

INTELIGÊNCIA ARTIFICIAL E PRECONCEITO DE IDENTIDADE DE GÊNERO: O PROBLEMA DO VIÉS NA CONSTRUÇÃO DAS IA'S E A PERPETUAÇÃO DAS DISCRIMINAÇÕES EM SOCIEDADES PREVIAMENTE DISCRIMINATÓRIAS

ARTIFICIAL INTELLIGENCE AND GENDER IDENTITY PREJUDICE: THE PROBLEM OF BIAS IN THE CONSTRUCTION OF AI'S AND THE PERPETUATION OF DISCRIMINATION IN PREVIOUSLY DISCRIMINATORY SOCIETIES

Anderson Jordan Alves Abreu¹
Kathya Cristhyna Silva Furtado²
Rennan Kevim Costa Santos³

Resumo: a presente pesquisa almejou analisar o problema do viés no aprendizado de máquina, e sua relação com a perpetuação de preconceitos de identidade de gênero em sociedades previamente discriminatórias. O problema que guiou a pesquisa foi o seguinte: a Inteligência Artificial favorece o preconceito de identidade de gênero em sociedades previamente discriminatórias? De que forma é possível solucionar a discriminação algorítmica? Nesse sentido, a pesquisa foi construída através de abordagem qualitativa e método indutivo, se realizando por meio de uma revisão de literatura que se deu por procedimento exploratório e bibliográfico, no qual se investigou a literatura pertinente ao tema. O trabalho está dividido em seis partes, nas quais se analisou o conceito de gênero e identidade de gênero, o conceito de Inteligência Artificial e suas aplicações, o aprendizado de máquina e a forma como as IAs aprendem, o problema do viés em aprendizado de máquina, a relação entre viés em aprendizado de máquina nas IAs e a perpetuação de preconceitos de identidade de gênero, e as possíveis soluções para superar o problema. Ao fim, concluiu-se que há relação entre o viés em aprendizado de máquina e a perpetuação de preconceitos de identidade de gênero, sendo uma mazela a ser superada por meio de regulação ética e princípios a serem estabelecidos e observados por instituições e governos, de modo a realizar a evolução tecnológica baseada no respeito e na igualdade.

Palavras-chave: inteligência artificial; viés; preconceito; identidade; gênero.

Abstract: the present research aimed to analyze the problem of bias in machine learning, and its relationship with the perpetuation of gender identity prejudices in previously discriminatory societies. The problem that guided the research was the following: does Artificial Intelligence favor gender identity prejudice in previously discriminatory societies? How is it possible to solve algorithmic discrimination? In this sense, the research was built through a qualitative approach and inductive method, being carried out through a literature review that took place through an exploratory and bibliographic procedure, in which the literature relevant to the topic was investigated. The work is divided into six parts, in which the concept of gender and gender identity, the concept of Artificial Intelligence and its applications, machine learning and the way in which AIs learn, the problem of bias in machine learning were analyzed. , the relationship between machine learning bias in AIs and the perpetuation of gender identity biases, and possible solutions to overcome the problem. In the end, it was concluded that there is a relationship between the bias in machine learning and the perpetuation of gender identity prejudices, being a problem to be overcome through ethical regulation and principles established and observed by institutions and governments, in order to carry out technological evolution based on respect and equality.

Keywords: artificial intelligence; bias; preconception; identity; gender.

¹Bacharel em Direito pelo Instituto de Ensino Superior do Sul do Maranhão (IESMA - UNISULMA). Especialista em Direito Digital, Gestão da Inovação e Propriedade Intelectual pela Pontifícia Universidade Católica de Minas (PUC Minas). Especialista em Direitos Humanos, Responsabilidade Social e Cidadania Global pela Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS).

²Bacharel em Direito pela Faculdade de Imperatriz (FACIMP Wyden). Pós graduanda em Direito Público pela Faculdade de Educação Santa Terezinha (FEST).

³Bacharel em Direito pela Pontifícia Universidade Católica de Goiás (PUC GO). Advogado. Pós Graduando em Direito Digital pela Universidade do Vale do Rio dos Sinos (UNISINOS).

1 INTRODUÇÃO

A presente pesquisa buscou analisar o problema do viés preconceituoso na construção das Inteligências Artificiais e como se relaciona com a perpetuação das discriminações baseadas em identidade de gênero em sociedades previamente discriminatórias. A pesquisa partiu do seguinte problema: a Inteligência Artificial favorece o preconceito de identidade de gênero nessas sociedades? De que forma é possível solucionar a discriminação algorítmica?

Nesse sentido, este trabalho teve por objetivo geral analisar se as IAs promovem a perpetuação do preconceito de identidade de gênero em sociedades previamente preconceituosas, e de que forma é possível superar esta questão. Como objetivos específicos, a pesquisa almejou analisar o preconceito de identidade de gênero; conceituar Inteligência Artificial; compreender o aprendizado de máquina; investigar o problema do viés no aprendizado de máquina; verificar a relação entre o aprendizado de máquina e o preconceito de identidade de gênero na construção das Inteligências Artificiais; e por fim, analisar de que formas é possível superar o preconceito algorítmico gerado por IAs enviesadas.

A pesquisa está dividida da seguinte forma: na primeira parte foi analisada a diferença entre gênero e identidade de gênero, com enfoque para o problema da discriminação LGBTQIA+. Na segunda parte, buscou-se compreender o que é uma Inteligência Artificial e como esta tem sido aplicada na sociedade atual. Na terceira parte, evidenciou-se o que é o aprendizado de máquina (machine learning) e a forma como as IAs aprendem. Na quarta parte, identificou-se o que é o viés em Inteligência Artificial e os malefícios que este traz. Na quinta parte, foi analisada a relação entre o viés em IAs e a perpetuação da discriminação baseada em identidade de gênero em sociedades preconceituosas. Na sexta parte, buscou-se expor alternativas para superar o problema do viés preconceituoso em IAs.

A pesquisa se deu por método qualitativo e indutivo, efetuada por meio de procedimento bibliográfico e exploratório, através de uma revisão de literatura, em que se utilizaram artigos e livros extraídos de fontes e bases de dados confiáveis como o Portal de Periódicos Capes e a biblioteca temática do Superior Tribunal de Justiça.

Ao termo, evidenciou-se que as IAs enviesadas podem não apenas perpetuar preconceitos em sociedades previamente preconceituosas, mas também aumentar estas discriminações, havendo já diversos casos de preconceitos e discriminações por meio de IAs preconceituosas, sobretudo preconceitos raciais e de identidade gênero. Ainda, o tema é de relevância e tem sido tratado como prioridade dentro das organizações, instituições e governos que buscam regular o uso ético das IAs.

2 GÊNERO, IDENTIDADE DE GÊNERO, E PRECONCEITO DE IDENTIDADE DE GÊNERO

O entendimento sobre os conceitos de gênero, identidade de gênero e sexualidade(s) são de fundamental importância para esta pesquisa. Dito isso, o conceito de gênero, de acordo com Cruz *et al.* (2021), difere diametralmente do conceito de sexo, sendo o primeiro uma construção social, moldada e estabelecida pela relação social com outros indivíduos, e o segundo se tratando de aspectos fisiológicos e anatômicos, os quais diferenciam o corpo entre masculino e feminino. Santos (2021) traz o conceito segundo o qual gênero é o “corpo em sua materialidade” (SANTOS, 2021, p. 20). Assim, pode-se entender que o gênero é um construto social de caráter interno, residindo no âmbito íntimo do indivíduo e a partir dele se externando. Neste sentido, Oyèrónké Oyèwùmí, afirma que o gênero é, acima de tudo, uma construção sociocultural, bem como critica que a ideia de gênero não é um conceito universal, mas sim está relacionado à raça e à classe (OYÈWÙMÍ, 2004, p. 04).

