REDES
Revista de estudios sociales de la ciencia y la tecnología

Neutros e objetivos? Uma análise do uso de algoritmos em processos de tomadas de decisão a partir das epistemologias feministas[1]

Maria Vitoria Pereira de Jesus[*]

Bruno Lucas Saliba de Paula [**]

Resumo

O objetivo deste trabalho é analisar as controvérsias em torno da neutralidade e da objetividade dos algoritmos utilizados em processos de tomadas de decisão. Apoiamo-nos nos Estudos Sociais da Ciência e Tecnologia, bem como nos insights provenientes das epistemologias feministas, a fim de problematizar a presença de valores e interesses em inovações sociotécnicas relacionadas às tecnologias de informação e comunicação. Demonstramos, assim, que, em contextos de

[1] Os/as autores/as agradecem à Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) pelo financiamento desta pesquisa.

[*] Universidade Estadual de Montes Claros (UNIMONTES). Universidade Estadual de Campinas (UNICAMP). Correo electrónico: maria.vi.toriap959@gmail.com

[**] Departamento de Ciências Humanas e Sociais Aplicadas ao Direito/Universidade do Estado de Minas Gerais (UEMG). Programa de Pós-Graduação em Sociologia da Universidade de Brasília (PPGSOL/UNB). Correo electrónico: bruno.paula@uemg.br

DOI: https://doi.org/10.48160/18517072re56.212

disparidades estruturais e de subrepresentação de minorias em áreas de tecnologia e inovação, análises e decisões automatizadas através de sistemas de inteligência artificial tendem a reproduzir desigualdades em termos de gênero, raça e classe. Do ponto de vista metodológico, valemo-nos das técnicas de análise de conteúdo a fim de investigarmos matérias jornalísticas pertinentes ao tema, além de relatórios e legislações. Nossos argumentos apontam para a necessidade de levar em consideração a situacionalidade dos sujeitos envolvidos com a programação de sistemas automatizados de tomadas de decisão, o que tensionaria critérios tradicionais de neutralidade e objetividade, tal como defendido pela noção de "objetividade forte", definida por Sandra Harding. Dessa forma seria possível levar em consideração os valores e interesses que perpassam as práticas de inovação, de modo a reconhecer, e talvez mitigar, os vieses políticos e socioculturais, bem como os efeitos discriminatórios, que caracterizam as decisões com base em algoritmos.

Palavras Chave

ESTUDOS SOCIAIS DA CIÊNCIA E TECNOLOGIA, EPISTEMOLOGIAS FEMINISTAS, INTELIGÊNCIA ARTIFICIAL, ALGORITMOS

Introdução

Em diversos países, o uso de algoritmos já é realidade em procedimentos decisórios, tanto no âmbito governamental quanto empresarial, que buscam a modernização e a neutralidade através de tecnologias e sistemas baseados em Inteligência Artificial (IA). Na Nova Zelândia, por exemplo, este fato se fez evidente na criação de um sistema utilizado para prever a possibilidade de que recém-nascidos sofram ou não maus-

tratos com base em variáveis como a idade dos pais, sua saúde mental e situação financeira (Pascual, 2019). Já no Brasil, o uso de algoritmos foi evidenciado em setores do judiciário, nos quais sistemas de IA realizam a triagem de processos e a tomada de decisões com base na digitalização de processos e sentenças dadas anteriormente (Ferreira, 2020). Deste modo, embora sejam mais conhecidos em mecanismos de busca e em redes sociais online, torna-se evidente a presença de algoritmos em instituições públicas e privadas a fim de que sejam tomadas decisões importantes, sendo adotados em processos pautados nos ideais de “neutralidade” e “objetividade”.

No entanto, recentemente, são vários os estudos, notícias e matérias jornalísticas em que se evidenciam as controvérsias em torno da neutralidade e da eficiência dos algoritmos presentes em tecnologias e sistemas automatizados, que, ao realizar análises e sugestões com base em perfis e predições algorítmicas, promovem a discriminação de indivíduos e corroboram os vieses e atitudes baseadas em preconceitos. Dessa forma, indagamos: “tecnologias têm qualidades políticas?” (Winner, 1986: 1). Ou ainda, é possível alcançar resultados neutros, objetivos e eficientes quando empregamos os algoritmos em procedimentos socioinstitucionais? Diante disso, as análises realizadas por Langdon Winner (1986) e Andrew Feenberg (2010) nos são de suma importância, pois revelam a política e as contradições existentes nas aplicações práticas das tecnologias e do conhecimento científico. Igualmente relevantes são as considerações de Evelyn Fox Keller (1991) de que tanto o gênero quanto a ciência são categorias socialmente construídas e, diante do sexismo imperante nas atividades científicas, é preciso avaliar “como a construção de homens e mulheres afeta a construção da ciência” (Fox Keller, 1991: 12). Em outras palavras, nosso objetivo, de modo geral, é analisar a política incorporada aos objetos técnicos em termos das inúmeras relações de poder e de desigualdades sociais,

sejam elas efeitos de gênero, raça ou classe. Não se trata, então, de identificar a política apenas nos “usos” que são feitos da tecnociência, mas sim nos próprios processos que a constituem, como no modo como são formulados problemas de pesquisa e inovação (ou, inversamente, as razões pelas quais alguns temas são simplesmente ignorados como “objetos” de investigação), descritos os fatos, alcançados os resultados e produtos, etc.

Em termos mais específicos, propomo-nos a analisar as controvérsias em torno da pretensão de neutralidade dos algoritmos presentes em tecnologias e sistemas de IA, que cada vez mais tem ganhado espaço em processos de tomadas de decisões relevantes em diferentes setores da sociedade. Poderiam os algoritmos, quando postos em funcionamento, nos fornecer soluções neutras, sem reproduzir qualquer viés, discriminação e preconceito? Poderiam eles ser objetivos e eficientes ao indicar intervenções com base em perfis e predições provenientes da mineração de dados? Diante dessas questões, apresentamos alguns estudos de caso - sobretudo relacionados à discriminação de gênero a partir do uso de algoritmos - que operam como fundamentos para nossa argumentação. Constatamos que, diante do sexismo, racismo e classismo estruturais, os algoritmos tendem a incorporar e a reproduzir esses mesmos valores em seu modus operandi. Para tanto, baseamos-nos em debates consolidados no âmbito dos Estudos Sociais da Ciência e Tecnologia (ESCT), como a política intrínseca aos artefatos tecnológicos, o contexto socioeconômico de sua construção e a possibilidade de que eles incorporem “diferentes graus de poder assim como diferentes níveis de consciência” (Winner, 1986: 9). Valemo-nos igualmente da literatura sobre tecnologias da informação, vigilância e governança algorítmica (Doneda; Almeida, 2018; Bruno, 2008, 2013; Zuboff, 2019). Mais especificamente, sustentamos nossas proposições a partir dos insights provenientes

das epistemologias feministas, especialmente através da noção de “objetividade forte” (Harding, 2019).

Por fim, no intuito de fundamentar empiricamente a nossa discussão, foram selecionadas e examinadas matérias jornalísticas que expõem breves retratos, ou estudos de caso, de situações significativas de discriminação algorítmica. Como a análise aqui empreendida volta-se a um tema contemporâneo, que ainda carece de maior consolidação na literatura especializada, o material proveniente da imprensa revelou-se uma base relevante de dados secundários. Ao todo, selecionamos nove reportagens, por meio de buscas nos mecanismos de pesquisa dos principais veículos brasileiros Folha de São Paulo, BBC Brasil e El País Brasil, e também em outros veículos de comunicação e notícia como o UOL, Reuters e Convergência Digital, entre os períodos de maio de 2017 a outubro de 2021. Uma vez realizado esse levantamento, analisamos o material selecionado através das técnicas de análise de conteúdo preconizadas por Martin Bauer (2008). A escolha desse método se deve à possibilidade de “produzir inferências de um texto focal para seu conteúdo social de maneira objetivada” (Bauer, 2008: 191), o que, em nosso caso, permitiu efetuar uma análise cientificamente relevante a partir de um material não-acadêmico, que são as matérias jornalísticas. Nosso corpus foi, inicialmente, categorizado e codificado conforme nossos referenciais teórico-conceituais (Bauer, 2008: 199), o que resultou, num segundo momento, numa análise comparativa entre os diversos materiais que compuseram nossa amostra. Em seguida, conduzimos nossa interpretação e análise a partir de um quadro tripartido de perspectivas: a dos pesquisadores, a decorrente dos materiais empíricos e aquela presente das vertentes teóricas que nos inspiraram. Finalmente, investigamos ainda, em menor quantidade, relatórios e legislações. Esses materiais foram analisados separadamente, já que não compuseram um volume mais extenso de dados.

O presente texto está organizado da seguinte forma. Num primeiro momento, são apresentadas definições elementares, tais como a de "algoritmos" e a de “inteligência artificial”. Essas noções são associadas criticamente ao contexto sociotécnico em que emergiram, baseados em ideais de eficiência, objetividade e neutralidade. Esse imaginário é novamente problematizado, na seção seguinte, a partir das epistemologias feministas, cujos conceitos são acionados a fim de evidenciarmos o caráter sexista e androcêntrico da construção da tecnociência. Também nessa seção argumentamos a favor da diversificação dos sujeitos envolvidos nos processos sociotécnicos de programação e uso da IA. Finalmente, apresentamos, na última parte do artigo, princípios de regulação e governança algorítmica potencialmente relevantes para as tentativas de reduzir os vieses e discriminações nas decisões automatizadas.

