

**Isadora Valadares  
Assunção**

*Universidade de São Paulo*

isadoravaladares@usp.br

## INTELIGÊNCIA ARTIFICIAL E O TRANSTORNO DO ESPECTRO AUTISTA: O POTENCIAL DISCRIMINATÓRIO

---

### RESUMO

O presente artigo é uma revisão integrativa de literatura que visa evidenciar o potencial discriminatório advindo de aplicações diagnósticas e terapêuticas para pessoas com deficiência, em especial para indivíduos com Transtorno do Espectro Autista (TEA). Tal potencial advém da formalização simplificada de complexos problemas da realidade, da escolha dos dados de treinamento - que podem conter vieses, sub- ou super-representações - e da ponderação de quais atributos serão analisados pelo modelo algorítmico. Além do risco discriminatório, tais modelos contribuem para a desumanização de pessoas com TEA por retirar sua autonomia e agência, tornando o quadro ético de IA insuficiente para mitigar o risco de discriminação algorítmica. Perspectivas técnicas de mitigação de risco também mostram-se insatisfatórias pela diversidade de apresentação do TEA, uma característica intrinsecamente não-observável e não-mensurável. Portanto, o aumento da representatividade de pessoas com TEA durante o desenvolvimento de tais ferramentas faz-se necessário, em uma perspectiva de desenvolver com pessoas autistas, não para elas.

**Palavras-chave:** Preconceito; Transtorno do Espectro Autista; autismo; discriminação social; Inteligência Artificial.

## ARTIFICIAL INTELLIGENCE AND AUTISM SPECTRUM DISORDER: THE DISCRIMINATORY POTENTIAL

---

### ABSTRACT

The present article is a integrative literature review aimed at evidencing the discriminatory potencial of diagnostic and therapeutical applications for people with disabilities, especially for people with Autism Spectrum Disorder (ASD). This potencial derives from the simplified formalization of complex real-life problems, from the choice of the training data - which might have biases, under- or over-representations - and from the ponderation of which attributes will be considered by the algorithmic model. Apart from the discriminatory risks, these systems also contribute to the dehumanization of people with ASD by removing their autonomy and agency, making IA ethics insufficient for mitigating the algorithmic discrimination risks. Technical perspectives of risk mitigation are also unsatisfactory because of the diversity of ASD presentations, as it is an intrinsically non-observable and non-measurable characteristic. Thus, an increase of the representativity of people with ASD during the development of these tools makes itself necessary, in a perspective of designing with people with autism, not for them.

**Keywords:** bias; Autism Spectrum Disorder; autism; social discrimination; artificial intelligence.

#### Correspondência/Contato

revistaneurodiversidade@gmail.com  
<https://www.revistaneurodiversidade.com/>

#### Editores responsáveis

Daniele Pendeza

Lucas Pontes

## 1 INTRODUÇÃO

O Transtorno do Espectro Autista pode ser caracterizado, segundo o DSM-V, por “déficits persistentes na comunicação social e na interação social” junto a “padrões restritos e repetitivos de comportamento, interesses ou atividades”, que existam precocemente e que causem prejuízo funcional ao indivíduo (APA, 2014). Clinicamente, o diagnóstico se dá por meio da realização de entrevistas com o paciente e seus familiares, além da aplicação de avaliações neuropsicológicas (Cardoso & Nogueira, 2021).

Recentemente, aplicações de inteligência artificial (IA) têm sido desenvolvidas para diagnosticar ou auxiliar no tratamento de questões associadas ao transtorno, como dificuldades de socialização. No entanto, aspectos como o potencial discriminatório de tais tecnologias não vêm sendo discutidas durante o processo de elaboração dessas. Logo, o presente trabalho visa responder ao questionamento sobre se as tecnologias de inteligência artificial voltadas ao Transtorno do Espectro Autista são discriminatórias, em especial, delimitando se elas têm potencial prejudicial em relação às pessoas com autismo.

Para melhor compreender a interseção entre a inteligência artificial e o campo da saúde, é primeiro necessário examinar o que é aquela. Segundo Norvig e Russell (2013), as definições de IA variam em torno de considerações seja sobre o comportamento ou pensamento análogo ao de um ser humano seja sobre o comportamento ou pensamento racional - isto é, que visa alcançar o melhor resultado esperado em determinado contexto. Dentre o campo de estudo da IA, destacam-se a existência de sistemas especialistas e sistemas baseados em aprendizagem de máquina.

Os sistemas especialistas (expert systems) ou sistemas baseados em conhecimento (knowledge systems) baseiam-se em uma combinação da compreensão teórica sobre um problema com uma gama de regras de resolução de problemas que se mostraram úteis em um domínio específico do conhecimento (Norvig & Russel, 2013). Dependem, portanto, da extração do conhecimento teórico de um especialista humano na área alvo, o que acarreta problemas como a dificuldade de se manter o sistema atualizado e a possibilidade de que vieses sejam introduzidos no sistema ao se basear na compreensão de um ser humano sobre uma área do conhecimento. Ademais, tem baixa flexibilidade na resolução de problemas, já que se são apresentados a problemas atípicos em relação à área de domínio, em geral são incapazes de resolvê-los (Luger, 2013).

Por outro lado, há os sistemas baseados em aprendizagem de máquina (machine learning), que induzem uma hipótese a partir da experiência passada visando relacionar os valores de entrada de

---

---

objetos (inputs) ao valor de saída destes (outputs), englobando métodos como redes neurais, modelos probabilísticos e algoritmos genéticos (Faceli et al., 2011).

Nas últimas décadas, aplicações de IA vêm sendo crescentemente usadas na área da saúde, seja para tarefas burocráticas quanto para auxílio a decisões clínicas, como de diagnóstico e tratamento a partir da análise de dados biomédicos, clínicos, fisiológicos e comportamentais (Preininger & Wang, 2019). Exemplos de sistemas de IA aplicados à saúde são: (i) detecção de nódulos de pulmão com maior rapidez que um painel de radiologistas especialistas (Koo et al., 2012); (ii) detecção de câncer de pele com acurácia equivalente à de dermatologistas (Esteve et al., 2017); (iii) aplicativos de saúde móvel que são utilizados para diagnosticar asfixia infantil no momento do parto (Ndiomu et al., 2017); (iv) sistemas especialistas para acompanhar dinamicamente a realização de primeiros socorros e melhorar sua qualidade (Ertl & Christ, 2007); entre outros. Na área da saúde mental, destacam-se algoritmos que, com o processo de digital phenotyping (Preininger & Wang, 2019), analisam dados comportamentais para aferir humor, cognição e padrões de comportamento, diagnosticando transtornos como o Transtorno do Déficit de Atenção e Hiperatividade (TDAH) (Ra et al., 2018).

No entanto, contrastando com tais experiências bem-sucedidas, diversas outras aplicações falharam em suas funções almejadas, como a IBM Watson, por dificuldades de generalização, de análise de dados incompletos ou ambíguos e pela dificuldade de formalização dos problemas médicos (Strickland, 2019). Ainda, tanto os sistemas especialistas quanto os de aprendizagem de máquina são, em geral, opacos, dificultando a análise de sua acurácia e sua adoção por equipes médicas, que exigem a explicabilidade das decisões e das etapas de raciocínio lógico realizadas pelos sistemas de IA para validá-las (Preininger & Wang, 2019).

Ademais, surgem preocupações com a privacidade dos usuários, já que a análise de IA depende da extração de quantidades massivas de dados para alimentar sistemas de aprendizagem de máquina, podendo tais dados serem usados para tornar aspectos íntimos analisáveis e comercializáveis, em um processo de renderização da experiência humana (Zuboff, 2020). Tal processo depende da objetificação de seres humanos, a serem vistos como outros a serem observados e mensurados, suscitando dilemas éticos como a dificuldade de que sistemas de IA ofereçam tratamentos humanos aos pacientes, por não participarem de um intercâmbio socioemocional com os pacientes (Coeckelbergh, 2010). Ainda, tal objetificação pode levar a análises acríticas dos sistemas de IA, que falham em reconhecer seus vieses e potenciais discriminatórios.

## 2 MÉTODO

Uma revisão integrativa da literatura foi conduzida, a partir da busca de artigos sobre o tema listados na plataforma Google Scholar relacionados às palavras-chave “Autism Spectrum Disorder”, “artificial intelligence” e “machine learning”. Um total de 5.490 artigos foram identificados na busca inicial. Artigos foram excluídos se fossem (i) artigos puramente teóricos e metodológicos sobre inteligência artificial, (ii) artigos com menos de 10 participantes, (iii) artigos que não usavam majoritariamente inteligência artificial, (iv) artigos que não fossem meta-análises ou revisões integrativas de literatura. Ao final, 28 artigos foram incluídos para análise do texto completo.

## 3 O TRANSTORNO DO ESPECTRO AUTISTA E A INTELIGÊNCIA ARTIFICIAL

Segundo Judy Singer, uma das criadoras do movimento da neurodiversidade, o advento da internet abriu caminho para uma coevolução dinâmica e simbiótica do homem com a máquina, facilitando a comunicação de autistas e promovendo um novo nicho para pessoas no espectro, cuja existência no ciberespaço pode ser comparada metaforicamente à figura de um ciborgue (Singer, 2017). Além de ferramentas que facilitam a interação social de autistas, tecnologias computacionais que centram-se na relação homem-máquina, como a inteligência artificial, vêm sendo desenvolvidas como ferramentas diagnósticas e terapêuticas para o Transtorno do Espectro Autista.

No entanto, o processo é demorado, o que prejudica a aplicação de terapias que visam desenvolver habilidades centrais durante janelas de desenvolvimento nas quais a neuroplasticidade é maior (Bosa, 2006). Segundo dados do National Health System (NHS), a idade média de diagnóstico é de 4 anos e 7 meses (Brett et al., 2016), mesmo havendo evidência de que é possível realizá-lo consistentemente a partir dos 18 meses (McCarthy, 2019). A demora persiste também nos diagnósticos tardios, ou seja, aqueles realizados com indivíduos acima de 18 anos e que tem duração média de 361 a 480 dias, a depender da faixa etária (NHS Digital, 2021). Ademais, o processo diagnóstico é permeado por vieses de gênero (Baron-Cohen, 2002; Constantino et al, 2012; Haney, 2016; Maenner et al., 2018), raça (Begeer et al., 2009; White et al., 2019), localização geográfica e classe (Brett et al., 2016), o que dificulta o acesso de minorias a atendimento especializado e amplifica a sub-representação de tais grupos, que muitas vezes recebem diagnósticos errados ou tardios.

Em uma tentativa de superar tais vieses e diminuir o tempo médio de diagnóstico, ferramentas diagnósticas baseadas em inteligência artificial foram desenvolvidas. Dentre tais iniciativas,

---

---

destacam-se três grupos: (i) o uso de IA aplicada a imagens de ressonâncias magnéticas, (ii) a coleta e análise de dados comportamentais e (iii) a aplicação automatizada de escalas diagnósticas.

Quanto ao primeiro grupo, em uma meta-análise de 119 estudos sobre o tema, Xu et al. (2021) concluíram que, desde 2015, o número de estudos baseados em ressonâncias magnéticas funcionais vêm aumentando em detrimento de ressonâncias estruturais, em grande parte devido ao compartilhamento de bancos de dados públicos como o Autism Brain Data Exchange (ABIDE).

Entretanto, ainda há uma correlação inversamente proporcional entre a acurácia dos sistemas e sua amostra, o que indica dificuldades de generalização dos modelos. Como exemplo, é de se notar que, em estudos com 1500 ou menos indivíduos, a acurácia máxima foi de 96.8% (Bi et al., 2018), enquanto que, nos estudos com mais de 1500 indivíduos na amostra, a acurácia máxima foi de 64% (Ke et al., 2020). Isso indica dificuldades na análise das imagens, o que pode se dar pela variedade de aparelhos e configurações nas quais os exames foram realizados, em especial em amostras maiores. Ainda, é de se destacar que os bancos de dados são compostos majoritariamente por imagens advindas de indivíduos do sexo masculino e pessoas com autismo nível 01 de suporte, o que pode dificultar o diagnóstico em mulheres e de pessoas com os demais níveis de suporte a partir dos modelos treinados nestas amostras, excluindo-as (Xu et al., 2021).

Quanto ao uso de dados comportamentais, destacam-se tentativas de diagnosticar o autismo por meio da gravação de vídeos caseiros dos indivíduos com sua posterior análise por meio da IA. Em Tariq et al. (2018), 162 vídeos de dois minutos foram analisados, com acurácia de mais de 90% em distinguir pessoas no espectro de crianças com desenvolvimento normal. Ademais, o modelo usado fornece uma indicação quantitativa do risco de autismo ao invés de uma classificação binária, o que favorece seu uso para triagens de possíveis pessoas com TEA. No entanto, testes realizados com crianças em Bangladesh mostraram uma redução na acurácia para 75%, indicando as particularidades culturais no diagnóstico (Tariq et al., 2019).

No mesmo sentido, destaca-se o aplicativo GuessWhat desenvolvido por pesquisadores de Stanford, que grava vídeos das crianças jogando as charadas propostas e usa os vídeos para fins diagnósticos (Collins, 2019). Ainda, é de se mencionar a ferramenta Canvas Dx, desenvolvida pela empresa Cognoa e com aprovação da Federal Drug Administration (FDA) dos Estados Unidos. A ferramenta visa auxiliar no diagnóstico de crianças de 18 a 72 meses por meio da realização de um questionário com os pais por meio de um aplicativo móvel, a análise de dois vídeos da criança e um questionário realizado por um médico especialista (Hackett, 2021).

Além de sistemas que se baseiam na observação do comportamento filmado dos indivíduos, é de se destacar estudos que utilizam movimentos motores das mãos e dos membros superiores, com acurácia de 86.7% (Li et al., 2017) e 96.7% (Crippa et al., 2015), respectivamente. Por fim, em Liu et al. (2016), a diferença no escaneamento facial foi utilizada para diferenciar entre crianças com e sem TEA com acurácia de 88.51%, chegando-se à conclusão de que aquelas demoravam-se mais analisando o olho esquerdo das faces e estas focavam mais no olho direito.

Por fim, quanto ao uso de escalas diagnósticas, é de se mencionar duas abordagens: a de aplicação automatizada das escalas com posterior análise por mecanismos de IA e a redução de itens nos questionários. No que tange à primeira abordagem, em Song et al. (2019) identificou-se que as aplicações de IA permitiram o diagnóstico precoce antes dos 3 anos de idade. Por exemplo, em Bussu et al. (2018), usaram-se as tabelas Mullen Scales of Early Learning (MSEL) e Vineland Adaptable Behavior Scale (VABS) para diagnosticar crianças de 14 meses. Ademais, tal abordagem ainda permite a detecção de outros transtornos do neurodesenvolvimento e a realização de diagnósticos diferenciais, como entre TEA e TDAH por meio da aplicação de 5 dos 65 itens da Social Responsiveness Scale (SRS) (Duda et al., 2016). Já quanto a segunda abordagem, estudos conseguiram reduzir a quantidade de itens da escala Autism Diagnostic Interview - Revised (ADI-R) em 92% (Wall et al., 2012) e eliminar 21 itens dos módulos da escala Autism Diagnostic Observational Schedule (ADOS) (Kosmicki et al., 2015), mantendo alta acurácia.

Ferramentas de IA também são usadas com funções terapêuticas, sendo que há uma dificuldade intrínseca entre separá-las das diagnósticas já que, muitas vezes, as mesmas ferramentas observam os indivíduos para diagnosticá-los e, com base nos dados coletados, modificam seu comportamento se adaptando ao comportamento do sujeito (Erden et al., 2020). Entre as ferramentas terapêuticas de IA aplicadas ao TEA, destaca-se o ensino de habilidades não-verbais, como expressões faciais, tom de voz, gestos e movimentos, além de habilidades sociais (Naser et al., 2021).

Alguns exemplos destas ferramentas são o TeachTown, que utiliza princípios da Análise do Comportamento Aplicada (ABA) para ensinar habilidades sociais, e o Emotion Trainer, que ensina pessoas a reconhecer emoções em terceiros (Jaliawala & Khan, 2019). Além disso, é de se mencionar iniciativas que utilizam robôs adaptativos para ensinar emoções e habilidades sociais para as crianças, como o NAO Robot (Rudovic et al., 2018), além de robôs que ensinam habilidades como imitação e atenção compartilhada, como o Robota (Robins et al., 2005).

Segundo Jaliawala e Khan (2019), tais ferramentas têm o potencial de favorecer a difusão do tratamento e da educação de pessoas com TEA, sendo intervenções eficazes por reforçarem apenas

---

---

os comportamentos desejados. No entanto, em uma revisão sistemática a partir de 31 estudos na área, concluiu-se que nenhum deles mostrou evidências concretas de sucesso terapêutico nem se mostrou escalável, em especial por serem aplicados apenas a pequenas amostras, por faltar um banco de dados amplo de treinamento, pelos robôs terem inteligência socioemocional insuficiente para interações sociais naturais e por serem pouco explicáveis para a equipe terapêutica (Jaliawala & Khan, 2019). No projeto Aurora, por exemplo, no qual foi usado o robô Robota, a previsibilidade do robô o permitiu ser um bom mediador de interações sociais, mas seu uso não conseguiu incentivar interações espontâneas e proativas (Robins et al., 2005).

Tanto as ferramentas diagnósticas quanto as terapêuticas oferecem sérios dilemas éticos, dentre os quais destaca-se, segundo Keyes, a medicalização do autismo como algo a ser corrigido sob uma ótica neoliberal de menor despendimento de tempo e recursos, sem se questionar sobre a humanidade e dignidade dos autistas, que podem ser, como resultado, alvo de discriminação algorítmica (Keyes, 2020).

#### 4 DISCRIMINAÇÃO ALGORÍTMICA E A APLICAÇÃO DE IA AO TEA

Segundo Selbst e Barocas (2016), a IA sempre envolve discriminações estatísticas, uma vez que almeja diferenciar indivíduos de forma racional, atribuindo aos sujeitos características compartilhadas por um determinado grupo que seja estatisticamente similar a ele. Entretanto, além dessa discriminação atuarial, a IA pode afetar certos indivíduos de forma desproporcional. Nesse sentido, como pontuam Mendes e Mattiuzzo (2019), a discriminação algorítmica tanto pode advir de afirmações estatisticamente inconsistentes quanto de afirmações lógicas, mas que são injustas, tomando o indivíduo unicamente como parte de um grupo e o desfavorecendo como resultado.

A partir de tal pontuação, Selbst e Barocas (2016) apresentam uma divisão de três momentos que podem gerar a discriminação algorítmica: (i) a definição da variável de destino (target variable); (ii) a coleta dos dados de treinamento; (iii) a definição de quais atributos serão analisados pelo modelo.

Inicialmente, é necessário pontuar que a variável de destino de um modelo é o seu resultado de interesse. Sua conceituação pode gerar discriminações, haja vista que decorre da formalização matemática de complexos e controversos problemas reais (Selbst & Barocas, 2016). Um exemplo disso é um algoritmo que determine quem é um bom empregado. A definição da variável de destino, o bom empregado, é controversa, sendo que, enquanto alguns pensam que bom empregado é aquele que fica por anos na mesma empresa, outros consideram que é aquele que sempre chega antes do horário. Caso

um bom empregado seja definido desta última forma, aqueles que morem longe do emprego, o que possivelmente correlaciona-se com renda, e empreguem maior tempo no trânsito serão desproporcionalmente prejudicados, já que são mais propensos a atrasos (Borgesius, 2018).

No que tange ao autismo, ferramentas diagnósticas dependem da formalização do que é ser autista, a variável de destino. Como o autismo tem etiologia obscura, o seu diagnóstico baseia-se não em testes objetivos e sim em observações clínicas, o que dificulta a formalização da variável de destino e insere significativa subjetividade nos modelos, em especial naqueles que dependam da definição de especialistas (expert systems). A situação é agravada ao considerar-se que o autismo é um transtorno de espectro e que, por isso, tem diversas manifestações, o que torna o problema de sua formalização ainda mais complexo. Mesmo que os modelos se atenham aos critérios diagnósticos do DSM, é de se pontuar que estes mudam, sendo que pessoas que antes se adequavam ao diagnóstico podem passar a não fazê-lo em edições posteriores. Por exemplo, 40% daqueles que eram diagnosticados com autismo clássico, síndrome de Asperger ou Transtorno Global do Desenvolvimento - Sem Outra Especificação (TGD-SOE) segundo o DSM-IV não continuariam a ter diagnóstico de TEA segundo os critérios do DSM-V (McPartland et al., 2012). Logo, a formalização dos modelos diagnósticos de TEA em si gera a possibilidade de que determinados indivíduos sejam discriminados, o que é agravado ao se considerar que o acesso a diversos direitos sociais é contingente ao diagnóstico formal.

Outro momento que pode ensejar a discriminação algorítmica é a coleta e a escolha dos dados de treinamento. Se o banco de dados for enviesado, pode levar à reprodução de vieses estruturais, como ocorreu com o algoritmo de contratação de médicos do St. George's Hospital, no Reino Unido, que desfavoreceu mulheres e negros por tal viés estar presente nos dados de treinamento (Selbst & Barocas, 2016).

Ademais, se inferências forem realizadas a partir de bancos de dados nos quais determinados grupos estejam sub- ou super-representados, pode ocorrer discriminação contra eles (Selbst & Barocas, 2016). Exemplo disso é a aplicação de policiamento preditivo conhecida como heat list de Chicago. Como pessoas negras historicamente têm maior contato com a polícia por causa do racismo estrutural, tal grupo era super-representado nos dados de treinamento, recebendo maiores escores de risco e maior atenção policial, em um ciclo retroalimentado de feedback (Ferguson, 2017).

Como já mencionado, o diagnóstico de autismo é permeado por vieses de gênero, classe, raça e localização geográfica. Portanto, se os dados de treinamento baseiam-se em diagnósticos já realizados, tais vieses serão reproduzidos e amplificados pela IA, excluindo grupos marginalizados e aumentando a taxa de falsos negativos. O cenário é agravado pela sub-representação de diversos

---

---

grupos nas amostras, o que dificulta, por exemplo, que o modelo se adeque a particularidades geográficas (como citado acima, com a diminuição da acurácia após a aplicação da ferramenta diagnóstica estadunidense em Bangladesh) ou diferenças na apresentação do TEA de acordo com o gênero ou com o grau de suporte, por exemplo. É importante ressaltar que a coleta de mais dados não é suficiente, já que oferece significativos riscos à privacidade, podendo tais bancos de dados serem usados para outras funções, possivelmente excludentes e discriminatórias (White, 2019). As ferramentas terapêuticas também podem ensejar discriminação algorítmica por dados enviesados, já que reproduzirá padrões de tratamento que muitas vezes são viciados em relação à condição socioeconômica do paciente, com acesso diferenciado a terapias.

Ainda, outra possibilidade de geração de discriminação algorítmica é na definição dos atributos que serão ponderados pelo algoritmo em suas análises. Se estes envolverem uma simplificação da realidade, falhando em atingir a granularidade dos aspectos que influenciam uma decisão, pode criar-se modelos discriminatórios baseados em inferências estatisticamente corretas, mas injustas, já que não levam em consideração todos os atributos do mundo real. Um exemplo disso é um algoritmo de contratação que atribua maior peso à universidade de origem do candidato. Haja vista a dificuldade de acesso de minorias a universidades de renome, tais grupos seriam discriminados, mesmo tendo as mesmas qualidades dos outros candidatos, uma vez que o modelo falhará em considerar outros aspectos além das credenciais acadêmicas (Selbst & Barocas, 2016).

No caso das ferramentas diagnósticas para autismo, a dependência de estereótipos do que é autismo pode gerar a supervalorização de certos atributos, como contato visual ou ausência de fala, em uma simplificação das diversas formas que o TEA pode assumir. Já quanto às aplicações terapêuticas, pode-se valorizar certas ações como medição de sucesso terapêutico ou de adesão à terapia em detrimento de uma avaliação geral e holística do comportamento.

Assim, verifica-se que há alto potencial discriminatório do uso da IA aplicado ao autismo, o que é agravado, no caso das ferramentas diagnósticas, pela possibilidade de ocorrer um diagnóstico não-solicitado, como na ferramenta patenteada pela Hirevue que visa diagnosticar pessoas com deficiência durante os processos seletivos (supostamente para não discriminá-las) (Whittaker et al., 2019). Isto pode aumentar o mascaramento (masking) dos indivíduos, receosos da estigmatização que ainda acompanha o diagnóstico de autismo (Keyes, 2020). Ainda, mesmo que este seja solicitado, pela opacidade inerente aos sistemas de IA, o diagnóstico ainda será de difícil explicabilidade e de difícil questionamento, podendo aumentar a taxa de falsos positivos e negativos.

Além disso, é de se notar que tanto as ferramentas diagnósticas quanto as terapêuticas tratam o autismo como algo a ser identificado e corrigido, adequando-se à normalidade que faz parte da própria lógica binária da linguagem de máquina. Assim, desumanizam-se os autistas como pessoas sem autonomia e agência devido às suas dificuldades de comunicação, dificultando que se posicionem e reajam a situações de discriminação (Keyes, 2020). Segundo Keyes (2020), o framework ético de IA tem como ponto central a justiça algorítmica (fairness), entendida como ausência de discriminação. No entanto, essa depende de que todos os indivíduos sejam igualmente considerados humanos e autônomos, o que torna tal quadro teórico insuficiente para analisar os danos dignitários advindos da discriminação algorítmica decorrente da aplicação de IA ao TEA. Nesse sentido, o autor observa que:

Profissionais de IA estão enquadrando discursivamente certas populações como não-humanas e sem agência em paralelo com quadros éticos que dependem em humanidade e agência para endereçar danos. [...]. Se a comunicação autista é inválida - se autistas são inválidos - e endereçar danos é uma prática discursiva entre pessoas válidas, então danos contra pessoas autistas não podem ser endereçados neste enquadramento. (tradução livre) (Keyes, 2020, p. 20).

Por fim, é de se notar que, além do framework ético de IA ser insuficiente para lidar com a discriminação algorítmica contra pessoas com deficiência como aquelas com autismo, também o são perspectivas técnicas que dependem de características observáveis, como a paridade de classificação - que consiste em equalizar predições entre grupos - e a paridade anticlassificação - que consiste em omitir atributos protegidos (Tomasev et al., 2021). Tais técnicas têm desafios em se adequar a características não-observadas, como deficiências invisíveis, que podem não ter dados por: (i) questões logísticas, como nem todas as pessoas declararem que as tem, tornando os bancos de dados incompletos e inconsistentes; (ii) questões legais, por leis antidiscriminação e de proteção à pessoa com deficiência; (iii) serem intrinsecamente não-mensuráveis, em especial aquelas que se apresentam em um espectro como o TEA.

Assim, se perspectivas técnicas e éticas mostram-se insuficientes, são necessários maiores estudos sobre novas diretrizes que abordem os danos causados pela discriminação algorítmica aplicada a pessoas com deficiência, como pessoas com TEA, primordialmente pelo aumento da representatividade de tais grupos durante o desenvolvimento de tais ferramentas, em uma concepção de desenvolver com, não para as pessoas com TEA (designing with, not for). Para tal, posteriores pesquisas são fundamentais para descobrir pontos específicos que gerem discriminação de pessoas com TEA por tecnologias de IA, além de explicitar como abordá-los concretamente, reduzindo o impacto discriminatório de tais tecnologias.

---

---

## CONSIDERAÇÕES FINAIS

Com o avanço da IA na área da saúde, foram desenvolvidas ferramentas diagnósticas para o TEA a partir de imagens de ressonâncias magnéticas, de dados comportamentais e de escalas diagnósticas, junto a ferramentas terapêuticas que focam no ensino de habilidades sociais e não-verbais, de reconhecimento de emoções, de atenção compartilhada e de imitação.

Tais ferramentas apresentam significativo potencial discriminatório pela formalização matemática simplificada de questões complexas, pela seleção dos dados de treinamento e pela escolha dos atributos a serem ponderados pelo modelo. Esses aspectos são agravados pela possibilidade dos algoritmos serem usados para outras funções, com diagnósticos não-solicitados. Ainda, os riscos discriminatórios são amplificados pela opacidade dos sistemas, que consistem em caixas-pretas (black-boxes) sem explicabilidade (explainability) adequada, o que dificulta sua adoção em larga escala.

Ainda, por tratar o autismo como algo a ser diagnosticado e corrigido, em uma visão medicalizada, as ferramentas retiram agência e autonomia das pessoas com TEA, desumanizando-as e dificultando que lutem contra situações de opressão algorítmica em um framework ético cuja centralidade reside justamente na autonomia e na agência presentes no conceito de fairness. Perspectivas técnicas de mitigação de risco discriminatório também são insuficientes já que o TEA é uma característica não-observada com dificuldade de mensuração, além de existirem obstáculos logísticos e legais à coleta de dados para que tais técnicas sejam realizadas.

Assim, uma nova diretriz anti-discriminação faz-se primordial, majoritariamente pela inclusão de pessoas com TEA durante o desenvolvimento de tais sistemas, em uma perspectiva de designing with, not for.

## REFERÊNCIAS

American Psychiatric Association. (2014). DSM-5: Manual diagnóstico e estatístico de transtornos mentais. Artmed Editora.

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *Calif. L. Rev.*, 104, 671.

Baron-Cohen, S. (2002). The extreme male brain theory of autism. *Trends in cognitive sciences*, 6(6), 248-254.

Begeer, S., El Bouk, S., Boussaid, W., Terwogt, M. M., & Koot, H. M. (2009). Underdiagnosis and referral bias of autism in ethnic minorities. *Journal of autism and developmental disorders*, 39(1), 142.

Bi, X. A., Liu, Y., Sun, Q., Luo, X., Tan, H., Chen, J., & Zeng, N. (2019). The genetic-evolutionary random support vector machine cluster analysis in autism spectrum disorder. *IEEE Access*, 7, 30527-30535.

Brett, D., Warnell, F., McConachie, H., & Parr, J. R. (2016). Factors affecting age at ASD diagnosis in UK: no evidence that diagnosis age has decreased between 2004 and 2014. *Journal of autism and developmental disorders*, 46(6), 1974-1984.

Bosa, C. A. (2006). Autismo: intervenções psicoeducacionais. *Brazilian Journal of Psychiatry*, 28, s47-s53.

Bussu G, Jones EJH, Charman T, Johnson MH, Buitelaar JK BASIS Team. Prediction of autism at 3 years from behavioural and developmental measures in high-risk infants: a longitudinal cross-domain classifier analysis. *J Autism Dev Disord*. 2018;48:2418–2433. doi: 10.1007/s10803-018-3509-x.

Cardoso, A., Nogueira, M. (2021). *Atenção Interdisciplinar ao Autismo*. Belo Horizonte: Ampla.

Carvalho, A., Faceli, K., Lorena, A., & Gama, J. (2011). *Inteligência Artificial – uma abordagem de aprendizado de máquina*. Rio de Janeiro: LTC, 45.

Crippa A, Salvatore C, Perego P, Forti S, Nobile M, Molteni M, et al. Use of machine learning to identify children with autism and their motor abnormalities. *J Autism Dev Disord*. 2015;45:2146–2156. doi: 10.1007/s10803-015-2379-8.

Coeckelbergh, M. (2010). Health care, capabilities, and AI assistive technologies. *Ethical theory and moral practice*, 13(2), 181-190.

Constantino, J. N., & Charman, T. (2012). Gender bias, female resilience, and the sex ratio in autism. *Journal of the American Academy of Child and Adolescent Psychiatry*, 51(8), 756-758.

Collins, N. (2019). A smartphone app to treat and track autism. Recuperado de: A smartphone app to treat and track autism | Stanford News. Acesso em 01 de fev. de 2022.

Duda M, Ma R, Haber N, Wall DP. Use of machine learning for behavioral distinction of autism and ADHD. *Transl Psychiatry*. 2016;6:e732. doi: 10.1038/tp.2015.221.

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *nature*, 542(7639), 115-118.

Ertl, L., and Christ, F., Significant improvement of the quality of bystander first aid using an expert system with a mobile multimedia device. *Resuscitation* 74:286-295, 2007.

Erden, Y. J., Hummerstone, H., & Rainey, S. (2021). Automating autism assessment: What AI can bring to the diagnostic process. *Journal of Evaluation in Clinical Practice*, 27(3), 485-490.

Ferguson, A. G. (2017). *The rise of big data policing*. New York University Press.

---

---

Hackett, M. (2021). FDA greenlights Cognoa's tool to detect autism spectrum disorder in kids. Recuperado de: <https://www.mobihealthnews.com/news/fda-greenlights-cognoas-tool-detect-autism-spectrum-disorder-kids>. Acesso em: 01 de fev. de 2022.

Haney, Jolynn L. Autism, females, and the DSM-5: Gender bias in autism diagnosis. *Social Work in Mental Health*, v. 14, n. 4, p. 396-407, 2016.

Imm, P., White, T., & Durkin, M. S. (2019). Assessment of racial and ethnic bias in autism spectrum disorder prevalence estimates from a US surveillance system. *Autism*, 23(8), 1927-1935.

Jaliaawala, M. S., & Khan, R. A. (2020). Can autism be catered with artificial intelligence-assisted intervention technology? A comprehensive survey. *Artificial Intelligence Review*, 53(2), 1039-1069.

Ke, F., Choi, S., Kang, Y. H., Cheon, K. A., & Lee, S. W. (2020). Exploring the structural and strategic bases of autism spectrum disorders with deep learning. *IEEE Access*, 8, 153341-153352.

Keyes, O. (2020). Automating autism: Disability, discourse, and Artificial Intelligence. *The Journal of Sociotechnical Critique*, 1(1), 8.

Koo, C. W., Anand, V., Girvin, F., Wickstrom, M. L., Fantauzzi, J. P., Bogoni, L., ... & Ko, J. P. (2012). Improved efficiency of CT interpretation using an automated lung nodule matching program. *American Journal of Roentgenology*, 199(1), 91-95.

Kosmicki JA, Sochat V, Duda M, Wall DP. Searching for a minimal set of behaviors for autism detection through feature selection-based machine learning. *Transl Psychiatry*. 2015;5:e514. doi: 10.1038/tp.2015.7.

Li B, Sharma A, Meng J, Purushwalkam S, Gowen E. Applying machine learning to identify autistic adults using imitation: an exploratory study. *PLoS One*. 2017;12:e0182652. doi: 10.1371/journal.pone.0182652.

Liu W, Li M, Yi L. Identifying children with autism spectrum disorder based on their face processing abnormality: a machine learning framework. *Autism Res*. 2016;9:888–898. doi: 10.1002/aur.1615.