Ainda, Costa (2020) destaca que sexo e sexualidade também não se confundem, sendo o sexo, novamente, remetido como aspecto anatômico e fisiológico, e a sexualidade sendo relativa à todo o espectro de relações de cunho sexual entre o indivíduo e outros. A Organização Mundial da Saúde define sexualidade da seguinte forma:

Aspecto central do ser humano ao longo da vida; ela engloba sexo, **identidades e papéis de gênero**, orientação sexual, erotismo, prazer, intimidade e reprodução. A sexualidade é vivida e expressada por meio de pensamentos, fantasias, desejos, crenças, atitudes, valores, comportamentos, práticas, papéis e relacionamentos. Embora a sexualidade possa incluir todas essas dimensões, nem sempre todas elas são vividas ou expressas. A sexualidade é influenciada pela interação de fatores biológicos, psicológicos, sociais, econômicos, políticos, culturais, jurídicos, históricos, religiosos e espirituais (OMS, 2020, p. 17, grifos nossos).

Cabe aqui apontar para as diversas sexualidades, como a bissexualidade, a pansexualidade e a assexualidade, dentre outras. Jaeger *et al.* (2019) aponta que a “bissexualidade tem sido usada como um termo ‘guarda-chuva’ para se referir à sexualidade de pessoas que sentem atração sexual e/ou afetiva por mais de um gênero, incluindo identidades não monossexuais como pansexualidade, polisssexualidade e sexualidade fluída” (JAEGER *et al.*, 2019, p. 03). Leão (2018), discorre que “são inúmeras as identidades que se enquadram no chamado ‘guarda-chuva bissexual’, a maior parte deles não é encontrada no cotidiano brasileiro, embora algumas venham ganhando projeção em determinados contextos como [...] o ativismo LGBT” (LEÃO, 2018, p. 30).

Entende-se assim que o escopo das sexualidades não mais comporta o “binarismo” de outrora, ao contrário, se alastra em diversas formas de expressão. Destarte, infere-se que o

conceito de sexualidade é amplo e geral, não se falando na atualidade em sexualidade, mas sim em sexualidades, podendo assim englobar, ou não, os aspectos relacionados ao gênero e à identidade de gênero.

É importante, ainda, não confundir estes dois últimos conceitos. Nesse sentido, identidade de gênero, conforme conceituam Gonçalves e Gonçalves (2021), são “as possibilidades dos seres humanos de reconhecerem a si como pertinentes, ou não, de determinado gênero ou de possuírem características femininas ou masculinas, sem que deixem de pertencer ao que se constitui por homem ou mulher” (GONÇALVES; GONÇALVES, 2021, p. 02).

Gênero, portanto, se relaciona com a construção social do indivíduo sobre si mesmo a partir de suas interações com os outros. Identidade de gênero tem relação com a expressão do gênero, da forma como o indivíduo se sente em relação ao próprio gênero e como externaliza esse sentimento. Feita essa conceituação, cabe apontar para o fato de que a discriminação de pessoas baseada nas sexualidades tem sido uma constante no Ocidente. Rocha, Carvalho Neto e Pio (2021) evidenciam que, ao longo da História, sobretudo por força da fé cristã, a homoafetividade tem sido associada a conceitos negativos, como pecado, distúrbio mental e crime. Zahed (2019) também afirma que, na tradição mulçumana, “aqueles descritos em árabe como *mukhanathun* (“efeminados”), são frequentemente associados com a homossexualidade” (ZAHED, 2019, p. 259), de modo que fora do eixo das religiões cristãs há também conceitos negativos relacionados aquilo que destoa da heteronormatividade.

Na atualidade, com a ascensão do debate sobre gênero e sexualidades, a discriminação baseada em gênero e sexualidades ganha novos contornos, desta vez associados justamente à estas novas discussões. Dentro destes preconceitos é possível conceituar também a LGBTfobia, entendida como “toda e qualquer conduta homofóbica ou transfóbica, real ou suposta que envolva aversão odiosa à orientação sexual ou à identidade de gênero de alguém, por traduzirem expressões de racismo, compreendido em sua dimensão social” (GONÇALVES *et al.*, 2020).

Diante disso, a homofobia e a bifobia contemporâneas, por exemplo, se voltam para as mais diversas manifestações das sexualidades que fujam ao padrão da heteronormatividade e ao binarismo. Nesse sentido é que Costa (2020) dispõe:

Observa-se que o padrão heterossexual de comportamento que se manifesta nas relações de poder não exclui socialmente apenas os sujeitos homossexuais, atingindo todos aqueles que não aderem à pseudonormalidade, ou seja, exclui e marginaliza as mulheres e homens bissexuais, os transgêneros e até mesmo mulheres e homens heterossexuais, bastando que estes não se enquadrem no modelo de conduta esperado ou não interpretem seus papéis de gênero (COSTA, 2020, p. 21).

O preconceito de identidade de gênero, surge, portanto, como uma manifestação da discriminação de gênero e sexual, e é extremamente prejudicial como externalização de tal tipo de discriminação.

Destarte, conforme se alastra a discussão sobre gênero e identidade de gênero, alastra-se também o preconceito. No contexto pátrio, conforme o Grupo Gay da Bahia (GGB), em 2020, 237 pessoas tiveram mortes violentas no Brasil, relacionadas à orientação sexual ou identidade de gênero (IBDFAM, 2021). Ainda segundo o mesmo grupo, uma pessoa LGBTQIA+ morre a cada 23 horas no Brasil (ARCOVERDE; SOUSA, 2019). Não obstante, estudo conduzido pelo GGB também evidenciou que o preconceito matou pelo menos cinco mil pessoas LGBTQIA+ em vinte anos (NICLEWICZ, 2021). A Fundação Getúlio Vargas, no estudo “A Violência LGBTQIA+ no Brasil”, publicado em dezembro de 2020 demonstrou que, durante a pandemia de Covid-19, 70 por cento da população LGBTQIA+ sofreu violência dentro de casa, cumprindo isolamento social com a família (GONÇALVES, 2020).

Tais dados evidenciam que o Brasil possui uma população LGBTfóbica. Nesse ínterim, a homofobia e os diversos sub preconceitos que dela decorrem encontram também, no meio digital e na Sociedade da Informação, ambiente propício para se propagarem. Rocha, Carvalho Neto e Pio (2021) afirmam que, “com o aumento de exposição, também aumenta os ataques de quem não tolera o diferente e vive submerso na homofobia virtual” (CARVALHO NETO; PIO, 2021, p. 02). É nesse contexto que se evidencia o problema da discriminação algorítmica (BORJA, 2021), uma mazela decorrente da Sociedade 4.0 e que tem na expansão da Inteligência Artificial, um de seus fatores.

3 INTELIGÊNCIA ARTIFICIAL: CONCEITOS INICIAIS

De início, cabe advertir: o conceito de inteligência não é pacificado, tampouco há consenso sobre se inteligência é algo uno e indivisível, ou se é algo diverso. Neste sentido é que Yuk Hui (2020) questiona: “de que modos é possível falar sobre inteligência artificial se considerarmos que esta é mais suscetível à mutações que a inteligência humana, cujo mecanismo estamos longe de compreender inteiramente?” (HUI, 2020). Diante disso, antes de adentrar na esfera da inteligência artificial, é necessário conceituar (ou pelo menos tentar conceituar) a inteligência humana.

Com efeito, Max Tegmark (2020), na obra Vida 3.0, define que inteligência é a capacidade de atingir objetivos complexos, reconhecendo, todavia, que seu conceito é limitado e assentindo que “nem entre os inteligentes pesquisadores de inteligência há consenso sobre o que é inteligência” (TEGMARK, 2020, p. 93). Teixeira e Cheliga (2021), partindo da análise de Ken Richardson sobre inteligência sob os aspectos genético e sócio-histórico, afirmam que inteligência é a “potencialidade descrita nos genes, que pode ou não ser ativada no meio, e constrói esquemas sócio-históricos de uma sociedade” (TEIXEIRA;

CHELIGA, 2021, p. 17).

Feita esta tentativa, ocorre que a Inteligência Artificial como se conhece hoje tem início com Alan Turing, pai da computação, em idos da primeira metade do séc. XX. Todavia, a ideia de máquinas inteligentes remonta a um passado bem mais distante, podendo sua arqueologia ser buscada até mesmo em Platão, e anteriormente a este, nos mitos (PERELMUTER, 2019).

Em Alan Turing, porém, tem início a abordagem contemporânea de Inteligência Artificial, através do seu famoso Teste de Turing, no qual uma máquina seria posta à prova de acordo com a capacidade de mimetizar um ser humano, e seria aprovada se, após um interrogador humano lhe fazer perguntas, este não fosse capaz de identificar se estava se comunicando com uma máquina ou com um outro ser humano (BOEING; ROSA, 2020).

Com o avançar da ciência da computação, as definições de Inteligência Artificial ganham novos contornos, e como previamente exposto, o conceito exato de Inteligência Artificial não é pacífico, como não é, também, a definição de inteligência humana. Assim, traz-se algumas das definições atualmente tidas como de respaldo na literatura, quais sejam as definições de Minsky (1985), Luger (2013), e Russell e Norvig (2013).

Para Minsky (1985), a Inteligência Artificial é definida como a ciência da construção de máquinas que fazem coisas que requerem inteligência, caso fossem feitas por homens. Luger (2013), por sua vez, conceitua que Inteligência Artificial é “o ramo da ciência da computação que se ocupa da automação do comportamento inteligente” (LUGER, 2013, p. 21). Russell e Norvig (2013), por conseguinte, ressaltam quatro dimensões para conceituar inteligência Artificial, sendo aquelas que se relacionam com processos de pensamento e raciocínio, comportamento, fidelidade ao desempenho humano e racionalidade.

Destaque-se que mesmo os citados autores reconhecem a insuficiência de suas definições, dada a amplitude do tema. Feita esta conceituação primária, é possível observar que muitos dos aspectos considerados pela literatura para conceituar Inteligência Artificial se relacionam com a experiência humana. É que a Inteligência Artificial se apoia na humana na medida em que a toma como base de definição. Teixeira e Cheliga (2021) dispõem que “a Inteligência Artificial conta com partes que simulam o cérebro humano, como os neurônios. Os neurônios, dentro de uma Inteligência Artificial, são capazes de simular o processamento de uma tarefa” (TEIXEIRA; CHELIGA, 2021, p. 31).

Nesse sentido é que funcionam as denominadas redes neurais artificiais, ou ANNs. Estas são algoritmos que simulam a estrutura do cérebro humano. Viegas (2020) leciona que “nas ANNs há ‘neurônios’, que possuem várias camadas e conexões com outros ‘neurônios’. Cada camada escolhe um recurso específico para aprender, como curvas e bordas no reconhecimento de uma imagem, por exemplo” (VIEGAS, 2020, p. 148). Destarte, a neurociência tem sido apontada como uma das bases para o entendimento sobre Inteligência Artificial, uma vez que é a partir da análise de como o cérebro humano

funciona, isto é, aprende, que a Inteligência Artificial se guia.

Há que se definir ainda que a Inteligência Artificial é, na atualidade, especializada e não geral. Isto significa dizer que o que se tem hoje são Inteligências Artificiais capazes de executar tarefas, em modelos específicos, de maneira melhor otimizada que um ser humano (TEIXEIRA; CHELIGA, 2021). Guy Perelmuter (2019) afirma que “quanto mais previsível e estável for uma tarefa, maior a chance de um robô ser capaz de desempenhá-la” (PERELMUTER, 2019, p. 65).

Nesse panorama, as Inteligências Artificiais (IAs) atuais são capazes de realizar processos racionais, especializados e lógicos, mas não alcançaram ainda o patamar da denominada “singularidade”, isto é, a capacidade de pensar por si mesmas. Não é por outro motivo que Tegmark (2020) afirma que “o santo graal da pesquisa em IA é criar uma ‘IA geral’, que seja extremamente ampla: capaz de realizar praticamente qualquer objetivo, inclusive o aprendizado” (TEGMARK, 2020, p. 95).

Nesse sentido, embora não se tenha alcançado ainda uma Inteligência Artificial Geral, o atual estágio de evolução das IAs demonstra que estas encontram aplicabilidade em praticamente todos os âmbitos atuais da vida humana no contexto globalizado.

4 MACHINE LEARNING, DEEP LEARNING E A FORMA COMO AS INTELIGÊNCIAS ARTIFICIAIS APRENDEM

Para entender como uma IA aprende, é necessário compreender o que é um algoritmo. Assim, um dos conceitos mais completos atualmente aceitos do termo é dado por Thomas H. Cormen, que define que “um algoritmo é qualquer procedimento computacional bem definido que toma algum valor ou conjunto de valores como entrada e produz algum valor ou conjunto de valores como saída” (CORMEN *et al.*, 2002). Perelmuter (2019) conceitua que algoritmo é o nome dado ao conjunto de instruções seguido por um computador para desempenhar determinada tarefa. É possível afirmar, portanto, que algoritmos resolvem fórmulas.

Compreendido o que é um algoritmo, ocorre que é através destes que se dá o *machine learning*, ou aprendizado de máquina. Esse termo *machine learning* foi cunhado por Arthur Samuel, que o conceituou como a capacidade de aprender sem ter sido explicitamente programado (BOEING; ROSA, 2020). Teixeira e Cheliga (2021) afirmam que o aprendizado de máquina é o que faz uma Inteligência Artificial ser inteligente, o processo pelo qual ela aprende com os erros, através de falseamento, construindo conhecimento até chegar numa resposta satisfatória por meio dos algoritmos.

