

Dados e Inteligência Artificial: os efeitos jurídicos da discriminação algorítmica

LUCAS CORTIZO*

Resumo: O texto propõe-se a fazer uma análise sobre conceitos básicos de inteligência, inteligência artificial (IA), discriminação e discriminação algorítmica, interligando-os a fim de analisar juridicamente os respetivos efeitos tecnológicos. Para isto, o artigo realiza um levantamento de como o uso dos algoritmos de IA (que tomam decisões automatizadas) podem adotar uma postura discriminatória, sem prejuízo do princípio da neutralidade tecnológica. Neste levantamento sobre uma possível discriminação pelo próprio sistema de IA, um dos elementos analisados são os dados. Dados estes que são utilizados no treinamento dos sistemas de decisão, mas no cenário dos mesmos estarem enviesados, analisa-se a possibilidade do algoritmo emitir uma decisão discriminatória. Neste sentido, casos concretos são trazidos para evidenciar na prática como uma discriminação algorítmica tem efeitos jurídicos, atingindo inclusive Direitos Humanos e Princípios Fundamentais.

Palavras-chave: *inteligência, artificial, discriminação, direito, algoritmo.*

Abstract: The text proposes an analysis on basic concepts of intelligence, artificial intelligence (AI), discrimination and algorithmic discrimination, interconnecting them in order to legally analyse their respective technological effects. To do so, the article

* Advogado na Autoridade Nacional de Proteção de Dados do País de Malta; Representante de Malta na European Data Protection Board (EDPB) nos seguintes expert subgroups: International Transfers e Social Media; Representante de Malta na Global Privacy Assembly e na Common Thread Network; Participante da estratégia nacional de implementação da Inteligência Artificial em Malta; Especialista e Mestre em Direito e Tecnologia pela Universidade do Minho, Braga, Portugal; Graduado em Direito pela Universidade Federal de Pernambuco, Brasil; Coautor do livro União Europeia Interop 2019. Capítulo sobre Blockchain e e-Government, tendência da tecnologia na Administração pública; Fundador do Podcast “Direito Digital Cast”; Professor de Direito e Tecnologia em diversas instituições; Membro da European Artificial Intelligence Alliance, Bélgica.

conducts inquiries on how the use of AI algorithms (which make automated decisions) can adopt a discriminatory behaviour, without prejudice to the principle of technological neutrality. In such analysis on the possible discrimination by the AI system itself, one of the elements analysed is data. These data are used to train the decision systems, but in the scenario that they are biased, it is analysed the possibility of the algorithm issuing a discriminatory decision. In this sense, concrete cases are brought to show in practice how an algorithmic discrimination has legal effects, affecting even Human Rights and Fundamental Principles.

Keywords: *intelligence, artificial, discrimination, law, algorithm.*

Introdução

Nos últimos anos, uma área que tem sido bastante comentada e discutida é a inteligência artificial (IA), entretanto não foi ainda atingido um consenso sobre o que seria entendido por IA¹. E dentro desta extensa área do conhecimento, o presente texto propõe-se, inicialmente, a fazer uma análise tanto do conceito de inteligência, como também aprofundar sobre a modalidade de inteligência que se enquadra como artificial. A análise conceitual também almeja definir o termo discriminação com intuito de, posteriormente, delimitar a discriminação algorítmica, para assim interligar conceitos de IA e discriminação, estudando os seus possíveis efeitos jurídicos.

Para estabelecer os mencionados efeitos jurídicos, o presente artigo realiza uma análise de quais seriam as causas para um eventual uso de algoritmos de IA² que possam resultar numa postura discriminatória³.

¹ De acordo com BERTOLINI: “Regular inteligência artificial requer sua definição. Ainda não existe nenhum consenso sobre o que é entendido por IA”. Ver mais em BERTOLINI, Andrea. *Artificial Intelligence and Civil Liability*. Policy Department for Citizens’ Rights and Constitutional Affairs. Parlamento Europeu, 2020. Tradução livre.

² Algoritmos de IA são aqueles que tomam decisões automatizadas.