As promessas e os futuros imaginados através da IA

Com o surgimento da cibernética e o desenvolvimento do behaviorismo radical, cujos experimentos se dedicam ao controle social do comportamento humano, o homem e suas condutas tornaram-se alvo da racionalidade do conhecimento científico. Sendo assim, depois de nos tornarmos “os mestres e senhores da natureza” (Feenberg, 2010: 53), restar-nos-ia desvendar ou resolver as “equações” da comunicação e do comportamento humano, combinando-os ao progresso e as exigências do desenvolvimento tecnológico. Na década de 1940, ao propor o campo de estudos denominado cibernética, Norbert Wiener e seus colaboradores apresentam-nos a descrição de um sistema eletromecânico que, segundo ele, seria capaz de desempenhar funções exclusivamente humanas (Kim, 2004) e, para isso, Wiener parte da ideia de que “certas funções de controle e de processamento de informações

semelhantes em máquinas e seres vivos [...] são, de fato, equivalentes e redutíveis aos mesmos modelos e [...] leis matemáticas" (Kim, 2004: 200). Tal proposição estabelece um fundamento importante para os experimentos em favor da inteligibilidade das máquinas, que passariam a se comunicar e a realizar funções e tarefas antes executadas por humanos, por meio de um complexo cálculo de operações matemáticas. Dessa forma, abre-se a possibilidade, ou a pretensão, de que funções e tarefas sejam realizadas de modo "neutro" e "objetivo" que, em consonância com os ideais de neutralidade da ciência e tecnologia (Dagnino, 2008), confirmariam as promessas e os valores da sociedade ocidental moderna. Trata-se, então, da constituição de um "imaginário sociotécnico", nos termos de Sheila Jasanoff (2015), que posiciona as tecnologias de informação no lugar da eficiência e da emancipação do que seriam as falhas ou limitações humanas. Como definido pela autora, "nós redefinimos imaginários sociotécnicos [...] como visões coletivas, institucionalmente estabilizadas e publicamente praticadas de futuros desejáveis, suscitados por entendimentos compartilhados sobre a vida social e a ordem social alcançáveis pelos avanços da ciência e tecnologia" (Jasanoff, 2015: 4 - tradução nossa). Em outras palavras, os imaginários sociotécnicos envolvem a aspiração de um futuro a ser alcançado por meio de inovações científicas e tecnológicas que convergem com os valores em geral positivos e otimistas em relação ao progresso social e tecnológico. Por outro lado, contrariamente a essas projeções desejadas, de como esperamos que seja o futuro, ou normativas, de como ele deve ser, seria possível analisar esse processo de emergência e posterior disseminação das tecnologias de informação como exemplo daquilo que Donna Haraway (2013) entende como "informática da dominação". Ao caracterizar as relações entre ciência, tecnologia e sociedade no capitalismo contemporâneo, a autora considera que haveria, sobretudo com o advento das tecnologias de informação e comunicação, uma

sobreposição entre formas tradicionais de dominação, hierárquicas e centralizadas, e modos contemporâneos, descentralizados e em rede. Processo semelhante é descrito por Gilles Deleuze (2013) através do conceito de “sociedade de controle”.

Entendemos a Inteligência Artificial como a capacidade de um sistema computacional de tomar decisões ou fornecer sugestões de tomada de decisão com base em classificações e predições (Vicentin, 2022) algorítmicas com base em um grande conjunto de dados. Os algoritmos são componentes essenciais das tecnologias de informação e sistemas de IA que paulatinamente vêm sendo adotadas por instituições públicas e privadas na condução de seus procedimentos. São eles que possibilitam com que a extração de padrões e a correlação entre os elementos contidos em um vasto conjunto de dados (Pasquinelli & Joler) sejam feitas de forma ágil, que ultrapasse o da cognição humana. Na web, eles são conhecidos por sua agilidade em extrair padrões e regularidades em meio a um conjunto de informações que são obtidas através dos rastros digitais (Bruno, 2013) deixados, por exemplo, quando acessamos notícias ou clicamos em anúncios e sites. Com o crescimento do comércio e da publicidade online, os algoritmos têm sido ferramentas cada vez mais solicitadas por empresas e grupos publicitários a fim de estabelecer produtos e práticas e direcioná-las aos internautas, potenciais compradores e/ou usuários das redes sociais online. Neste contexto, “os algoritmos de recomendação mapeiam nossas preferências em relação a outros usuários, trazendo ao nosso encontro sugestões” (Gillespie, 2018: 97) de produtos, hábitos e informações consideradas relevantes para nós com base no conhecimento já acumulado sobre a nossa atividade e a de outros usuários considerados “parecidos” conosco em termos preferenciais, probabilísticos e demográficos (Beer apud Gillespie, 2018).

A implementação de tecnologias de informação para a realização de diferentes procedimentos fez com que os algoritmos se tornassem elementos constitutivos

importantes, por exemplo, desde os sistemas de policiamento preditivo, passando pelas câmeras de reconhecimento facial até programas utilizados na previsão de maus-tratos contra crianças e bebês recém-nascidos. Reconhecer, prever e predizer, nesses casos, só é possível depois da máquina ser submetida a um processo de aprendizagem, também conhecido como Machine Learning, que ocorre quando ela é a "alimentada" com um volume significativo de dados, tornando possível que os seus algoritmos forneçam padrões e perfis sobre os indivíduos e as situações da realidade.

O processo de análise de dados, sobretudo para e em tomadas de decisão, que até então ficava a cargo da atividade e interpretação humana, agora é feito por sistemas de IA, deixando dessa forma as decisões por conta da classificação e da seleção algorítmica. Os resultados fornecidos pelos algoritmos ofereceriam bases sólidas para a tomada de decisões, de forma pretensamente objetiva, na medida em que "as informações geradas são analisadas de modo matemático" (Aragão & Benevides, 2019: 7), o que inviabilizaria a ocorrência de vieses e interpretações subjetivas, marcando, assim, o sucesso do pensamento racional sobre a interpretação humana, permeada por falhas e por preconceitos e vieses econômicos, sociais e políticos (Rouvroy, 2012). Portanto, além da ideia de modernização, a IA surge com a promessa de contornar a subjetividade e os valores socioculturais fortemente imbricados nas ações e no comportamento humano, sendo adotadas por diferentes instituições públicas e privadas que cada vez mais buscam fornecer resultados ágeis, "neutros" e "objetivos" na realização de análises e procedimentos.

No Brasil, por exemplo, as propostas de "modernização" e de um governo digital, que deseja "ampliar o acesso e a qualidade dos serviços públicos e promover a transformação digital da gestão e dos serviços" (Decreto nº 10.609, 2021: 2) também já sinalizam o uso de IA. No decreto 10.332 publicado no dia 28 de abril de 2020, o governo federal institui a Estratégia de um Governo Digital para os órgãos e demais

setores da administração pública, no qual propõe a oferta de “serviços públicos digitais simples e intuitivos”. Além disso, pretende “disponibilizar a identificação digital ao cidadão”; realizar a promoção de “políticas públicas baseadas em dados e evidências; e oferecer serviços preditivos e personalizados, com a utilização de tecnologias emergentes” (Decreto nº 10.332, 2020: 3) como o Big Data e a IA, que vem sendo amplamente utilizada nos serviços de atendimento ao público. Já no que tange à Política Nacional de Modernização do Estado, o “Moderniza Brasil”, publicado no dia 21 de janeiro de 2021, os objetivos apontam para a “articulação, o monitoramento e a avaliação de políticas, programas e iniciativas de modernização do Poder Executivo” (Decreto nº 10.332, 2020: 1), que pretende promover a simplificação das relações entre Cidadão e Estado, a agilidade dos serviços e a eficiência da gestão pública por meio da implementação de um “governo e uma sociedade digital” (Decreto nº 10.609, 2021: 2). Além disso, através de tecnologias digitais, ambas as propostas buscam promover a “desburocratização” do Estado, tornando o atendimento, os serviços e as políticas públicas mais ágeis, eficientes, econômicas (Aragão & Benevides, 2019) e menos burocráticas. Os resultados de correlações algorítmicas já são bases relevantes também para a concessão e cortes de benefícios sociais do governo federal, tal como o programa de transferência de renda “Bolsa Família”. Segundo Grossmann (2018), desde o lançamento da plataforma Govdata, do governo federal, em abril de 2018, já havia sido cancelados 5,2 milhões de benefícios do Bolsa Família por meio do cruzamento de dados. O GovData abriga informações de diferentes setores e órgãos públicos do Estado, e, por meio de ferramentas de análise de dados, promete auxiliar no fornecimento de resultados objetivos e estratégicos para a formulação de políticas públicas e a concessão de benefícios sociais. No Govdata, são os algoritmos os responsáveis por realizar o cruzamento de dados e, por conseguinte, identificar a correlação e a veracidade das informações contidas nas

bases do governo federal. Portanto, neste caso, nas situações em que há cortes de benefícios, as análises sobre a veracidade das informações declaradas pelos beneficiários seriam apoiadas em correlações algorítmicas que, com base no Big Data do governo federal, dariam permissão ou indicariam as possíveis fraudes na concessão dos benefícios. Por fim, ainda no caso brasileiro, nota-se o esforço no desenvolvimento de programas "inteligentes", para além do Poder Executivo, também no âmbito do Judiciário. Segundo Bruno Bioni, Mariana Rielli e Marina Kitayama (2021: 62), "a Defensoria Pública [...] está empenhada em um projeto de sistematização e geração de inteligência sobre os processos de todos os seus órgãos distribuídos pelo estado". Tal projeto pretende "gerar inteligência para a propositura de ações civis públicas" (Bioni; Rielli & Kitayama, 2021: 62), com base no conjunto de informações sobre os litígios individuais. Além disso, o programa teria o objetivo não só de servir de argumento convincente para a firmação de acordos, como também de aumentar as chances de sucesso das ações da instituição no judiciário (Bioni; Rielli & Kitayama, 2021). Sendo assim, o uso de "Inteligência" auxiliaria na triagem e na sugestão de andamento com os processos, visando à maximização de ganhos das causas e dos processos protocolados pela Defensoria Pública.

Em todos esses casos, percebe-se a tentativa de tornar eficientes e econômicos os serviços e procedimentos realizados no âmbito do Estado. A tônica discursiva é baseada nos ideais de eficiência, facilidade, agilidade e neutralidade que poderiam ser alcançados através da adoção de novos recursos sociotécnicos que, de forma alinhado às concepções do determinismo tecnológico (Marx & Smith, 1994), por si só seriam capazes de "transformar" o Estado brasileiro e suas relações com os cidadãos. Com a conexão digital "inteligente" estabelecida entre governo e sociedade, o governo poderia, assim, "enxugar" os gastos com o serviço público e os cidadãos seriam beneficiados com a desburocratização do Estado (Aragão & Benevides, 2019). No

serviço de atendimento ao público, o uso da IA possibilitaria a obtenção de respostas rápidas, quase automáticas, com base em algoritmos que realizam a predição e a personalização dos serviços. Já na Defensoria Pública, os algoritmos do programa "inteligente" deveriam identificar os padrões e regularidades dos processos contidos nas bases de dados da instituição, orientando os advogados sobre as chances de sucesso e a melhor maneira de dar prosseguimento às causas.

A utopia e seus impasses: reflexões sobre a reprodução de preconceitos e discriminações por algoritmos com base nas epistemologias feministas

Os algoritmos são treinados para reconhecer os padrões implícitos no conjunto de dados utilizados para a aprendizagem da IA. No entanto, se os dados utilizados para o seu treinamento contarem com vieses de gênero, raça ou classe, é provável que a máquina os reproduza (Pasquinelli, 2017) e até discrimine com base em inferências estatísticas. Nos Estados Unidos, por exemplo, foi identificado que um sistema utilizado na prevenção da reincidência de presos no país emitia vieses racistas ao induzir que "os acusados negros eram duas vezes mais propensos a serem mal rotulados como prováveis reincidentes" (Salas, 2017: 3) do que os acusados brancos. Matteo Pasquinelli (2017) também salienta sobre as falhas ocorridas nos sistemas de reconhecimento facial que, ao serem treinados com dados enviesados – por exemplo, dados que contemplem apenas os rostos de pessoas brancas – fracassam consideravelmente em reconhecer pessoas negras. O contrário também acontece nos sistemas de reconhecimento facial utilizados pela polícia que insistem em "reconhecer" os negros como prováveis criminosos, aumentando com isso, os riscos de prisões injustas e perseguições racistas.

Os erros e vieses emitidos pela IA ocorrem devido ao tratamento dado à coleta de dados utilizados no treinamento de seus algoritmos ou, em alguns casos, é possível que eles estejam implícitos na maneira com que eles são programados. Quando programamos um algoritmo para que ele desempenhe uma função que geralmente é feita por um ser humano, definimos um problema e o inscrevemos em uma sequência de passos para que ele o resolva (Lucena, 2019). Resta-nos saber se uma série de passos inscritos por um sujeito localizado em um lugar corporal e social específicos, na contramão da pretensão de objetividade abstrata e descorporificada (Haraway, 2009), poderia fornecer soluções que contemple os interesses e a heterogeneidade de um grupo ou de uma população. Na maioria dos casos, a programação é feita sem muita preocupação com os interesses e riscos de discriminação e violação de direitos de mulheres, negros e demais sujeitos tidos como “minorias”. Homens e mulheres, embora sejam de uma mesma cultura, interagem de maneira diferente com os ambientes sociais e “à medida que se ocupam de diferentes tipos de atividade, desenvolverão e manterão padrões distintos de conhecimento” (Harding, 2007: 167) sobre os problemas e a realidade. Tal proposição nos faz pensar em como seria o desenvolvimento de uma IA e, por conseguinte, a programação de algoritmos feita, por exemplo, por mulheres, sobretudo nos casos em que o uso dessas tecnologias é feito e aplicado a públicos diversos em termos de “raça” e gênero. Como seria o funcionamento e a eficácia de um sistema cujos algoritmos foram programados e treinados por mulheres para prever probabilisticamente, por exemplo, a chance de pais e mães cuidarem bem ou não de seus filhos?

A fim de problematizarmos essas questões, valemo-nos das reflexões provenientes das epistemologias feministas em torno dos vieses de gênero presentes na produção da tecnociência. Além de questionar as relações de poder sexistas presentes nas práticas da ciência e da tecnologia, as perspectivas feministas chamam

a atenção para aspectos epistêmicos que perpassam esses campos. Nesse sentido, são estabelecidos critérios de objetividade a partir da constatação das tendências sexistas e androcêntricas que caracterizam as práticas de pesquisas acadêmicas. Assim, não se trata de adotar o princípio de neutralidade em relação a valores e interesses como definidor do que seria uma boa ciência capaz de produzir resultados válidos e confiáveis, até porque, numa sociedade marcada pela dominação masculina, aquilo que seria considerado “neutro” e “objetivo” seria precisamente estabelecido conforme o ponto de vista dos homens (Harding, 2019: 143-146). Não por acaso, Harding observa que as epistemologias tradicionais descartam as mulheres enquanto “agentes do conhecimento” e, consequentemente, fazem com que o sujeito produtor da ciência e das narrativas históricas e sociais seja necessariamente uma figura masculina e dominante em termos de raça e classe (Harding, 1987: 3).

No mercado de trabalho, em áreas como a ciência da computação e no desenvolvimento de tecnologias de informação, a presença de homens brancos, em sua maioria provenientes das classes mais favorecidas, confirmam a ideia de que existe uma homogeneidade de gênero, raça e classe na produção de ciência, tecnologia e inovação, sobretudo em áreas consideradas “duras”. Na maioria dos casos, a ausência e o baixo número de mulheres na área da computação e de tecnologia da informação (TI) se explica pelo fato delas serem, frequentemente, desencorajadas a seguirem tais profissões com base em estereótipos e qualidades “naturais” (Maia, 2016), que, além de serem atribuídas a gêneros específicos, promovem a divisão sexual do trabalho e a desigualdade de gênero na academia e no mercado.

Os estereótipos e qualidades atribuídas às mulheres, ao longo da história, fizeram com que elas fossem sempre submetidas a atividades e profissões relacionadas ao cuidado (Maia, 2016) e afastadas da produção da ciência e tecnologia

por serem menos “racionais” e “objetivas”. As mulheres, por serem demasiadamente “sentimentais”, “emotivas” e “subjetivas”, não seriam capazes de realizar um trabalho tão bom quanto os indivíduos do sexo masculino, tidos como mais racionais, objetivos (Lima, 2013) e, portanto, mais aptos para construir uma “ciência pura”, que fornece resultados e soluções “neutras” para os inúmeros problemas da realidade. Outro argumento justifica que existem profissões que são mais adequadas para os homens do que para as mulheres e, nesse caso, se destacam as áreas de tecnologia, computação e ciências exatas. Dessa forma, se ainda observamos um baixo número de mulheres na TI, é possível que um dos motivos da baixa adesão e permanência delas nesse campo sejam os padrões socioculturais e ideais sexistas, que definem não só atitudes e comportamentos, como também atividades e profissões segundo o sexo.

A presença de mulheres em atividades de pesquisa e em áreas de engenharia e ciência da computação ainda é baixa, se comparada à presença dos homens nos mesmos campos do conhecimento. Segundo o relatório divulgado pela UNESCO (2021), em 2018, as mulheres representavam 33% dos pesquisadores em todo o mundo. Nas áreas de engenharia e computação, elas representavam 28% e 40%, respectivamente, e, na produção de tecnologias como a IA, elas representavam apenas 22% dos profissionais. Esses números podem ser ainda mais baixos se considerarmos alguns países individualmente e o incentivo que é dado às meninas e mulheres para que exerçam a atividade de pesquisa e se dediquem à produção da ciência e tecnologia. O baixo número de mulheres em atividades de pesquisa, desenvolvimento e inovação indica não só a presença de um contexto de desigualdades de gênero e oportunidades, como também contribui para que tenhamos uma ciência “particularista [...] e sexista” (Lima, 2013: 795) e resultados baseados em visões masculinas e androcêntricas (Lima, 2013) sobre a realidade.

É notável, portanto, a homogeneidade das comunidades acadêmicas hegemônicas, que “atraem e admitem apenas cidadãos de um conjunto específico de valores e interesses sociais da elite e os treina para práticas de pesquisa que levam adiantes tais valores e interesses específicos” (Harding, 2019: 146). Reconhecer e encarar a adesão a esses valores e interesses, bem como suas interferências sobre as atividades científicas, não faria com que as pesquisas se tornassem distorcidas ou subjetivas. Pelo contrário, a consideração consciente dos mesmos possibilitaria o alcance de uma “verdadeira objetividade”, que, longe de ser assentada no ideal da neutralidade em relação a valores, seria precisamente aquela que potencializa a confiabilidade dos resultados de investigações na medida em que, ciente dos preconceitos e interesses que atravessam as atividades científicas, mostra-se capaz de lidar com retidão diante das evidências, críticas e objeções que constituem uma pesquisa (Harding, 2019: 148). Finalmente, o afastamento e a autonomia em relação às pautas, valores e pontos de vista dominantes, bem como adoção de experiências e perspectivas contra-hegemônicas, seria outro gesto potencializador da “objetividade forte”, nos termos de Harding.

> A abordagem que leva em conta as perspectivas propõe que os pesquisadores deveriam começar suas investigações fora dos quadros conceituais dominantes – especificamente, nas vidas cotidianas dos grupos oprimidos tais como as mulheres –, a fim de obter relatos mais objetivos das relações naturais e sociais. Aqui eu tenho em vista a proposta de “objetividade forte”, que surge das teorias das perspectivas (Harding, 2019: 146).

A autora pontua que, raramente, grupos oprimidos levantam questões apenas a fim de alcançar uma “verdade pura”, mas sim em decorrência de uma necessidade e desejo de transformar suas condições de vida (Harding, 1987: 8). Isso evidencia o caráter ético-político das pesquisas formuladas e desenvolvidas “desde baixo”. Assim, enquanto a filosofia da ciência tradicional toma como irrelevantes para a

avaliação da qualidade de pesquisas as origens e o contexto em que são formuladas perguntas e hipóteses de investigação, as epistemologias feministas conferem relevância a quem, como e em quais condições foram projetadas e executadas as atividades científicas. Assim, a ciência deixa de ser vista como uma atividade abstrata e desincorporada, capaz de produzir enunciados universais e descolados do mundo, e passa a ser tratada a partir da situacionalidade dos/as pesquisadores/as (Harding, 1987: 6-7; Haraway, 2009). De acordo com Harding, ao deixarem de ser tratados de forma anônima, invisível e abstrata e serem considerados em sua historicidade, posição social, com interesses e valores específicos, os/as pesquisadores/as não perdem, mas ganham objetividade em suas práticas acadêmicas. “A introdução do elemento ‘subjetivo’ nas investigações de fato aumenta a objetividade das pesquisas e diminui o ‘objetivismo’ que esconde esse tipo de evidência [sobre as crenças e comportamentos do pesquisador] do público” (Harding, 1987: 9 - tradução nossa). Semelhantemente, Fox Keller (1991) pontua que a busca por objetivismos, pautada na exclusão do sujeito e de suas dimensões pessoais, acaba por produzir, contraditoriamente, perspectivas e explicações pobres e limitadas.

> Uma ideologia objetivista, que proclama prematuramente o anonimato, o desinteresse e a impessoalidade, e que exclui radicalmente o sujeito, impõe um véu sobre essas práticas [cotidianas]. [...] O esforço em prol da universalidade se fecha em si mesmo, e com isso se protege a estreiteza do olhar. Assim, a ideologia da objetividade científica trai seus próprios propósitos, subvertendo tanto o significado quanto o potencial da investigação objetiva (Fox Keller, 1991: 20).

É preciso reconhecer que as contribuições feministas aos ESCT não constituem um bloco homogêneo e monolítico, mas se dividem em inúmeras abordagens e vertentes. Além disso, uma das críticas direcionadas às epistemologias feministas produzidas no Norte Global tem a ver com sua incapacidade de levar em consideração as estruturas

historicamente consolidadas do imperialismo e do colonialismo, deixando de lado as condições das diversas populações indígenas, bem como a variedade dos conhecimentos por elas produzidos (Subramanian et al., 2017: 409). Esse é, aliás, o movimento empreendido pelas linhas pós-coloniais dos estudos em ciência e tecnologia que, a partir de perspectivas não-eurocentradas, reconsideram as narrativas excepcionalistas e triunfalistas da tecnociência do Norte Global. Daí deriva uma "reflexividade robusta" diante dos projetos e promessas da modernidade ocidental (Harding, 2008). Por outro lado, estudos pós-coloniais da ciência e tecnologia tendem a obliterar as questões de gênero em suas abordagens. Diante dessas limitações, surgem as tentativas de articular as duas perspectivas: "Os estudos pós-coloniais nos ESCT se diferenciam [...] por sua atenção primária ao colonialismo e imperialismo, geralmente deixando o gênero de lado. O feminismo nos ESCT compartilha uma limitação parecida na sua falta de atenção ao colonialismo e às identidades indígenas. Para lidar com essas limitações, pesquisadores começaram a explorar as conjunções entre ESCT, feminismo e estudos pós-coloniais" (Subramanian et al., 2017: 409 - tradução nossa). Nesse sentido, as autoras argumentam que "questões de gênero, raça, colonialidade e identidades indígenas não são variáveis opcionais que cada campo pode escolher se leva ou não em consideração" (Subramanian et al., 2017: 422 - tradução nossa). Concordamos com esse posicionamento e incluiríamos ainda a categoria "classe" como outra que não pode ser obliterada sob pena de reproduzir desigualdade nos processos de produção de conhecimento. Assim, a articulação entre ESCT, feminismo e pós-colonialismo, por propor reflexões fora dos "quadros conceituais dominantes", nos termos de Harding, parece-nos um tanto prolífica para a produção de uma tecnociência mais justa e inclusiva. No caso específico das tecnologias de informação e comunicação, acreditamos que a heterogeneidade do corpo profissional de programadores/as, bem

como a reconsideração dos ideais de neutralidade a partir das problematizações suscitadas pelo conceito de “objetividade forte”, potencializaria não só a justiça, mas a qualidade das decisões tomadas por meio do uso de algoritmos e Inteligência Artificial. No Brasil, a necessidade de uma equipe diversificada em termos de gênero, profissões e raça já é assegurada pelo Conselho Nacional de Justiça (CNJ), que, na Resolução nº 332 (CNJ, 2020), fornece disposições gerais e orientações para um desenvolvimento da IA no judiciário.

Diante dos argumentos apresentados, poderíamos problematizar não apenas as questões que levam à criação de modelos de IA, mas também alguns dos resultados que eles emitem. Na maioria dos casos, não levamos em consideração que a identificação de um problema ou mesmo a sugestão de como solucioná-lo, são processos que podem estar permeados por vieses de gênero, raça e classe. Se, por exemplo, na tentativa de resolver o problema de maus-tratos contra crianças, um sistema é utilizado para prever a possibilidade de que as mães cuidem bem dos seus filhos (Pascual, 2019), é necessário que o modelo criado compreenda o perfil de uma mãe passível de maltratar os seus filhos, para que, depois, ele possa identificar e sugerir a melhor decisão. No exemplo da reportagem de Pascual (2019), o sistema utilizado pelo governo da Nova Zelândia para prever maus-tratos contra crianças e bebês recém-nascidos foi acusado de erro na maioria dos casos em que foi usado para análise. No entanto, não fica claro quais dados foram usados para treinar o algoritmo que criou o modelo que sugeria o perfil das mães passíveis de cometer maus-tratos.

Em entrevista à Rede Lavits (Rede latino-americana de estudos sobre vigilância, tecnologia e sociedade), Cristina Plamadeala (2021) salienta a situação de mulheres grávidas e em pós-parto que preferem não revelar à equipe médica que estão passando por conflitos que interferem em sua saúde mental com medo de serem

consideradas como “mães ruins ou incapazes de cuidar dos filhos” (p. 4) e com isso, terem eles levados pelos serviços sociais. Em um contexto de adoção de sistemas de IA e uso de Big Data, é possível que dados pessoais sensíveis, como os relacionados à saúde mental, sejam minerados e processados pelos algoritmos, que fornecem sugestões e resultados para a tomada de decisões com base na definição do problema a ser solucionado. Numa situação em que programas de IA, como o da Nova Zelândia, são utilizados para analisar e prever a capacidade dos pais cuidarem bem ou não de seus filhos, dados de diferentes bases do governo podem ser minerados, utilizados e processados pelos algoritmos, e, se definimos a saúde mental e a condição financeira como fatores determinantes para que mães e pais cuidem bem ou não de suas crianças, é provável que o sistema sugira o acompanhamento e a tutela do exercício da maternidade nos casos de mulheres em que se verifica o histórico de problemas relacionados à saúde mental.

Num contexto em que a participação das mulheres na programação ainda é pequena em relação à dos homens, o perfil do que seria uma mãe inapta aos cuidados e capaz de maltratar o seu filho seria traçado por programadores que, na maioria dos casos, possuem percepções machistas e sexistas em relação à mulher e à maternidade e desconsideram situações específicas, contextos e particularidades. Para o algoritmo considerar que uma mulher com o histórico de problemas de saúde mental será uma má mãe ou incapaz de cuidar dos seus filhos, é necessário que seja fornecido a ele características de uma mulher-mãe “ideal” e das condições mentais necessárias ao exercício da maternidade. Tais características e condições a serem fornecidas requerem, portanto, uma imaginação moral que só pode ser feita por humanos (O’neil, 2016) que, estando socialmente situados, fornecem visões parciais ou limitadas sobre determinadas situações ou problemas da realidade. Para esse caso específico, por exemplo, um grupo heterogêneo de programadores em termos de

gênero, raça e classe possivelmente pensaria de forma multifacetada numa série de questões que desencadeiam os problemas da saúde mental em diferentes contextos e momentos da vida dos indivíduos, principalmente das mulheres que lidam com pressões externas e a idealização da maternidade (Cordellat, 2018), momento em que, no imaginário social, a mulher-mãe só experimentaria sentimentos de auto realização e felicidade.

Em outros casos em que sistemas de IA são usados por instituições públicas, é comum que surjam questões como as de quais dados considerar e quantos pontos atribuir a cada um deles para que um sistema possa sugerir o que fazer ou como agir diante de determinada situação. É nesse momento também que, segundo O'neil (2016), preconceitos e vieses são codificados, fazendo, assim, com que a máquina os reproduza em seus resultados. Em algumas situações, as classificações feitas por seus algoritmos apoiam-se em crenças e suposições sobre personalidades e comportamentos, o que aumenta a relevância de um dado específico sobre o indivíduo. Nos Estados Unidos, por exemplo, O'neil (2016) afirma que são muitos os empregadores que verificam o histórico de crédito dos candidatos para decidir sobre uma contratação. Com o aumento da capacidade de armazenamento digital dos dados em nuvens, os algoritmos podem minerar diversas informações e sugerir decisões com base nos dados fixados como principais para a tomada de decisão.

Alega-se que as classificações e sugestões algorítmicas são feitas de forma objetiva e levam em consideração os padrões e regularidades observadas em cada caso. No entanto, é preciso observar posições e disposições dos sujeitos responsáveis pela programação, que, embora não sejam (ou nem sempre são) conscientes, inserem dados ou fixam modelos discriminatórios que não compreendem a complexidade dos casos em questão. Além da programação, a coleta e o tratamento conferido aos dados utilizados no treinamento dos sistemas são fatores consideráveis

quando se trata dos vieses implícitos nos resultados e sugestões algorítmicas. Os algoritmos treinados a partir de um conjunto de dados totalmente “homogêneos” ou “enviesados” tendem a fixar padrões extremamente específicos quanto ao contexto e à população sobre a qual irá atuar. Os algoritmos treinados, por sua vez, a partir de um conjunto de dados “heterogêneos”, cujo volume e variedade compreendam as características do Big Data, tendem a operar em uma lógica infra-individual e supra-individual (Rouvroy, 2012; Bruno, 2008), que, embora baseiem em fragmentos de informações pessoais de vários sujeitos, distribuídas e fragmentadas em redes e relações interpessoais (Bruno, 2008), também replicam discriminações e preconceitos.

Se considerarmos a primeira opção, em que os algoritmos são treinados com dados homogêneos e enviesados, isso se faz evidente, por exemplo, nas situações em que se constata que “os programas usados nos departamentos de contratação de algumas empresas mostram uma inclinação por nomes usados por brancos e rejeitam os dos negros” (Salas, 2017: 4). O mesmo ocorre nos casos em que se identifica uma disparidade significativa na seleção de homens ao invés de mulheres em vagas de emprego. Essa última situação compreende, por exemplo, o caso da Amazon, que, ao utilizar uma ferramenta de IA para fazer a seleção de candidatos para entrevista à vaga de desenvolvedor de software, evidenciou que o sistema tendia a selecionar mais currículos de indivíduos do sexo masculino do que do sexo feminino (Dastin, 2018). Mais tarde, tornou-se evidente que o viés emitido pela máquina era proveniente do banco de dados processado pelos algoritmos, composto por currículos enviados à empresa nos últimos dez anos (Dastin, 2018). Assim, compreendemos que a tendência do sistema em selecionar mais currículos de indivíduos do sexo masculino ao invés de indivíduos do sexo feminino revela não apenas a possibilidade de que a discriminação algorítmica tenha origem no conjunto de dados, como também

evidência a predominância de um gênero específico (o masculino) na aréa de desenvolvimento de software. São inúmeros os estudos e relatos sobre discriminação de gênero e raça por sistemas de IA, que evidenciam que os vieses podem estar no conjunto de dados processados pelos algoritmos. Um estudo feito pela Universidade da Virgínia, por exemplo, ao identificar que um algoritmo de análise de imagem associava sempre às mulheres a imagem de indivíduos na cozinha (Sayuri, 2019), nos sugere um possível "padrão" dos dados (imagens) usados no treinamento desse algoritmo para que ele classificasse os indivíduos presentes na cozinha como mulheres, embora nem sempre fossem elas nas imagens submetidas para análise. Nesse caso, se o sistema foi treinado com dados "homogêneos" e os algoritmos "ensinados" a classificar como mulheres os indivíduos presentes no espaço da cozinha, é provável que ele vá associar a cozinha às mulheres e discriminar os indivíduos que destoem do perfil já fixado ou do padrão das imagens contidas no banco de dados.

Já no segundo caso, em que os algoritmos são treinados com dados heterogêneos, a discriminação do sistema ocorre devido à falta de critérios na seleção dos dados que fazem com que os algoritmos operem em uma lógica infra ou supra-individual, selecionando padrões e produzindo predições que dizem pouco a respeito de um indivíduo ou de sua pessoa identificável (Bruno, 2013). Nesses casos, os algoritmos podem fornecer resultados completamente aleatórios e prejudicar indivíduos que se submetem a processos de seleção que consideram importantes para a sua trajetória pessoal e profissional. Recentemente, no Reino Unido, por exemplo, um algoritmo utilizado para julgar a nota dos estudantes foi acusado de fornecer resultados injustos e inesperados que fez com que inúmeros jovens perdessem a chance de entrar na Universidade (BBC, 2020).

Portanto, em tais situações, não seria de se estranhar a existência de pontos de vista opostos em relação à instalação de sistemas algorítmicos em análises e procedimentos de tomadas de decisão. No caso do Brasil, os sistemas de Inteligência Artificial já atuam em tribunais, como o Superior Tribunal de Justiça (STJ), sugerindo decisões e fornecendo súmulas das ações judiciais (Sakai, Galdino & Burg, 2021). No entanto, para este fim, o ex-desembargador do Tribunal de Justiça [de São Paulo], José Roberto Neves Amorim (apud Ferreira, 2020) salienta os possíveis riscos de processos criminais e de guardas serem analisados pela IA, pois, para ele, causas como essas "jamais poderão passar por máquinas" (p. 3), uma vez que envolvem uma série de circunstâncias que podem ser negligenciadas na mineração dos dados ou não corresponderem aos profilings identificados pelos algoritmos.

Tornando os algoritmos auditáveis para decisões mais justas, transparentes e responsáveis

É certo que as opiniões em torno da implementação de sistemas de IA em processos de tomada de decisão têm sido múltiplas e suscitado debates fortemente profícuos sobre a sua eficiência, neutralidade e objetividade. Se, por um lado, temos inúmeras propostas como a instituição de uma "Estratégia de um Governo Digital" e iniciativas baseadas no uso de Inteligência Artificial em diferentes setores do judiciário, com o objetivo de "contribuir com a agilidade e coerência do processo de tomada de decisão" (CNJ, 2020: 1), por outro, temos a preocupação com a regulação dos algoritmos que, além de reproduzirem formas de discriminações e preconceitos em seus resultados, emitem sugestões para a tomada de decisões que podem se revelar obscuras ou pouco inteligíveis.

Diante das últimas polêmicas que apontam para a existência de vieses e possibilidade de erros em decisões automatizadas, somos cada vez mais convidados a pensar sobre os constrangimentos sofridos pelos indivíduos que se submetem ou são submetidos às análises e predições fornecidas pelos algoritmos. Sabemos que os erros e vieses emitidos pela IA podem não só ocasionar em sentenças e demissões injustas e colaborar para que estudantes percam a oportunidade de entrar em uma Universidade, como também fazer com que mães se sintam amedrontadas pela possibilidade de perderem a tutela dos seus filhos ao serem acusadas como prováveis cometedoras de maus-tratos. Nesses casos, geralmente, quando são fornecidos os resultados ou tomadas as decisões, não são claros quais foram os critérios e os dados utilizados para se chegar àquele resultado ou tomar àquela decisão.

Em uma reportagem recente, publicada no jornal *El País*, Miquel Echarri (2021) relata, por exemplo, a situação de trabalhadores demitidos por programas de IA que cada vez mais são adotados por empresas, seja para realizar possíveis contratações, ou demitir funcionários. Esse foi o caso de um dos funcionários da Amazon, que, em 2019, foi demitido por um algoritmo que o considerou com um desempenho insatisfatório para o trabalho. No entanto, ele considerou essa decisão injusta e pouco clara, pois ninguém o esclareceu sobre os critérios utilizados no resultado emitido pela máquina (Echarri, 2021). Nos Estados Unidos, um sistema semelhante foi utilizado para realizar a avaliação de professores e, a partir disso, sugerir demissões com base no perfil esperado de um “bom” professor. Nesse caso, ao fixar uma pontuação específica para o que seria um “bom” professor, o sistema demitiu centenas de profissionais cujas pontuações ficaram abaixo do padrão estabelecido (O’neil, 2016). Entretanto, os indivíduos que receberam os comunicados de demissão com base nos resultados fornecidos pelos algoritmos que realizam a correlação e o processamento

de dados, os consideraram inexplicáveis diante das opiniões positivas da comunidade escolar em relação ao trabalho desempenhado por eles.

A opacidade aliada à crença na eficiência e na objetividade dos algoritmos baseados em cálculos e operações matemáticas fazem com que os seus resultados sejam tidos como “incontestáveis” e funcionem como bases relevantes para a tomada de decisões. Segundo Canclíni (2020), vivemos em um contexto de informatização, em que cada vez mais confiamos nas decisões e na autoridade dos algoritmos de macrodados. Confiamos tanto nos resultados algorítmicos que não pensamos duas vezes em aceitar as sugestões da IA de como prosseguir em determinados procedimentos ou realizar a tomada de decisões de forma neutra e objetiva. No entanto, quando os vemos fornecer sugestões e resultados completamente aleatórios ou que destoam da realidade, provocando constrangimentos, é necessário que o modelo seja revisto e o passo a passo dado nos processos de decisão se tornem transparentes e auditáveis.

Posicionamento semelhante é assumido por Arrieta *et al*. (2020: 83-84), que argumentam que modelos de IA baseado em *Machine Learning* operam como caixas-pretas e, comumente, suas decisões e predições carecem de interpretabilidade, os que as torna incompreensíveis para os humanos. Diante disso, os autores discutem os esforços analíticos e conceituais feitos no âmbito da Explainable Artificial Intelligence (XAI), cujas propostas visam tornar a IA mais transparente e explicável. Contudo, para Arrieta *et al*. (2020), pelo menos duas questões surgem desse objetivo. A primeira é que a explicabilidade nem sempre tem como foco os públicos para os quais a IA deve ser inteligível e explicada, o que faz com que esse princípio perca parte de sua efetividade. A segunda é que a explicabilidade é frequentemente impraticável, já que há muitos sistemas que não são transparentes em si mesmos, isto é, que não embutem em sua programação, ou desenho técnico, a possibilidade

de serem explicados. Por essa razão, os autores defendem que as proposições da XAI sejam baseadas, além da explicabilidade, também em princípios como justiça, accountability e privacidade[2], associados ao que eles consideram "IA responsável", um paradigma mais abrangente e capaz de lidar de modo mais efetivo com situações que envolvem informações sensíveis e confidenciais, sobretudo aquelas atreladas aos modelos de "fusão de dados", ou seja, que se alimentam de múltiplas e variadas fontes de dados (Arrieta *et al*., 2020). Finalmente, apesar de várias instituições demonstrarem preocupações com os efeitos negativos e imprevisíveis decorrentes da utilização de IA em suas práticas, a implementação de princípios de "IA responsável" depende de mudanças na cultura organizacional de cada órgão ou corporação (Arrieta *et al*., 2020: 108). Acrescentaríamos, nesse ponto, a partir do que argumentamos por meio das epistemologias feministas, que essas mudanças passam pela diversificação dos sujeitos envolvidos tanto na construção - social e técnica - dos sistemas de IA quanto em âmbitos institucionais públicos e corporativos, de modo a torná-los menos homogêneos em termos de gênero, classe e raça.

Do ponto de vista legal, já há, no Brasil, dispositivos voltados à falta de transparência, ou de responsabilidade, de predições e decisões provenientes dos uso de IA. Por exemplo, o artigo 20 da Lei Geral de Proteção de Dados (Lei nº 13.709, publicada no dia 14 de agosto de 2018), define que, nas situações em que os indivíduos são alvos de análises discriminatórias ou recebem resultados que

[2] Além desses, dois outros princípios de governança algorítmica são recorrentes na literatura especializada e parecem-nos igualmente relevantes, a saber, o de "precisão" e o de "conscientização". Enquanto o primeiro prevê a necessidade de que as fontes de erros e imprecisões de um algoritmo sejam facilmente detectadas, o segundo tem a ver com um processo educativo através do qual programadores e usuários da IA tornem-se cientes das consequências - especialmente as discriminatórias - dessas tecnologias (Mendes; Mattiuzzo, 2019: 56).

comprometem os seus interesses e o alcance de oportunidades, “o titular dos dados tem direito a solicitar a revisão de decisões tomadas unicamente com base em tratamento automatizado de dados pessoais” (Lei nº 13.709, 2018: 10). Além disso, o artigo também inclui os casos em que as decisões são tomadas com base em predições, que insistem em prever perfis de crédito, pessoais e profissionais com base nos padrões e tendências retiradas de um grande conjunto de dados. A lei reconhece as possibilidades de falhas e vieses virem implícitos nas predições e sugestões de tomadas de decisão fornecidas pelos algoritmos, por isso, garante aos sujeitos, alvos de discriminações e injustiças, que tenham não só os seus resultados revistos por humanos, como também determina que sejam dadas “informações claras e adequadas a respeito dos critérios e dos procedimentos utilizados para a decisão automatizada” (Lei nº 13.709, 2018: 10).

Além disso, é a LGPD (Lei Geral de Proteção de Dados) que garante a transparência dos dados processados pelos algoritmos e impede o uso de alguns deles que, por algum motivo, possam prejudicar os indivíduos em determinadas análises. A ANPD (Autoridade Nacional de Proteção de Dados) estabelece que, nos casos em que se verificam vieses e incoerências nos resultados e predições algorítmicas, é necessário que sejam realizadas auditorias para a “verificação de aspectos discriminatórios em tratamento automatizado de dados pessoais” (Lei nº 13.709, 2018: 10). Portanto, nessas situações, a existência de uma legislação e a criação de órgãos de supervisão, como a ANPD, é fundamental para que possamos pensar na possibilidade de governança dos algoritmos e em como promover decisões justas, transparentes e explicáveis.

Cathy O’Neil (2016), em seu livro *“Weapons of math destruction: how big data increases inequality and threatens democracy”*, defende que precisamos exigir a transparência dos resultados e predições fornecidas pelos algoritmos. Temos o direito

de saber quais dados e informações pessoais estão sendo utilizadas para que o algoritmo realize a análise ou sugira determinada tomada de decisão. Nesse sentido, a criação de legislações e órgãos de regulação é de suma importância para que se determine a auditoria e a revisão dos processos de tomadas de decisão. Segundo a autora, nos casos em que sistemas de IA são utilizados para realizar a concessão de crédito aos consumidores, os EUA dispõem de legislações específicas para que se alcance resultados justos e igualitários. Dessa forma, o FCRA (Fair Credit Reporting Act) garante que o consumidor tenha acesso e verifique a veracidade dos dados processados pelos algoritmos. Do mesmo modo, o ECOA (Equal Credit Opportunity Act) proíbe que dados como raça, gênero (O'neil, 2016) e estado civil sejam correlacionados com o objetivo de traçar um perfil e indicar os indivíduos aptos a receberem empréstimos.

A reivindicação de posse de um sigilo econômico e industrial não pode continuar a servir de argumento para que empresas e instituições impeçam o prosseguimento de ações que solicitam que os algoritmos sejam auditáveis. Da mesma forma, entendemos que o argumento de que os sistemas de IA são "autônomos" e "incontestáveis", não pode continuar funcionando como pretexto para que os seus algoritmos permaneçam opacos, inexplicáveis e incontroláveis (Pasquinelli & Joler, 2020). A ideia de que os algoritmos são "neutros" e "objetivos" não deve permitir que fechemos os olhos ou desconsideremos a possibilidade de falhas e constrangimentos sejam gerados pelo uso da IA. Até porque, como argumentamos, a concepção do que é "neutralidade" e "objetividade" deve ser tensionada a partir da noção de "objetividade forte", isto é, ao invés de negar a presença de interesses, valores e outros fatores de ordem pessoal, as práticas de construção de artefatos sociotécnicos devem, na verdade, reconhecer conscientemente a existência desses aspectos subjetivos, além de suas interferências

sobre nossas práticas e visões de mundo. Portanto, cientes disso, precisamos estar atentos à operatoriedade algorítmica, a base de dados com a qual foi treinada e a forma com que os algoritmos foram programados. Quem inscreve ao algoritmo o "passo a passo" ou apresenta os dados considerados "imprescindíveis" para definir se um sujeito é um "bom" professor ou um funcionário inapto para o trabalho? Em que conceito, visão de mundo ou experiências se baseia a definição de "bom" ou "mau" professor ou mesmo a aptidão necessária para um bom desempenho no emprego? Além disso, quem define se a IA é ou não eficiente para a realização dessas análises? A expectativa é a de que as respostas a essas perguntas, inspiradas nos insights proporcionados pelas epistemologias feministas, nos possibilitem esclarecer e tornar mais justas as decisões baseadas nas sugestões emitidas pela máquina.

Os vieses e a controvérsia em torno da imparcialidade dos algoritmos podem estar contidos já no momento de seleção dos dados que servirão como base de treinamento. "O ato de selecionar uma fonte de dados em vez de outra é a marca profunda da intervenção humana" (Pasquinelli & Joler, 2020: 8) no processo de construção da IA e uma das principais causas de discriminação e preconceito emitidos pelos algoritmos. Por exemplo, se, ao treinarmos um sistema de IA criado para a seleção de candidatos a vagas de emprego, utilizamos uma base de dados em que os últimos contratados do ano foram homens brancos, com idade entre 25 a 40 anos e estado civil solteiro, não poderíamos ficar surpresos se, ao ser posto em funcionamento, o sistema discriminar homens negros, mulheres e demais indivíduos cuja idade seja inferior a 25 ou superior a 40 anos ou cujo estado civil seja o de casado, separado ou divorciado. Uma base de dados pode, então, conter inúmeros vieses (Pasquinelli & Joler, 2020) e estar permeada de preconceitos e discriminações de cunho racista, sexista e misógino, e, nesses casos, os algoritmos de *Machine Learning* podem contribuir para que tais estruturas continuem de forma permanente em nossas

sociedades. Logo, trata-se de um exemplo emblemático de como as tecnologias incorporam qualidades políticas, nos termos de Winner (1986).

Os conceitos a partir dos quais os algoritmos são programados podem não só esconder inclinações morais (O'neil, 2016), como também corresponder a perfis muito específicos, por exemplo, de indivíduos esperados para que ocupem determinados postos de trabalho. Por isso, a tentativa de regular o uso de sistemas de IA em instituições públicas deve não apenas envolver dimensões técnicas, mas também incorporar certa diversidade nas diferentes etapas de seu desenvolvimento até o momento de sua implementação. Na verdade, a Resolução n° 332, publicada no dia 21 de agosto de 2020, ao defender a diversidade e "a participação representativa [...] em todas as etapas do processo, tais como planejamento, coleta e processamento de dados, construção, verificação, validação e implementação" (CNJ, 2020: 4) da IA, já evidencia a necessidade da diversidade na construção de IAs no setor judiciário brasileiro. A participação de um grupo diversificado em termos raça, gênero, profissão e etnia pretende possibilitar um treinamento mais cuidadoso da IA, que considere a pluralidade de indivíduos e as inúmeras realidades e pontos de vista, para que, assim, os algoritmos forneçam resultados mais justos e comprometidos com os princípios de equidade e justiça. O conceito ou ponto de vista econômico de eficiência em termos de custo e agilidade não pode continuar a servir como critério exclusivo para a implementação e uso da IA. É necessário, ao invés disso, considerar a noção de eficiência que está em disputa e a necessidade de que o uso de sistemas de IA também estejam comprometidos com a diversidade, equidade e justiça.

Considerações finais

A inserção da IA em procedimentos de análises e tomadas de decisão se dá a partir da pretensão de isenção em relação a valores e em conformidade com ideais de eficiência e progresso tecnológico (Feenberg, 2010). Apesar da construção desse "imaginário sociotécnico" (Jasanoff, 2015), a utilização de algoritmos comumente resulta na discriminação de indivíduos e na fixação de padrões específicos que reproduzem desigualdades historicamente produzidas em termos de gênero, raça e classe. É premente, diante disso, "abrir a caixa-preta" dos sistemas automatizados de tomadas de decisão. Iniciativas nesse sentido envolvem, em parte, várias dimensões: técnica, legal, e sobretudo, política e epistêmica, já que, tal como argumentamos, é preciso levar em consideração quem, como e com quais objetivos são programados treinados os sistemas automatizados de tomadas de decisão. Em outras palavras, é preciso compreender a situacionalidade e as perspectivas dos sujeitos envolvidos nos processos de programação.

Do ponto de vista técnico, é preciso compreender que o resultado ou a omissão dos algoritmos em determinadas análises requer o conhecimento dos dados utilizados no *Machine Learning*, bem como a observância dos conceitos a partir dos quais os algoritmos operam para realizarem a análise e o processamento de dados. Já a dimensão legal está relacionada à relevância da criação de normas, legislações e órgãos governamentais de regulação e supervisão (Almeida & Doneda, 2018) que estejam comprometidos com a responsabilidade social e tecnológica e com a transparência dos resultados e sugestões que os algoritmos fornecem para a tomada de decisão. Por fim, quanto aos aspectos políticos e epistêmicos, destacamos que em uma sociedade em que padrões de atitudes específicos são atribuídos às mulheres, opiniões sexistas podem ser "amplamente sustentadas por instituições e pela sociedade como um todo" (Harding, 2007: 165) e, com isso, serem apresentadas para o processamento dos algoritmos como sendo formas de conhecimento e

classificações “objetivas”, baseadas em observações, cálculos matemáticos, testes e revisões por pares.

Nesses casos, apoiados nas epistemólogas feministas, argumentamos que, para que não haja discriminação com base nessas visões, é necessário nos questionarmos de que lugar se observa, “quem revisa” e “para quem é feito”, assim como a diversidade de sujeitos no processo de construção da IA. Isso porque, num contexto em que as mulheres ainda são minorias nos laboratórios destinados a produção ciência e tecnologia, sobretudo nas áreas de computação e engenharia, é possível que tenhamos modelos e classificações algorítmicas utilizadas em processos de decisão sendo definidas por homens com base em visões masculinas sobre, por exemplo, o que é uma “boa” mãe, que, provavelmente, não irá maltratar o seu filho. Seria o caso também da associação direta entre um espaço como o da cozinha a um indivíduo específico que, segundo Sayuri (2019), seria uma mulher.

“Tecnologias têm qualidades políticas?” (Winner, 1986: 1). Essa foi uma das perguntas que orientaram nossa perspectiva diante das controvérsias decorrentes das análises algorítmicas. Salientamos serem de suma importância os questionamentos realizados sobre a política incorporadas nas tecnologias e nos objetos técnicos que se apresentam controversos em relação aos seus objetivos e diante do contexto e da situação em que são construídos. Nesse ponto, são bastante prolíficas as proposições das epistemologias feministas quanto à situacionalidade dos sujeitos envolvidos nos processos de pesquisa e inovação. Compreendemos que tais reflexões se tornam fundamentais se o objetivo é fazer com que tenhamos decisões justas frente à digitalização crescente dos processos sociais e de tomada de decisão através de sugestões de sistemas de IA.

Referências bibliográficas

Aragão, F. A. A e Benevides, P. S. (2019), “Governamentalidade algorítmica e Big data: o uso da correlação de dados como critério de tomada de decisão”, IV Simpósio internacional lavits: assimetrias e (in)visibilidades: vigilância, gênero e raça, Salvador, 26-28 junho. Disponível em: https://lavits.org/wp-content/uploads/2019/12/Araujo_Benevides-2019-LAVITS.pdf

Arrieta, A. B. *et al* (2020), “Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI”, *Information fusion,* v. 58, pp. 82-115.

BBC (2020), “Algoritmo roubou meu futuro’: solução para ‘Enem britânico’ na pandemia provoca escândalo”. Disponível em: https://www.bbc.com/portuguese/internacional-53853627

Bauer, M. W. (2008), “Análise de conteúdo clássica: uma revisão”. In: Bauer, M. and Gaskell, G. *Pesquisa qualitativa com texto, imagem e som: um manual prático*. Petrópolis, Vozes.

Bioni, B. R.; Rielli, M. e Kitayama, M. (2021), *O legítimo interesse na LGPD: quadro geral e exemplos de aplicação.* São Paulo, Associação Data Privacy Brasil de Pesquisa.

Bruno, F. (2013), *Máquinas de ver, modos de ser: vigilância, tecnologia e subjetividade.* Porto Alegre, Sulina.

Bruno, F. (2008), “Monitoramento, classificação e controle nos dispositivos de vigilância digital”, *Revista FAMECOS*, n. 36, pp. 10-16.

Canclini, N. G. (2020), Ciudadanos reemplezados por algoritmos, México, Calas.

Conselho Nacional de Justiça (2020). Resolução nº 332, de 21 de agosto. Disponível em: https://atos.cnj.jus.br/atos/detalhar/3429

Cordellat, A. (2018), “Doença mental materna ainda é associada a ser ‘uma mãe ruim”. *El País*. Madrid, 11 de maio de 2018. Disponível em: https://brasil.elpais.com/brasil/2018/05/05/ciencia/1525542192_173178.html#?rel=listaapoyo

Dagnino, R. P. (2008), *Neutralidade da ciência e determinismo tecnológico: um debate sobre a tecnociência*. Campinas, Editora da Unicamp.