É pelo aprendizado de máquina, portanto, que a IA adquire suas capacidades e viabiliza o exercício das funções para as quais foi programada. Wolkart e Becker (2020) definem que “o *machine learning* opera de forma probabilística e seus algoritmos são auto

programáveis, isto é, aprendem por conta própria, sem a necessidade de programação prévia e explícita” (WOLKART; BECKER, 2020, p. 09). A ideia de aprendizado de máquina, embora tenha sua aplicabilidade nos tempos atuais como nunca antes experimentado, não é um conceito novo. Rover (2021) afirma:

O *machine learning* é uma técnica de longa data, em uso nos sistemas computacionais desde há muito. O que é novidade é o fato de hoje ser possível o acesso a dados massivos, estruturados e não estruturados, e sobre eles interpretar, modelar e fazer previsões automáticas, possibilitando assim ao sistema computacional “aprender” de forma independente de sua programação (ROVER, 2021, p.22).

Entende-se que o aprendizado de máquina, no contexto atual, encontrou o ambiente ideal para se desenvolver, dada a extensa quantidade de dados que hoje se troca e se coloca na rede mundial diariamente.

Mas como funciona o aprendizado de máquina? Perelmuter (2019) explica que, “em *machine learning*, procura-se apresentar aos algoritmos a maior quantidade possível de dados, permitindo que os sistemas desenvolvam a capacidade de tirar conclusões automaticamente” (PERELMUTER, 2019, p. 83). Wolkart e Becker (2020), afirmam que a técnica (do *deep learning*) funciona a depender do *input* de dados que funcionam como combustível para a sua aprendizagem, os dados alimentam a máquina, subsidiariamente o algoritmo analisa-os, e daí floresce o resultado esperado (*output*).

Desta forma, faz-se necessário expor a IA a um *dataset*, ou banco de dados, que servirá de “alimento” ao sistema, de modo que este possa, através da análise algorítmica, produzir uma saída ou resultado. Cabe ressaltar que, em *machine learning*, a IA não apenas adquire o “conhecimento” a partir dos inputs, mas aprimora tal conhecimento. Lage e Peixoto (2021) apontam que:

O *machine learning* é uma técnica que permite que os sistemas de computação internalizem e façam previsões baseadas em dados históricos. O processo é alimentado por um algoritmo de aprendizado de máquina, uma função capaz de melhorar seu desempenho com o tempo, treinando-se por meio de métodos de análise de dados e modelagem analítica (LAGE; PEIXOTO, 2021, p. 267).

Diante disso, através de um processo de efetivo aprendizado algorítmico, a IA evolui o tratamento de dados, tornando-se mais eficiente e “inteligente” conforme o fluxo de dados que a alimenta. Deve-se apontar, nesse contexto, também para o *deep learning*, ou aprendizado profundo. Este é definido como:

Um subdomínio do aprendizado de máquina, que consiste em múltiplas camadas, [...] conhecida como rede neural, permitindo que um sistema se treine usando dados históricos, reconhecendo padrões e fazendo interferências probabilísticas (LAGE; PEIXOTO, 2021, p. 270).

O *deep learning* busca se aproximar ainda mais da forma de pensar do ser humano, fazendo uso das já referidas redes neurais artificiais (ANNs), que segundo Wolkart e Becker (2020), “têm modelo de aprendizado baseado no cérebro humano, [...] baseado em erros e acertos, com reforço paulatino dos caminhos e decisões mais corretas para atingir determinados objetivos” (WOLKART; BECKER, 2020, p. 10). Dessa maneira, as redes neurais artificiais são a expressão do denominado aprendizado profundo, ou *deep learning*, representando um novo paradigma no aprendizado de máquina. Todavia, Perelmuter (2019) adverte que:

É fundamental lembrar que sua decisão (do *deep learning*) é altamente dependente da qualidade e do tipo de dados que utilizam durante a fase de aprendizado. E esse é um dos maiores fatores de risco para o uso dessa tecnologia: se o treinamento não for feito cuidadosamente, os resultados podem ser perigosos (PERELMUTER, 2019, p. 85).

Neste sentido, aponta-se para um dos maiores percalços no que tange às IAs, quais seja o problema do viés.

5 O VIÉS NO APRENDIZADO DE MÁQUINA (OU COMO INTELIGÊNCIAS ARTIFICIAIS SE TORNAM PRECONCEITUOSAS)

Antes de tudo cabe esclarecer que, nesta pesquisa, ao se falar que uma máquina se torna preconceituosa, busca-se usar a expressão com mero caráter explicativo. Não se está afirmando que uma IA, de fato, adquiriu consciência de preconceito. Como Waldman (2020) dispõe, “a IA não tem comportamento e ação características humanas [...], suas ações e comportamentos não podem retroagir a nenhuma interioridade, a nenhuma racionalidade, a nenhuma essencialidade” (WALDMAN, 2020, p. 125).

Ainda segundo Cruz *et al.* (2021), “o pré-julgamento ou como o nome indica, preconceito, está relacionado ao julgamento negativo sobre um determinado membro de um grupo social (etnia, religião, moral ou sexual)” (CRUZ *et al.*, 2021, p. 18). Assim, dado o fato de que um sistema, por mais inteligente que seja, não é capaz de julgar efetivamente, uma IA não se torna de fato preconceituosa, nem adquire consciência de preconceitos. Discriminar a própria espécie é uma execrável característica própria do ser humano.

Todavia, pelo fato de a IA ser hoje uma ferramenta de considerável aplicação na

sociedade globalizada, presente em praticamente todos os contextos da vida atual e cujas dimensões as previsões científicas indicam que somente irão se expandir (TEGMARK, 2020), há que se apontar para os riscos sociais que existem quando uma IA é construída eivada de vieses preconceituosos. Segundo Coutinho (2021), “o uso de algoritmos nunca estará imune aos valores humanos de seus programadores, sendo que eventuais resultados indesejáveis dizem respeito ao caráter enviesado e discriminatório da própria sociedade” (COUTINHO, 2021, p. 146).

Diante disso, consta que o viés em IAs trata-se justamente do problema da subjetividade humana interferindo na alimentação (*inputs*) no processo de aprendizado de máquina, o que pode deixar os algoritmos parciais e viciados, sobretudo por conterem os vícios dos seus criadores (VIEGAS, 2021). Borja (2021) afirma que:

Os algoritmos são um produto humano, de modo que sua programação pode estar entrelaçada à visão de mundo dos seus programadores, assim como seu próprio funcionamento pode servir como um repetidor ou amplificador de preconceitos e assimetrias, a exemplo da desigualdade de gênero (BORJA, 2021).

Identifica-se nesse prisma, que um dos pilares do problema do viés em aprendizado de máquina reside na subjetividade do programador ou da equipe de programadores que constrói os algoritmos e dita a forma como estes analisam os dados que irão alimentar a IA. Cathy O’Neil, em sua obra *Algoritmos de Destruição em Massa (Weapons of Math Destruction)*, apontou para os danos catastróficos que o viés em aprendizado de máquina pode gerar, denominando os sistemas inteligentes enviesados como Armas de Destruição Matemática, ou ADMs (O’NEIL, 2020). A autora dispõe:

As aplicações matemáticas fomentando a economia de dados são baseadas em escolhas feitas por humanos falíveis. Algumas dessas escolhas sem dúvida são feitas com as melhores das intenções. Mesmo assim, muitos desses modelos programam preconceitos, equívocos e vieses humanos nos sistemas de software que cada vez mais gerem nossas vidas (O’NEIL, 2020, p. 06).