³ Fala-se em “uso de algoritmos” e não algoritmos em si em respeito à neutralidade tecnológica, prevista no Considerando 15 do Regulamento Geral sobre a Proteção de Dados (RGPD), cf. União Europeia. Regulamento (UE) 2016/679 do Parlamento Europeu e do Conselho de 27 de abril de 2016 relativo à proteção das pessoas singulares no que diz respeito ao tratamento de dados pessoais e à livre circulação desses dados

Neste levantamento sobre uma possível discriminação pelo próprio produto do sistema de IA, um dos elementos analisados são os dados. Dados estes que são a matéria-prima no treinamento dos sistemas de decisão. Por isso é necessário estabelecer até que ponto o treinamento de IA com dados enviesados pode conduzir o algoritmo a uma decisão discriminatória.

E diante de uma possível relação entre dados de treinamento e resultados discriminatórios, o presente artigo traz casos concretos para analisar, na prática, se uma discriminação algorítmica pode ter repercussões na esfera jurídica dos indivíduos e atingir inclusive Direitos Humanos e Princípios Fundamentais.

1. Considerações iniciais sobre inteligência

Primeiramente, é importante traçar as noções basilares do que seria inteligência, não apenas do ponto de vista artificial, mas também do ponto de vista humano. A partir do momento que o próprio ser humano se intitulou o único ser racional do planeta, o monopólio da sabedoria foi definido pelo mesmo, sendo o único ser vivo a falar sobre inteligência ou considerar que é inteligente.

No momento em que seres humanos começaram a ter a consciência de que eram seres inteligentes, passou a existir um interesse crescente em desenvolver o conhecimento científico⁴. Uma competição histórica entre povos, que movidos por desenvolvimento científico, passaram a medir, comparar, e competir por conhecimento – todavia, ainda antes de amplamente definir o conceito de inteligência. Não obstante, pela forte relação que existe entre desenvolvimento científico e poder de conquista, historicamente, as grandes fontes da inteligência, representadas pelas grandes mentes da ciência, sempre foram cobiçadas pelas grandes civilizações⁵.

⁴ WARNER defende que esta corrida acontece ao longo da História e remete a ciclos. Ver mais em WARNER, Michael. *The Rise and Fall of Intelligence: an international security History*. Georgetown University Press, 2014.

⁵ Referência ao fenômeno do “intellectual theft” cf. MULKAY, Michael J. “Norms and ideology in science”. *Social Science Information*, v. 15, n. 4-5, 1976, p. 637-656.

É difícil encontrar um parâmetro objetivo para medir a inteligência. Principalmente, no contexto da geopolítica e da corrida científica, o que vai existir é uma análise de possível ganho, ou uma análise dos benefícios de uma inteligência individual que se juntará a outras para criar uma inteligência geral. E este foi o critério que fez com que cientistas, artistas e várias mentes notáveis das mais diversas áreas fossem disputadas por impérios na busca de expandir o seu domínio, não apenas territorial, mas intelectual⁶.

A Guerra Fria representou uma parte da História em que cresceu uma cobiça pelo desenvolvimento científico⁷, mas a própria II Guerra Mundial, em momento anterior, também foi resolvida pela ciência: a exemplo da invenção de Alan Turing, o computador, que foi uma das mais valiosas armas do Reino Unido contra a Alemanha⁸. A obra de Turing, por sua vez, traça um paralelo entre o conceito de inteligência sob a perspectiva da máquina, ao fazer a indagação de se seria possível as máquinas pensarem.

O paradoxo que surge é que, à medida que as grandes civilizações e impérios foram evoluindo, nunca houve tanto esforço no desafio de conceituar e avaliar de forma objetiva o que seria inteligência e qual sua extensão em cada pessoa. Os Estados Unidos e a União Soviética ao disputar uma grande mente não estavam prioritariamente em busca de provar, em caráter objetivo, o grau de inteligência de algum cientista estrangeiro, nem se aquele indivíduo oferecia um risco para os demais pensadores nacionais. E é por isso que a inteligência não costumava ser vista como uma ameaça interna, mas sim como objeto de desejo para aquisição externa pelos grandes países. Então, desde a revolução científica, a inteligência passou a ser vista como *commodity*⁹.

⁶ Para aprofundar sobre a corrida científica, principalmente no contexto da Guerra Fria, veja-se NELSON, Richard R.; ROMER, Paul M. "Science, economic growth, and public policy". *Challenge*, v. 39, n. 1, 1996, p. 9-21.

⁷ A cobiça pelo desenvolvimento científico foi marcada por um patrocínio à pesquisa e ao desenvolvimento acadêmico, muitas vezes distorcido e aumentado pelos Estados Unidos e União Soviética, nos termos de HOUNSHELL, David. "The Cold War, RAND, and the generation of knowledge, 1946-1962". *Historical Studies in the Physical and Biological Sciences* 27.2 (1997): 237-267.

⁸ Sobre este contexto, recomenda-se a obra: HODGES, Andrew. *Alan Turing: The Enigma*. Random House, 2012.

⁹ Neste sentido, PRITCHARD defende que historicamente o conhecimento sempre foi um valor distintivo. Veja-se mais em PRITCHARD, Duncan. "The value of knowledge". *The Harvard Review of Philosophy*, v. 16, n. 1, 2009, p. 86-103.

Contudo, quando a inteligência chega à máquina, ela passa a ser atacada sob outro ponto de vista¹⁰. Se há quem considere a obra de Turing como a primeira que traz as bases do amplo ramo que seria posteriormente denominado Inteligência Artificial (IA)¹¹, a preocupação em definir e medir a capacidade das máquinas pensarem acompanhou a evolução deste conceito de IA. Indo além, construiu-se um ideal de apreensão, pois pela primeira vez foi criada uma tecnologia que poderia ultrapassar a inteligência humana¹².

Por isso, ao surgir uma inédita entidade que computa e processa informações de uma forma mais rápida que a humana, passou a haver uma preocupação em desenvolvê-la, mas simultaneamente entender seus limites. Mas, entender os limites de uma tecnologia pelo senso comum tem sido um tarefa que usa uma “régua humana” limitante: devem os limites de qualquer inteligência ser mesmo a própria inteligência humana? Até porque o conceito de inteligência surgiu por autoproclamação¹³, e em consequência disto, existiu sozinho por centenas de anos em monopólio humano.

Neste momento, pode-se questionar que plantas e animais também são inteligentes. E é de admirar a complexidade e as características únicas que alguns deles possuem, mas o presente objetivo não é fechar uma conceituação de algo que jamais foi pacificamente conceituado. Busca-se mostrar que pela primeira vez, o ser humano sentiu-se ameaçado por reconhecer que o monopólio da inteligência, autoproclamado pelo “único animal racional”¹⁴, agora poderia ser um oligopólio em que o ser

¹⁰ Ponto de vista que supõe a inteligência artificial como ameaça, sustentado por grandes nomes como Stephen Hawking. Ver mais em: CELLAN-JONES, Rory. “Stephen Hawking warns artificial intelligence could end mankind”. BBC news, v. II, 2014, p. 10.10.

¹¹ Neste sentido, MOOR, James (Ed.). *The Turing test: the elusive standard of artificial intelligence*. Springer, Science & Business Media, 2003. & “the father of AI and computer science” em GUO, Ting. “Spirituality as reconceptualisation of the self: Alan Turing and his pioneering ideas on artificial intelligence”. *Culture and Religion*, v. 16, n. 3, 2015.

¹² Neste sentido, há várias obras como KURZWEIL, Ray. *The age of spiritual machines: When computers exceed human intelligence*. Penguin, 2000. & FRENCH, Robert M. *The Turing Test: the first 50 years. Trends in cognitive sciences*, v. IV, n. 3, 2000, p. 115-122.

¹³ Nenhum outro ser vivo jamais se proclamou inteligente cf. PEARCE, John M. “Animal learning and cognition: an introduction”. Psychology press, 2013.

¹⁴ Conceito de Deleuze acerca da unicidade da racionalidade humana. Ver mais em: ROCHA, André Menezes. “Formação da razão na ética de Espinosa, segundo Deleuze”. *Cadernos Espinosanos*, v. XVI, 2007, p. 89-100.

humano deveria aprender a viver em relação simbiótica com os dispositivos inteligentes criados por si próprio.

Ao longo dos tempos, alguns métodos foram criados para objetivar a inteligência e um dos mais conhecidos é o método do “Quociente de Inteligência” (Q.I.)¹⁵. