtual. In: MACEDO, Humberto Gomes; BENTES, Dorinethe dos Santos; ANDRADE, Pedro Gustavo Gomes (coord.). **Os direitos humanos na era tecnológica III**. III Congresso Internacional de Direito e Inteligência Artificial (III CIDIA). Belo Horizonte: Skema Business School, 2022, p. 53-60.

DIAKOPOULOS, Nicholas. Algorithm accountability: journalistic investigation of computational power structures. **Digital Journalism**, v. 3, n. 3, 2015, p. 398-415. Disponível em: <https://doi.org/10.1080/21670811.2014.976411>. Acesso em: 15 jul. 2023.

DUARTE, Alan. **A Antidiscriminação no contexto da inteligência artificial**: possibilidades de governança mediante a normatização de algoritmos. Fortaleza: Mucuripe, 2021.

EARTHY, Jonathan; JONES, Brian Sherwood; BEVAN, Nigel. The improvement of human-centered processes - facing the challenge and reaping the benefit of ISO 13407. **International Journal of Human-Computer Studies**, v. 55, n. 4, oct. 2021, p. 553-585. Disponível em: <https://doi.org/10.1006/ijhc.2001.0493>. Acesso em: 15 jul. 2023.

ESCRICH, Rafael; REIS, Guilherme. O panorama geral entre Inteligência Artificial e a Sociologia. In: LONGHI, João Vitor Rozzatti.; FALEIROS JÚNIOR, José Luiz de Moura.; BORGES, Guilher Oliveira de Aguiar.; REIS, Guilherme (coords.). **Fundamentos do Direito Digital**. Uberlândia: LAECC, 2020. p. 339-362.

ESTEVES, Andresa Silveira. **Um estudo sobre a construção da Inteligência Artificial de Confiança sob o enfoque dos Direitos Humanos**. 2022. 164 f. Dissertação (Curso de Mestrado em Ciência Jurídica) - Programa de Pós-Graduação Stricto Sensu em Ciência Jurídica, Universidade do Vale do Itajaí, Itajaí, 2022.

FONSECA, Ana Taveira. Smart Contracts. In: BARBOSA, Mafalda Miranda; BRAGA NETTO, Felipe; SILVA, Menezes Cordeiro; FALEIROS JÚNIOR, José Luiz Moura. (coords.). **Direito Digital e Inteligência Artificial**: Diálogos entre Brasil e Europa. Indaiatuba: Foco, 2021, p. 743-752.

FRAZÃO, Ana; GOETTENAUER, Carlos. Black box e o direito face à opacidade algorítmica. In: BARBOSA, Mafalda Miranda; BRAGA NETTO, Felipe; SILVA, Menezes Cordeiro; FALEIROS JÚNIOR, José Luiz Moura. (coords.). **Direito Digital e Inteligência Artificial**: Diálogos entre Brasil e Europa. Indaiatuba: Foco, 2021, p. 27-42.

## DIÁLOGOS POSSÍVEIS

ISSN 2447-9047  
VOLUME 23, Nº 2- JUL/DEZ 2024

Pág: 97 -121

GABRIEL, Martha. **Inteligência Artificial: Do Zero ao Metaverso**. São Paulo: Atlas, 2022.

GILLESPIE, Tarleton. A relevância dos algoritmos. **Parágrafo**, [S.l.], v. 6, n. 1, jan./abr. 2018, p. 95-121. Disponível em: <https://revistaseletronicas.fiamfaam.br/index.php/recicofi/article/view/722>. Acesso em 08 jul. 2023.

GREEN, Ben. Data Science as Political Action: Grounding Data Science in a Politics of Justice. **Journal of Social Computing**, v. 2, n. 3, sept. 2021, p. 249-265. Disponível em: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3658431](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3658431). Acesso em 09 jul. 2023.

HOFFMANN-RIEM, Wolfgang. **Teoria Geral do Direito Digital**. 2. ed. Trad. Italo Furhmann. Rio de Janeiro: Forense, 2021.

KAPLAN, Jerry. **Artificial Intelligence: what everyone needs to know**. Oxford: Oxford University Press, 2016.

KAUFMAN, Dora. **Desmistificando a Inteligência Artificial**. Belo Horizonte: Autêntica, 2022.

KLEINBERG, John; LUDWIG, Jens; MULLAINATHAN, Sendhil; SUNSTEIN, Cass. Discrimination in the Age of Algorithms. **Journal of Legal Analysis**, v. 10, 2018, p. 113–174. Disponível em: <https://doi.org/10.1093/jla/laz001>. Acesso em 08 jul. 2023.

KOVÁCS, Zsolt. **Redes Neurais Artificiais: Fundamentos e Aplicações**. 4. ed. rev. São Paulo: Livraria da Física, 2006.

KROLL, Joshua; HUEY, Joanna; BAROCAS, Solon; FELTEN, Edward; REIDENBERG, Joel; ROBINSON, David; YU, Harlan. Accountable Algorithms. **University of Pennsylvania Law Review**, v. 165, 2017, p. 633-706. Disponível em: [https://scholarship.law.upenn.edu/penn\\_law\\_review/vol165/iss3/3/](https://scholarship.law.upenn.edu/penn_law_review/vol165/iss3/3/). Acesso em: 15 jul. 2023.

LAGE, Fernanda de Carvalho. **Manual de Inteligência Artificial no Direito Brasileiro**. 2. ed. rev., atual. e ampl. Salvador: JusPodivm, 2022.

LEE, Kai-Fu. **Inteligência artificial: como os robôs estão mudando o mundo, a forma como amamos, nos relacionamos, trabalhamos e vivemos**. Trad. Marcelo Barbão. Rio de Janeiro: Globo Livros, 2019.

MAGALHÃES, Alice Azevedo; GOMES, Tércio Spínola. Regulação de sistemas de reconhecimento facial para fins de segurança pública no Brasil: riscos e desafios. **Revista Humanidades e Inovação**. v. 8, n. 47, 2021, p. 168-182. Disponível em: <https://revista.unitins.br/index.php/humanidadesinovacao/article/view/5639>. Acesso em: 15 jul. 2023.

MAGRANI, Eduardo. New perspectives on ethics and the law of artificial intelligence. **Internet Policy Review**, v. 8, n. 3, 2019. Disponível em: <https://doi.org/10.14763/2019.3.1420>. Acesso em 08 jul. 2023.

MAYER-SCHÖNBERGER, Viktor; CUKIER; Kenneth. **Big Data, como extrair volume, variedade e valor**. Trad. Paulo Polzonoff Junior. Rio de Janeiro: Elsevier, 2013.

## DIÁLOGOS POSSÍVEIS

ISSN 2447-9047  
VOLUME 23, Nº 2- JUL/DEZ 2024

Pág: 97 -121

MCCULLOCH, Warren, PITTS, Walter. A logical calculus of the ideas immanent in nervous activity. **Bulletin of Mathematical Biophysics**, n. 5, 1943, p. 115-133. Disponível em: <https://doi.org/10.1007/BF02478259>. Acesso em: 09 jul. 2023.

MENDES, Laura Schertel; MATTIUZZO, Marcela. Discriminação Algorítmica: Conceito, Fundamento Legal e Tipologia. **Direito Público**, [S.l.] v. 16, n. 90, p. 39-64, nov-dez 2019. Disponível em: <https://www.portaldeperiodicos.idp.edu.br/direitopublico/article/view/3766>. Acesso em: 02 jul. 2023.

MINISTÉRIO DAS RELAÇÕES EXTERIORES. **Panorama Internacional: Políticas Nacionais e Institutos de Inteligência Artificial**. Dezembro, 2022. Brasília. Disponível em: <https://www.gov.br/mre/pt-br/assuntos/ciencia-tecnologia-e-inovacao/PanoramaInternacionalPolíticasNacionaisInstitutosdeInteligenciaArtificialV2.pdf>. Acesso em: 29 jul. 2023.

MITTELSTADT, Brent Daniel; ALLO, Patrick; TADDEO, Mariarosaria; WACHTER, Sandra; FLORIDI, Luciano. The ethics of algorithms: Mapping the debate. **Big Data & Society**. v. 3, n. 2, jul./dez., 2016, p. 1-21. Disponível em: <https://doi.org/10.1177/2053951716679679>. Acesso em: 23 jul. 2023.

MULHOLLAND, Caitlin; FRAJHOF, Isabella. Entre as leis da robótica e a ética: regulação para o adequado desenvolvimento da inteligência artificial. In: BARBOSA, Mafalda Miranda; BRAGA NETTO, Felipe; SILVA, Menezes Cordeiro; FALEIROS JÚNIOR, José Luiz Moura. (coords.). **Direito Digital e Inteligência Artificial: Diálogos entre Brasil e Europa**. Indaiatuba: Foco, 2021, p. 65-80.

MULHOLLAND, Caitlin; GOMES, Rodrigo Dias de Pinho. Inteligência Artificial e Seus Principais Desafios para os Programas de Compliance e as Políticas de Proteção de Dados. In: CUEVA, Ricardo Villas Bôas; FRAZÃO, Ana (coords.). **Compliance e Política de Proteção de Dados**. São Paulo: Thomson Reuters Brasil, 2021. p. 161-180.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. **Recommendation of the Council on Artificial Intelligence**. OECD/LEGAL/0449. Paris: OECD, 2019.

PASQUALE, Frank. **The Black Box Society**. Cambridge: Harvard University Press, 2015.

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

PINTO, Henrique Alves. A utilização da inteligência artificial no processo de tomada de decisões: por uma necessária accountability. **Revista de Informação Legislativa**, a. 57 n. 225, Brasília, jan./mar. 2020, p. 43-60. Disponível em: [https://www12.senado.leg.br/ril/edicoes/57/225/ril\\_v57\\_n225\\_p43.pdf](https://www12.senado.leg.br/ril/edicoes/57/225/ril_v57_n225_p43.pdf). Acesso em: 02 jul. 2023.