Luger, G. F. (2013). *Inteligência Artificial*.

Maenner, M. J., Shaw, K. A., Bakian, A. V., Bilder, D. A., Durkin, M. S., Esler, A., ... & Cogswell, M. E. (2021). Prevalence and characteristics of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, United States, 2018. *MMWR Surveillance Summaries*, 70(11), 1.

McCarthy, C. (2019). How early can you — and should you — diagnose autism? Recuperado de: <https://www.health.harvard.edu/blog/how-early-can-you-and-should-you-diagnose-autism-2019082317653>>. Acesso em 01 de fev. de 2022.

McPartland, J. C., Reichow, B., & Volkmar, F. R. (2012). Sensitivity and specificity of proposed DSM-5 diagnostic criteria for autism spectrum disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, 51(4), 368-383.

Naser, M., Nasr, M. M., & Shehata, L. H. (2021). Artificial Intelligence (AI) And Autism Spectrum Disorder (ASD) Literature review. *International Journal of Progressive Sciences and Technologies*, 29(2), 396-415.

NHS Digital. (2021). Autism Waiting Time Statistics. Recuperado de: Autism Waiting Time Statistics - NHS Digital. Acesso em 01 de fev. de 2022.

Norvig, P., & Russell, S. (2004). *Inteligência artificial*. Editora Campus, 20.

Onu, C. C., Udeogu, I., Ndiomu, E., Kengni, U., Precup, D., Sant'Anna, G. M., ... & Opara, P. (2017). Ubenwa: Cry-based diagnosis of birth asphyxia. *arXiv preprint arXiv:1711.06405*.

Wang, F., & Preininger, A. (2019). AI in health: state of the art, challenges, and future directions. *Yearbook of medical informatics*, 28(01), 016-026.

Ra, C. K., Cho, J., Stone, M. D., De La Cerda, J., Goldenson, N. I., Moroney, E., ... & Leventhal, A. M. (2018). Association of digital media use with subsequent symptoms of attention-deficit/hyperactivity disorder among adolescents. *Jama*, 320(3), 255-263.

Robins, B., Dautenhahn, K., Te Boekhorst, R., & Billard, A. (2005). Robotic assistants in therapy and education of children with autism: can a small humanoid robot help encourage social interaction skills?. *Universal access in the information society*, 4(2), 105-120.

Rudovic, O., Lee, J., Dai, M., Schuller, B., & Picard, R. W. (2018). Personalized machine learning for robot perception of affect and engagement in autism therapy. *Science Robotics*, 3(19).

Singer, J. (2017). *Neurodiversity: The birth of an idea*.

Song, D. Y., Kim, S. Y., Bong, G., Kim, J. M., & Yoo, H. J. (2019). The use of artificial intelligence in screening and diagnosis of autism spectrum disorder: a literature review. *Journal of the Korean Academy of Child and Adolescent Psychiatry*, 30(4), 145.

Strickland, E. (2019). IBM Watson, heal thyself: How IBM overpromised and underdelivered on AI health care. *IEEE Spectrum*, 56(4), 24-31.

Tariq Q, Daniels J, Schwartz JN, Washington P, Kalantarian H, Wall DP. Mobile detection of autism through machine learning on home video: a development and prospective validation study. *PLoS Med*. 2018;15:e1002705. doi: 10.1371/journal.pmed.1002705.

Tomasev, N., McKee, K. R., Kay, J., & Mohamed, S. (2021). Fairness for Unobserved Characteristics: Insights from Technological Impacts on Queer Communities. *arXiv preprint arXiv:2102.04257*.

Wall DP, Dally R, Luyster R, Jung JY, Deluca TF. Use of artificial intelligence to shorten the behavioral diagnosis of autism. *PLoS One*. 2012;7:e43855. doi: 10.1371/journal.pone.0043855.

White, J. J. (2020). Fairness of AI for people with disabilities: problem analysis and interdisciplinary collaboration. *ACM SIGACCESS Accessibility and Computing*, (125), 1-1.

---

---

Whittaker, M., Alper, M., Bennett, C. L., Hendren, S., Kaziunas, L., Mills, M., ... & West, S. M. (2019). Disability, bias, and AI. AI Now Institute.

Xu, M., Calhoun, V., Jiang, R., Yan, W., & Sui, J. (2021). Brain imaging-based machine learning in autism spectrum disorder: methods and applications. *Journal of neuroscience methods*, 361, 109271. <https://doi.org/10.1016/j.jneumeth.2021.109271>

Zuboff, S. (2021). *A era do capitalismo de vigilância*. Editora Intrínseca.

Zuiderveen Borgesius, F. (2018). *Discrimination, artificial intelligence, and algorithmic decision-making*.

---

***Isadora Valadares Assunção***

Estudante de direito da Universidade de São Paulo, Coordenadora do Núcleo de Proteção de Dados (NPD-TechLab) da USP, Participante do Núcleo de Estudos Internacionais da USP (NEI-USP).

**Recebido em 01 de fevereiro de 2022**

**Aceito em 25 de abril de 2022**