Ocorre, então, que os sistemas inteligentes podem já nascer enviesados, em decorrência das subjetividades de seus programadores. É o programador quem seleciona quais dados serão usados e de que forma eles vão ser processados, e neste processo, coloca suas próprias convicções, determina subjetivamente quais são os dados considerados importantes e quais não são, sendo uma escolha de cunho enviesado desde o nascimento do algoritmo (O’NEIL, 2020).

Os riscos práticos da construção de IAs com viés preconceituoso são evidentes: as Inteligências Artificiais estão presentes em praticamente todo sistema que efetua troca de dados, desde smartphones até eletrodomésticos, redes sociais, bancos digitais, assistentes virtuais e câmeras de vigilância (ITFORUM, 2022). Silva (2020) também afirma que, “crescentemente, os algoritmos e inteligência artificial ganham destaque no debate sobre comunicação, sociedade e tecnologia por regirem cada vez mais esferas da vida” (SILVA, 2020, p. 431), o que ressalta a importância do tema em comento, dada a extensão do uso de IAs na atualidade em diversos aspectos da vida humana.

É perceptível, portanto, que o enviesamento de sistemas inteligentes baseados em *machine learning* produz efeitos no mundo real em níveis impactantes, e ainda é um problema a ser superado, levantando consigo questionamentos sobre imparcialidade, neutralidade e transparência na construção de IAs. Carvalho (2021) pontua que, “quando um algoritmo de aprendizado de máquina é aplicado a um conjunto de dados, ele busca por padrões nos dados e reforça-os na criação de um modelo que represente os dados” (CARVALHO, 2021, p. 29). Neste sentido, o problema do viés é um paradigma a ser discutido, dado o potencial que possui de causar danos diversos a minorias e grupos menos favorecidos e socialmente excluídos.

6 APRENDIZADO DE MÁQUINA E IDENTIDADE DE GÊNERO: ENSINANDO UMA IA A DISCRIMINAR EM UMA SOCIEDADE DISCRIMINATÓRIA

Uma vez apontada a subjetividade dos programadores como um dos pilares que sustentam o problema do viés no aprendizado de máquina, consta que o outro pilar é o *dataset*, isto é, o banco de dados de inputs de uma IA. Carvalho (2021) afirma que “uma das principais fontes de preconceito está no conjunto de dados utilizados para gerar modelos, como é o caso das aplicações que geram modelos por meio da aplicação de algoritmos de aprendizado de máquina aos dados” (CARVALHO, 2021, p. 29). Diante disso, quando há viés em um dataset, haverá provavelmente viés em um sistema inteligente.

Coutinho (2021) avalia que “algoritmos de *machine learning* são capazes de gerar resultados discriminatórios em relação a gênero e raça se alimentados com inputs que não deflagram a diversidade verdadeira que existe” (COUTINHO, 2021, p. 147). Assim, o banco de dados que alimenta uma IA é a sua realidade, a realidade que os algoritmos irão processar. A depender de quais sejam estes dados que a IA está recebendo como alimentação, esta realidade já pode, por si só, estar enviesada, e conseqüentemente irá gerar uma IA discriminatória (BOEING; ROSA, 2020).

Foi o que ocorreu com a Inteligência Artificial Tay, um *chatbot* criado pela *Microsoft*, que em menos de 24 horas exposta aos usuários de redes sociais, tornou-se racista e foi desativada:

O robô, que deveria se tornar mais esperto e perspicaz ao conversar com os humanos, rapidamente passou a reproduzir o racismo e a ignorância dos trolls da internet. Instigada por usuários no Twitter, Tay usou termos impubescíveis para se referir a negros e mulheres, declarou suporte ao genocídio e demonstrou apoio à causa dos supremacistas brancos (VEJA, 2016).

Ainda a título de exemplificação, a IA do aplicativo Google Fotos, da empresa global Google, em meados de 2015 também apresentava viés racista relacionado à cor da pele:

A IA do Google Fotos etiquetava todos os seus amigos negros como gorilas, simplesmente, por não conseguir distinguir a diferença da pele dos seres humanos e a de macacos. O erro foi corrigido, mas demonstrou como os algoritmos refletem o viés e o preconceito dos seres humanos que estão presentes nas bases de dados utilizadas pelos programadores (ROSSETTI; ANGELUCCI, 2021, p. 10).

Diante disso, é certo que os danos colaterais causados por IAs preconceituosas podem alcançar proporções desastrosas. No panorama brasileiro, em que, como supracitado tem-se uma sociedade predominantemente homofóbica (GONÇALVES, 2020), surge o contexto ideal para a criação de uma IA igualmente homofóbica, diante da quantidade massiva de inputs discriminatórios que é possível auferir ao se alimentar um sistema inteligente com os dados do povo brasileiro. Neste sentido, como aponta Fornasier (2020):

Pesquisadores de Stanford já desenvolveram uma IA que usa fotografias para identificar a orientação sexual de uma pessoa; tem uma taxa de precisão preditiva acima de 90 por cento. Isso pode ser muito útil para mais de 70 países que ainda criminalizam a homossexualidade (FORNASIER, p. 173).

É evidente que a IA pode ser, desse modo, programada propositalmente para ser preconceituosa. Não se pode olvidar que a IA é uma ferramenta, e que enquanto ferramenta, pode ser usada tanto para fins legítimos quanto para fins maliciosos (FORNASIER, 2020). Diante disso, ao se tomar como *dataset*, dados e informações enviesadas para ensinar a IA a identificar pessoas LGBTQIA+ por este meio, tal modelo inteligente não apenas irá identificar pessoas LGBTQIA+, mas irá identificá-las de maneira preconceituosa, e dessa forma, ajudará a perpetuar um preconceito já existente.

Inobstante, os agentes artificiais tomam decisões acerca de acesso e visibilidade, os quais podem ser influenciados por vieses de gênero e de outros tipos, gerando aquilo que Silva denomina “opressão algorítmica” (SILVA, 2020, p. 431). Tal quadro ameaça os direitos e garantias fundamentais em diversos níveis, sobretudo de grupos vulneráveis.

Como Santos (2021) aponta, dados e informações acerca de gêneros, sexualidades, raças e outros, “podem servir como insumos para classificações de indivíduos através da elaboração de códigos que incorporem determinados valores sociais, julgamentos ideológicos ou critérios mercadológicos” (SANTOS, 2021, p. 24).

No que toca ao preconceito relacionado à identidade de gênero e o enviesamento de IAs, o caso brasileiro mais famoso é aquele relacionado ao “Kit Gay”, uma campanha de desinformação ocorrida nas redes sociais e que serviu de motor para minar a reputação de grupos minoritários nas eleições presidenciais brasileiras de 2018 (SANTOS, 2021). No caso em tela, a temática da ideologia de gênero foi utilizada por grupos sectários e homofóbicos que, por meio de algoritmos enviesados, manipularam ideias preconceituosas sobre identidade de gênero (BARAJAS, 2021).

A campanha alcançou níveis de discriminação algorítmica nunca antes observados no contexto pátrio, de modo que a própria corporação responsável pelo aplicativo *Whatsapp* admitiu que a eleição brasileira de 2018 teve uso de envios maciços de mensagens, com sistemas automatizados contratados por empresas (SANTOS, 2021). O que se evidencia, portanto, é que a preocupação aqui exposta não somente pode acontecer, como efetivamente já aconteceu, isto é, IAs enviesadas já foram, de fato, utilizadas para lesar direitos e garantias fundamentais de grupos LGBTQIA+ baseadas em viés discriminatório relativo à identidade de gênero (BARAJAS, 2021).

Diante disso, o panorama de exclusão e discriminação baseada em identidade de gênero que há em uma sociedade como a brasileira, através do viés aplicado no aprendizado de máquina para construir uma IA e da manipulação de algoritmos, pode se manter ou mesmo se expandir, num círculo vicioso em que exclusão gera cada vez mais exclusão. O’Neil (2020) acena para tal círculo retroalimentativo, informando que ao invés de o sistema buscar pela verdade, ele apenas perpetua o status quo, definindo a sua própria realidade usando seus próprios resultados para justificar um modelo destrutivo.

À vista disso, preconceitos são conservados, perdurando através de um sistema de exclusão que, pelo caráter enviesado, não busca ser mais igualitário, mas sim justifica sua realidade com mais desigualdade, a realidade que foi aprendida através do *machine learning* e dos dados que lhe alimentaram.

7 É POSSÍVEL SUPERAR O PROBLEMA DO VIÉS PRECONCEITUOSO NO APRENDIZADO DE MÁQUINA?

De modo a assegurar um futuro/presente mais isonômico, e a construção de sociedades cada vez menos discriminatórias, há que se apontar para as possíveis soluções para o problema do enviesamento de sistemas inteligentes. Coutinho (2021) defere que:

Apesar da expectativa de que a tecnologia venha a melhorar a vida das pessoas, não faltam dúvidas acerca dos limites que devam ser estipulados para o uso da IA, considerando a desconfiança em relação à segurança e robustez dos sistemas e a possibilidade de uso inescrupuloso da ferramenta (COUTINHO, 2021, p. 143).

Isto posto, faz-se necessário pensar em formas de superar os problemas relacionados à IA, uma vez que os fatores negativos que a envolvem são evidentes e possuem potencial prejudicial de larga escala, dado o contexto globalizado das sociedades humanas atuais. Perelmuter (2019) afirma que:

Se de fato os chamados sistemas inteligentes continuarem sua expansão para diversas áreas e negócios, serviços e governos, será imprescindível que não sejam contaminados pelos preconceitos que seres humanos acabam desenvolvendo, consciente ou inconscientemente (PERELMUTER, 2019, p. 85).

É certo, portanto, que tal problema deve ser levado a sério, e neste panorama muito se tem discutido sobre ética e regulamentação no uso das IAs (CARVALHO, 2021). Lima, Pessoa e Cleger (2019) afirmam que “a possibilidade de criar máquinas inteligentes levanta uma série de questões éticas das quais pouco são discutidas e difundidas no meio tecnológico e social” (LIMA; PESSOA, CLEGER, 2019, p. 02). A ética em IA tem sido assim, tratada como um novo ramo de estudo relacionando a ética com a evolução tecnológica.

Nesse contexto, Hartmann e Piaia (2021) destacam que o Fórum Econômico Mundial, em 2016, elencou várias questões éticas para a inteligência artificial, entre elas o racismo robótico, definido pela pergunta “como eliminar preconceitos e parcialidades causados pela inteligência artificial?” (HARTMANN; PIAIA, 2021, p. 97). Delineia-se, desse modo, uma preocupação crescente no contexto global quanto à questão do uso de IAs, sobretudo em relação à questão do enviesamento algorítmico de modelos inteligentes.

A Conferência Internacional de Proteção de Dados, em 2018, definiu alguns princípios para o uso e desenvolvimento das IAs, entre estes os princípios da proteção e da representação, sendo que o primeiro é voltado a mitigar eventuais efeitos negativos provocados em decorrência de gêneros, sexualidades, raças, etnias e origens ou inclinações políticas, e o segundo busca se opor a preconceitos e assegurar a paridade de representação na rede (HARTMANN; PIAIA, 2021).

A exposição destes princípios demonstra que há, por parte dos atores globais, um aumento da preocupação sobre o uso ético de sistemas inteligentes, e ao elencar tais diretrizes, percebe-se uma tentativa de superar os problemas sociais que claramente são trazidos pela evolução tecnológica (ROSSETTI; ANGELUCI, 2021). A regulação das IAs pela Lei também tem sido tratada como fator essencial para superar as questões prejudiciais do

uso dos modelos inteligentes. Carvalho (2021) destaca que na Inglaterra, em 2017, a Câmara dos Lordes criou um comitê de IA para estudar questões sobre a IA, incluindo economia, avanço social e implicações éticas.

No Brasil, há atualmente em curso uma Comissão de Juristas junto ao Senado Federal, a CJSUBIA, responsável por elaborar a regulamentação legal brasileira sobre uso de Inteligência Artificial (BRASIL, 2022). Não obstante, há atualmente três projetos de Lei em curso no país para regulamentar a IA, o PL nº 5051/2019, o PL nº 21/2020 e o PL nº 872/2021 (BRASIL, 2022). Com efeito, a regulamentação legal é fator de preponderância, e a quantidade de projetos de lei em curso demonstram a importância do tema.

A responsabilidade dos programadores também é fator a ser observado como de sumária importância para superar o problema do enviesamento em *machine learning*. Nesse sentido, Lage e Peixoto (2021) afirmam que “há o aprendizado de máquina supervisionado e o aprendizado de máquina não supervisionado” (LAGE; PEIXOTO, 2021, p. 270). Wolkart e Becker (2020) definem o aprendizado de máquina supervisionado como aquele que é alimentado com dados lapidados e previamente escolhidos pelos seres humanos. Desta forma, o modelo, durante o próprio aprendizado, vai ajustando suas variáveis para mapear as entradas (*inputs*) e chegar à saída correspondente (*output*).

Nesse panorama, é possível apontar para o uso do aprendizado de máquina supervisionado como técnica para suplantar o viés em IAs. Perelmuter (2019) também propõe a colaboração entre máquinas e seres humanos a fim de decidir sobre temas com nuances e complexidades ainda não compreendidas pelos algoritmos (PERELMUTER, 2019). O programador, nessa proposta, é figura central do processo de *machine learning*, e deve, como tal, se esquivar ao máximo de lançar suas subjetividades no processo de aprendizado de máquina e de construção dos algoritmos, a fim de evitar o enviesamento do sistema, assumindo assim responsabilidade pela construção de uma sociedade menos discriminatória.

8 CONSIDERAÇÕES FINAIS

O *machine learning* tem sido largamente utilizado na construção de novas tecnologias no contexto contemporâneo, diante da quantidade imensa de dados à disposição, e tem gerado inovações surpreendentes, entre estas o avanço das Inteligências Artificiais. Todavia, observou-se que embora novas e brilhantes possibilidades surjam com a evolução das IAs, nascem também problemas de ordem social e ética, entre estes o preconceito algorítmico.

Analisou-se aqui como as IAs são construídas e como o problema do viés em *machine learning* pode ser destrutivo para grupos minoritários e vulneráveis. Identificou-se que os problemas decorrentes do viés preconceituoso em IAs não são meras possibilidades, mas que evidentemente já causaram efeitos negativos concretos no mundo real. No caso do

preconceito decorrente de identidade de gênero, demonstrou-se que o viés algorítmico tem causado danos a grupos LGBTQIA+ em diversos lugares pelo mundo, entre estes o Brasil.

Concluiu-se que as IAs, quando expostas às sociedades com altos níveis de preconceito, são capazes de perpetuar e mesmo alavancar a discriminação baseada em identidade de gênero, bem como outros tipos de preconceitos que ali estejam instalados, por meio do viés algorítmico.

Faz-se necessário um maior rigor quanto à ética em IA, e esta tem sido uma preocupação de grupos, instituições e governos pelo mundo, que cada vez mais se mobilizam para estabelecer standards para o uso ético das IAs. É preciso ainda que sejam implementadas medidas de regulamentação legal, dada a urgência do tema e o potencial lesivo do viés algorítmico para a sociedade.

Novos estudos também se mostram necessários para aprofundar o tema e ressaltar junto à Academia e à sociedade a importância deste debate, de modo a conduzir a evolução tecnológica pelo caminho da igualdade e do respeito às diferenças.

REFERÊNCIAS BIBLIOGRÁFICAS

ARCOVERDE, Léo; SOUSA, Viviane. Brasil registra uma morte por homofobia a cada 23 horas, aponta entidade LGBT. **G1**: São Paulo, 2019. Disponível em:

<https://g1.globo.com/sp/sao-paulo/noticia/2019/05/17/brasil-registra-uma-morte-por-homofobia-a-cada-23-horas-aponta-entidade-lgbt.ghtml>. Acesso em: 15 maio 2022.

BARAJAS, Karina Bárcenas. Infraestructuras algorítmicas de la “ideología de género” y sus procesos de desinformación: una mirada hacia América Latina desde Brasil. *In*: BARAJAS, Karina Bárcenas. DELGADO-MOLINA, Cecilia. **Religión, género y sexualidad: entre movimientos e instituciones**. Ciudad de México: Universidad Nacional Autónoma de México, Instituto de Investigaciones Sociales, 2021. Disponível em:

https://www.researchgate.net/profile/Karina-Barcenas-Barajas/publication/357575826_Infraestructuras_algoritmicas_de_la_ideologia_de_genero_y_sus_procesos_de_desinformacion_una_mirada_hacia_America_Latina_desde_Brasil/links/61d4faafda5d105e551cecc9/Infraestructuras-algoritmicas-de-la-ideologia-de-genero-y-sus-procesos-de-desinformacion-una-mirada-hacia-America-Latina-desde-Brasil.pdf. Acesso em: 15 maio 2022.

BOEING, Daniel Henrique Arruda. ROSA, Alexandre Morais da. **Ensinando um robô a julgar**: pragmática, discricionariedade, heurísticas e vieses no uso de aprendizado de máquina no Judiciário. Florianópolis: Emais Academia, 2020.

BORJA, Letícia. **Vieses de gênero em algoritmos opaquos**: Um olhar sobre a regulação jurídica da Inteligência Artificial. SBC Horizontes, jun., 2021. Disponível em:

<http://horizontes.sbc.org.br/index.php/2022/01/vieses-de-genero-em-algoritmos-opaque-box-um-olhar-sobre-a-regulacao-juridica-da-inteligencia-artificial/>. Acesso em: 15 maio 2022.

BRASIL, Câmara dos Deputados. **CJSUBIA**: Comissão de juristas responsável por subsidiar elaboração do substitutivo sobre inteligência artificial no Brasil. Brasília, DF: 2022.

Disponível em: <https://legis.senado.leg.br/comissoes/comissao?codcol=2504>. Acesso em: 15 maio 2022.

CARVALHO, A. C. P. L. F. Inteligência Artificial: riscos, benefícios e uso responsável. **Estudos Avançados**, v. 35, n. 101, 2021. Disponível em: <https://www.scielo.br/j/ea/a/ZnKyrclrLVqzhZbXGgXTwDtn/?lang=pt&format=pdf>. Acesso em: 15 maio 2022.

CHELIGA, Vinicius. TEIXEIRA, Tarcisio. **Inteligência Artificial: aspectos jurídicos**. 3 ed. Salvador: JusPodivm, 2021.

COSTA, Diego Carneiro. **O viés do algoritmo e a discriminação por motivos relacionados à sexualidade**. 2020. Dissertação (Mestrado) – Universidade Federal da Bahia, Salvador, 2020. Disponível em: <https://repositorio.ufba.br/bitstream/ri/34394/1/DISSERTA%3%87%3%830-%20Diego%20Carneiro%20Costa.pdf>. Acesso em: 15 maio 2022.

CORMEN, T. H. Et al. **Algoritmos: Teoria e Prática**. Rio de Janeiro: Elsevier, 2002.

COUTINHO, Marina de Alencar Araripe. Considerações sobre Inteligência Artificial e tomada de decisão. *In*: PEIXOTO, Fabiano Hartmann (org). **Inteligência Artificial: Estudos de Inteligência Artificial**. Curitiba: Alteridade, 2021.

CRUZ, Luan Chagas. Et al. Vulnerabilidade da população trans no acesso à atenção à saúde: contexto pandêmico. *In*: SIQUEIRA, L. F. S. SILVA, M. C. O (org). **Diálogos contemporâneos: gênero e sexualidade na pandemia**. São Luís: Editora Expressão Feminista, 2021. Disponível em: https://www.researchgate.net/publication/351038099_DIALOGOS_CONTEMPORANEOS_GENERO_E_SEXUALIDADE_NA_PANDEMIA_wwweditoraexpressaofeministacombr. Acesso em: 15 maio 2022.

FORNASIER, Mateus de Oliveira. Inteligência artificial e democracia: oportunidades e desafio. **Revista do Programa de Pós-Graduação em Direito da UFC**, v. 40.1, jan. – jun., 2020. Disponível em: https://repositorio.ufc.br/bitstream/riufc/62264/1/2021_art_mofornasier.pdf. Acesso em: 15 maio 2022.

GONÇALVES, Alice Calixto. **A violência LGBTQIA+ no Brasil**. São Paulo: FGV Direito SP, 2020. Disponível em: <https://bibliotecadigital.fgv.br/dspace/bitstream/handle/10438/29886/A%20Viol%3%aancia%20LGBTQIA%2b%20no%20Brasil.pdf?sequence=1&isAllowed=y>. Acesso em: 15 maio 2022.

GONÇALVES, Marllon Caceres. GONÇALVES, Josiane Peres. Gênero, identidade de gênero e sexualidade: conceitos e determinações em contexto social. **Revista Ciências Humanas - Educação e Desenvolvimento Humano**, v. 14, e. 25, 2021. Disponível em: <https://www.rchunitau.com.br/index.php/rch/article/view/600/382>. Acesso em: 15 maio 2022.

HARTMANN, Gabriel Henrique. PIAIA, Thami Covatti. Ética e inteligência artificial: desafios e perspectivas à web semântica. *In*: PEIXOTO, Fabiano Hartmann (org). **Inteligência Artificial: Estudos de Inteligência Artificial**. Curitiba: Alteridade, 2021.

HUI, Yuk. **Tecnodiversidade**. São Paulo: Ubu Editora, 2020.

IBDFAM. **Dia Internacional contra a Homofobia**: Brasil registrou 237 mortes violentas de LGBTI em 2020. IBDFAM: 2020. Disponível em: <https://ibdfam.org.br/noticias/8488/Dia+Internacional+contra+a+Homofobia%3A+Brasil+registrou+237+mortes+violentas+de+LGBTI+em+2020>. Acesso em: 15 mai 2022.

ITFORUM. **Uma inteligência artificial sem ética pode arruinar a sociedade**. ITForum: 2022. Disponível em: <https://itforum.com.br/coluna/uma-inteligencia-artificial-sem-etica-pode-arruinar-a-sociedade/>. Acesso em: 15 maio 2022.

JAEGER, Melissa Bittencourt *et al.* Bissexualidade, bifobia e monossexismo: problematizando enquadramentos. **Revista Periódicus**, n. 11, v. 2, maio - out., 2019. Disponível em: <https://periodicos.ufba.br/index.php/revistaperiodicus/article/view/28011/20100>. Acesso em: 10 jul. 2022.

LAGE, Fernanda de Carvalho. PEIXOTO, Fabiano Hartmann. Inteligência Artificial e Direito: desafios para a regulação do uso da Inteligência Artificial. *In*: PEIXOTO, Fabiano Hartmann (org). **Inteligência Artificial: Estudos de Inteligência Artificial**. Curitiba: Alteridade, 2021.

LEÃO, Maria. **Os unicórnios no fim do arco-íris: bissexualidade feminina, identidades e política no Seminário Nacional de Lésbicas e Mulheres Bissexuais**. 2018. Dissertação (Mestrado) – Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2018. Disponível em: <https://www.bdt.d.uerj.br:8443/bitstream/1/4449/1/Dissertacao%20Maria%20Leao.pdf>. Acesso em: 10 jul. 2022.

LIMA, Jéssica Luana Oliveira de. PESSOA, Marcela. CLEGER, Sergio. Viés em Aprendizagem de Máquina: como a inteligência Artificial pode prejudicar as minorias. *In*: **Encontro regional de computação e sistemas de informação**, 8, 2019, Manaus. Anais [...] Manaus: 2019. Disponível em: https://www.researchgate.net/publication/339064294_Vies_em_Aprendizagem_de_Maquina_como_a_inteligencia_Artificial_pode_prejudicar_as_minorias. Acesso em: 15 maio 2022.

LUGER, George F. **Inteligência Artificial**. 6. Ed. São Paulo: Pearson Education do Brasil, 2013.

MINSKY, Marvin. **A Sociedade da Mente**. São Paulo: Editora Francisco Alves, 1985.

NICLEWICZ, Manuella. Preconceito matou mais de 5 mil LGBTQIA+ em 20 anos, diz estudo. **CNN**: São Paulo, 2021. Disponível em: <https://www.cnnbrasil.com.br/internacional/preconceito-matou-mais-de-5-mil-lgbtqia-em-20-anos-diz-estudo/>. Acesso em: 15 maio 2022.

OMS. **Saúde sexual, direitos humanos e a lei**. Porto Alegre: UFRGS, 2020. Disponível em: <https://apps.who.int/iris/bitstream/handle/10665/175556/9786586232363-por.pdf?ua=1>. Acesso em: 15 maio 2022.

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

OYĚWŪMÍ, Oyèrónké. **Conceituando o gênero: os fundamentos eurocêntricos dos conceitos feministas e o desafio das epistemologias africanas**. CODESRIA Gender Series, v. 1, Dakar, CODESRIA, 2004. Disponível em: https://filosofia-africana.weebly.com/uploads/1/3/2/1/13213792/oy%C3%A8r%C3%B3nk%C3%A9_oy%C4%Bw%C3%B9m%C3%AD_-_conceitualizando_o_g%C3%AAnero_os_fundamentos_euroc%C3%AAntrico_dos_conceitos_feministas_e_o_desafio_das_epistemologias_africanas.pdf. Acesso em: 10 jul. 2022

PERELMUTER, Guy. **Futuro presente: o mundo movido à tecnologia**. Barueri: Companhia Editora Nacional, 2019.

ROCHA, Thaynara Ferreira. CARVALHO NETO, Emanuel de Jesus. PIO, Marco Aurélio de. A (des)construção social da homofobia e os efeitos da pandemia da covid-19 na comunidade LGBTQIA+. *In*: SIQUEIRA, L. F. S. SILVA, M. C. O (org). **Diálogos contemporâneos: gênero e**

sexualidade na pandemia. São Luís: Editora Expressão Feminista, 2021. Disponível em: https://www.researchgate.net/publication/351038099_DIALOGOS_CONTEMPORANEOS_GENERO_E_SEXUALIDADE_NA_PANDEMIA_wwweditoraexpressaofeministacombr. Acesso em: 15 maio 2022.

ROSSETTI, Regina. ANGELUCI, Alan. Ética Algorítmica: questões e desafios éticos do avanço tecnológico da sociedade da informação. **Galáxia**, n. 46, 2021.

ROVER, Aires José. Machine Learning no Poder Judiciário: uma biblioteca temática, ano 2020. *In*: PEIXOTO, Fabiano Hartmann. **Inteligência Artificial: Estudos de Inteligência Artificial**. Curitiba: Alteridade, 2021.

RUSSELL, Stuart J. NORVIG, Peter. **Inteligência Artificial**. Rio de Janeiro: Elsevier, 2013.

SANTOS, Cintia Regina Ribeiro dos. Sentidos minerados: o discurso do kit gay e a circulação algorítmica no controle do gênero. **Leitura**, Maceió, n. 71, set. – dez., 2021. Disponível em: <https://www.seer.ufal.br/index.php/revistaleitura/article/view/12420/9404>. Acesso em: 15 maio 2022.

SILVA, Tarcízio da. Visão computacional e racismo algorítmico: branquitude e opacidade no aprendizado de máquina. **Revista da ABPN**, v. 12, n. 31, dez. – fev., 2020. Disponível em: https://www.researchgate.net/publication/339514173_Visao_Computacional_e_Racismo_Algoritmico_Branquitude_e_Opacidade_no_Aprendizado_de_Maquina. Acesso em: 15 maio 2022.

TEGMARK, Max. **Vida 3.0: o ser humano na era da inteligência artificial**. São Paulo: Benvirá, 2020.

VIEGAS, Claudia Mara de Almeida Rabelo. Inteligência Artificial: uma análise da sua aplicação no Judiciário Brasileiro. *In*: ALVES, Isabella Fonseca (org.). **Inteligência Artificial e Processo**. São Paulo: D'Plácido, 2020.

WALDMAN, Ricardo. **Inteligência Artificial: repensando a natureza humana**. Rio de Janeiro: Lumen Juris, 2020.

WOLKART, Erik Navarro. BECKER, Daniel. Tecnologia e precedentes: do portão de Kafka ao panóptico digital pelas mãos da jurimetria. *In*: ALVES, Isabella Fonseca (org.). **Inteligência Artificial e Processo**. São Paulo: D'Plácido, 2020.

ZAHED, Ludovic Mohamed. Homossexualidade e identidade trans no Islã. **Momento: diálogos em educação**, v. 28, n. 1, p. 245-261, jan./abr., 2019. Disponível em: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwjEmav4-O_4AhUtslUCHeQCC9MQFnoECAoQAO&url=https%3A%2F%2Fperiodicos.furg.br%2Fmomento%2Farticle%2Fdownload%2F8798%2F5792&usg=AOvVaw3CNo1ih8657q70q5e3EKLD. Acesso em: 10 jul. 2022.

Recebido em 15/05/2022
Aceito em 25/07/2022