REIS, Paulo Victor Alfeo. **Algoritmos e o Direito**. São Paulo: Almedina, 2020.

RIBEIRO, Júlia Melo Carvalho. Inteligência artificial e responsabilidade civil sob a perspectiva do direito do consumidor. In: CHAVES, Natália Cristina; COLOMBI, Henry (orgs.). **Direito e Tecnologia: novos modelos e tendências**. Porto Alegre: Fi, 2021. p. 313-336.

ROSA, Fernanda. Entendendo os algoritmos: propriedades e dilemas. In: **ICT households 2017**. Núcleo de Informação e Coordenação do Ponto BR, São Paulo: CGI.br. 2017, p. 67-73.

## DIÁLOGOS POSSÍVEIS

ISSN 2447-9047  
VOLUME 23, Nº2- JUL/DEZ 2024  
Pág: 97 -121

RUDIN, Cynthia. Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. **Nature machine intelligence**, v. 1, n. 5, 2019, p. 206-215. Disponível em: <https://doi.org/10.1038%2Fs42256-019-0048-x>. Acesso em: 23 jul. 2023.

RUSSELL, Stuart; NORVIG, Peter. **Inteligência Artificial**. 3. ed. Trad. Regina Célia Simille de Macedo. Rio de Janeiro: Elsevier, 2013.

SAMUEL, Arthur. Some Studies in Machine Learning Using the Game of Checkers. **IBM Journal of Research and Development**. v. 3. n. 3, jul. 1959, p. 206-226. Disponível em: <https://citeseerx.ist.psu.edu/doc/10.1.1.368.2254>. Acesso em: 09 jul. 2023.

SCASSA, Teresa. Regulation AI in Canada: A critical look at the proposed Artificial Intelligence and Data Act. **The Canadian Bar Review**, v. 101, n. 1, 2023: Emerging Issues in Technology Law and Intellectual Property, p. 1-30. Disponível em: <https://cbr.cba.org/index.php/cbr/article/view/4817>. Acesso em: 29 ju. 2023.

SCHIPPERS, Laurianne-Marie. **Algoritmos que discriminam**: uma análise jurídica da discriminação no âmbito das decisões automatizadas e seus mitigadores. 2018. 57f. Trabalho de Conclusão de Curso (Bacharelado em Direito) – Escola de Direito de São Paulo, Fundação Getúlio Vargas, São Paulo, 2018.

SELBST, Andrew; BOYD, Danah; FRIEDLER, Sorelle; VENKATASUBRAMANIAN, Suresh; VERTESI, Janet. Fairness and Abstraction in Sociotechnical Systems. In: **FAT '19: Proceedings of the Conference on Fairness, Accountability, and Transparency**, jan. 2019, p. 59-68. Disponível em: <https://doi.org/10.1145/3287560.3287598>. Acesso em: 08 jul. 2023.

SIEGMANN, Charlotte; ANDERLJUNG, Markus. **The Brussels Effect and Artificial Intelligence**: How EU regulation will impact the global AI market. Oxford: Centre for the Governance of AI, 2022.

SIMÃO, Ana Luisa Teotônio Josafá. Ética e Discriminação no desenvolvimento de algoritmos de Machine Learning. In: PARENTONI, Leonardo (coord.). **Direito, tecnologia e inovação IV**: Estudos de caso. Belo Horizonte: Centro DTIBR, 2022, p. 27-52.

SKINNER, Burrhus Frederic. **Sobre o Behaviorismo**. 15. ed. Trad. Maria da Penha Villalobos. São Paulo: Cultrix, 2009.

UNIÃO EUROPEIA. **Processo 2021/0106/COD**. Proposta de Regulamento do Parlamento Europeu e do Conselho que estabelece regras harmonizadas em matéria de inteligência artificial (Regulamento Inteligência Artificial) e altera determinados atos legislativos da União. Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/HTML/?uri=CELEX:52021PC0206>. Acesso em: 15 jul. 2023.

VALE, Simone do. Inteligência Artificial & Redes Sociais: notas sobre um bot que odiava humanos. In: **Águila**: Revista Interdisciplinar da Universidade Veiga de Almeida, n. 15, jul./dez. 2016, p. 36-48.

VEALE, Michael. A Critical Take on the Policy Recommendations of the EU High-Level Expert Group on Artificial Intelligence. **European Journal of Risk Regulation**, v. 11, n. 1, mar. 2020. Disponível em: <https://doi.org/10.1017/err.2019.65>. Acesso em: 23 jul. 2023.

YEUNG, Karen. Introductory notes to Recommendation of the Council on Artificial Intelligence. **International Legal Materials**, v. 59, n. 1, fev. 2020, p. 27-34. Disponível em: <https://doi.org/10.1017/ilm.2020.5>. Acesso em: 23 jul. 2023.

ZAMPIER, Bruno. **Bens Digitais**. 2. ed. Indaiatuba: Foco, 2021.