Dastin, J. (2018), “Amazon scraps secret AI recruiting tool that showed bias against women”. *Reuters*. São Francisco - EUA, 10 de outubro de 2018. Disponível em: https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

Decreto nº 10.332, de 28 de abril de 2020 (2020), Brasília – DF. Disponível em: https://www.in.gov.br/web/dou/-/decreto-n-10.332-de-28-de-abril-de-2020-254430358

Decreto nº 10.609, de 26 de janeiro de 2021 (2021), Brasília – DF. Disponível em: http://www.planalto.gov.br/ccivil_03/_Ato2019-2022/2021/Decreto/D10609.htm#art18

Deleuze, G. (2013), “Post-scriptum sobre as sociedades de controle”. In: Deleuze, G. *Conversações*. 3. ed. Trad. Peter Pál Pelbart. São Paulo, Editora 34, p. 223-230.

Doneda, D. e Almeida, V. A. F. (2018), “O que é a governança dos algoritmos?”, In Bruno, F., Cardoso, B, Kanashiro, M., Guilhon, L. e Melgaço, L. (Eds.). *Tecnopolíticas da vigilância*. São Paulo: Boitempo, pp. 141-148.

Echarri, M. (2021), “150 demissões em um segundo: os algoritmos que decidem quem deve ser mandado embora”. *El País*. Barcelona, 10 de outubro de 2021. Disponível em: https://brasil.elpais.com/tecnologia/2021-10-10/150-demissoes-

em-um-segundo-assim-funcionam-os-algoritmos-que-decidem-quem-deve-ser-mandado-embora.html

Feenberg, A. (2010), “Racionalização democrática, poder e tecnologia”, In Needer, R. T. (org.). *Ciclo de conferências Andrew Feenberg*.

Ferreira, F. (2020), “Inteligência Artificial atua como juiz, muda estratégia de advogado e ‘promove’ estagiário”. Folha de S. Paulo. São Paulo, 10 de março. Disponível: https://www1.folha.uol.com.br/poder/2020/03/inteligencia-artificial-atua-como-juiz-muda-estrategia-de-advogado-e-promove-estagiario.shtml

Fox Keller, E. (1991), *Reflexiones sobre género y ciencia*. Valencia, Edicions Alfons el Magnànim.

Gillespie, T. (2018), “A relevância dos algoritmos”, *Parágrafo*, v. 6, n. 1, pp. 95-121.

Grossman, L. O. (2018), “Big Data do Governo Federal levou ao corte de 5 milhões do Bolsa Família”. *Convergência Digital*. São Paulo, 16 de abril. Disponível em: https://ciab.convergenciadigital.com.br/cgi/cgilua.exe/sys/start.htm?infoid=47765&sid=11

Haraway, D. (2009), “Saberes localizados: a questão da ciência para o feminismo e o privilégio da perspectiva parcial”, *Cadernos Pagu*, n. 5, pp. 7–41.

Haraway, D. (2013), “Manifesto ciborgue: ciência, tecnologia e feminismo-socialista no final do século XX”, In Haraway, D.; Kunzru, H. e Tadeu, T. *Antropologia do ciborgue: as vertigens do pós-humano*. Belo Horizonte, Autêntica.

Harding, S. (1987), “Introduction: is there a feminist method?” In Harding, S. (Ed.). *Feminism and Methodology: social science issues*. Bloomington, USA, Indiana University.

Harding, S. (2007), “Gênero, democracia e filosofia da ciência", *RECIIS – Revista Eletrônica de Comunicação, Informação & Inovação em Saúde*. Rio de Janeiro, v. 1, n. 1, pp. 163-168.

Harding, S. (2008), “Postcolonial science and technology studies. Are there multiple sciences?” In Harding, S. *Sciences from Below: Feminisms, Postcolonialities, and Modernities.* Durham and London: Duke University Press.

Harding, S. (2019), “Objetividade mais forte para ciências exercidas a partir de baixo”, *Construção: arquivos de epistemologia histórica e estudos de ciência*, n. 5, pp. 143-162.

Jasanoff, S. (2015), “Future imperfect: Science, technology, and the imaginations of modernity”, In Jasanoff, S.; Kim, S-H. *In Dreamscapes of modernity: Sociotechnical imaginaries and the fabrication of power*, pp. 1-33. Chicago, University of Chicago.

Kim, J. H. (2004), “Cibernética, ciborgues e ciberespaço: notas sobre as origens da cibernética e sua reinvenção cultural". *Horiz. antropol*., Porto Alegre, v. 10, n. 21, pp. 199-219.

Lei nº 13.709, de 14 de agosto de 2018 (2018), Brasília – DF. Disponível em: http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2018/Lei/L13709compilado.htm

Lima, M. P. (2013), “As mulheres na Ciência da Computação”, *Revista Estudos Feministas*, Florianópolis, v. 21, n. 3, pp. 793-816.

Lucena, P., A. C. de (2019), “Policiamento preditivo, discriminação algorítmica e racismo: potencialidade e reflexos no Brasil”, *VI Simpósio internacional lavits: assimetrias e (in)visibilidades: vigilância, gênero e raça*, Salvador, 26-28 junho. Disponível em: https://lavits.org/wp-content/uploads/2019/12/Lucena-2019-LAVITSS.pdf

Maia, M. M. (2016), “Limites de gênero e presença feminina nos cursos superiores brasileiros no campo da computação”, *Cadernos Pagu*, Campinas, n. 46, pp. 223-224.

Mendes, L. e Mattiuzzo, M. (2019), “Discriminação algorítmica: conceito, fundamento legal e tipologia”, *Direito Público*, 16(90), pp. 39-64.

Marx, L. e Smith, M. (1994), “Introduction”, In Marx, L and Smith, M. (Eds.). *Does Technology Drive History? The Dilemma of Technological Determinism.* Cambridge (Massachusetts), MIT Press.

O’neil, C. (2016), *Weapons of math destruction: how big data increases inequality and threatens democracy*. New York, Crown Publishers.

Pascual, M. G. (2019), “Quem vigia os algoritmos para que não sejam racistas ou sexistas?” *El País*. Madrid, 17 de março de 2019. Disponível em: https://brasil.elpais.com/brasil/2019/03/18/tecnologia/1552863873_720561.html

Pasquinelli, M. (2017), “Machines that Morph Logic: Neural Networks and the Distorted Automation of Intelligence as Statistical Inference”, *Logic Gate: the Politics of the Artifactual Mind* [Online]. Disponível em: https://www.glass-bead.org/article/machines-that-morph-logic/?lang=enview

Pasquinelli, M. e Joler, V. (2020), “O manifesto Nooscópio: Inteligência Artificial como Instrumento de Extrativismo do Conhecimento”. *Lavits*, Rio de Janeiro, 30 de julho. Disponível em: https://lavits.org/o-manifesto-nooscopio-inteligencia-artificial-como-instrumento-de-extrativismo-do-conhecimento/?lang=pt

Plamadeala, C. (2021), “Vigilância de dossiê e a produção de arquivos como ferramentas de controle". [Entrevista concedida a] Paulo Faltay. *Lavits,* Rio de Janeiro, 23 de abril. Disponível em: https://lavits.org/entrevista-cristina-plamadeala/?lang=pt

Rouvroy, A. (2012), “The end(s) of critique: data-behaviourism vs. due-process”, In Rouvroy, A. *Privacy Due Process and the Computational Turn*. Philosophers of Law Meet Philosophers of Technology, Routledge.

Sakai, J.; Galdino, M. e Burg, T. (2021), “Governance recommendations: Use of Artificial Intelligence by public authorities", *Transparência Brasil*. São Paulo. Disponível em: https://www.transparencia.org.br/downloads/publicacoes/Governance_Recommendations.pdf

Salas, J. (2017), “Se está na cozinha é uma mulher: como os algoritmos reforçam preconceitos”. *El País.* Madrid, 23 de setembro. Disponível em: https://brasil.elpais.com/brasil/2017/09/19/ciencia/1505818015_847097.html

Sayuri, J. (2019), “Os algoritmos tentam identificar seu gênero, mas muitas vezes reforçam representações sexistas”, *Revista Trip*. São Paulo, 26 de abril. Disponível em: https://revistatrip.uol.com.br/tpm/os-algoritmos-tentam-identificar-seu-genero-mas-muitas-vezes-reforcam-representacoes-sexistas

Subramanian, B.; Foster, L.; Harding, S.; Roy, D. e Tallbear, K. (2017), “Feminism, Postcolonialism, Technoscience”, In Felt, Ulrike et al. (Eds.) *Handbook of Science and Technology Studies* - Fourth edition. Cambridge, MA e Londres: The MIT Press, pp. 407-434.

UNESCO (2021), *UNESCO Science Report: the Race Against Time for Smarter Development,* Unesco Institute for Statistics and animated by Values Associates. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000375429

Vicentin, D. (2022), “Esboço para o aprofundamento da Inteligência Artificial”, *Ideias.* Campinas - SP, v. 13, p. 1-28.

Winner, L. (1986), *Artefatos têm política?* (Traduzido por Fernando Manso). Chicago: The University of Chicago Press, pp. 19-39.

Zuboff, S. (2019), “The age of surveillance capitalismo: The fight for a human future at the new frontier of power”. New York: Public Affairs, 2018. Resenha de: Evangelista, R. *Surveilance & Society*, New York, v. 17, n. 1/2, pp. 246-251.

Artículo recibido el 1 de marzo de 2022
Aprobado para su publicación el 1 de junio 2023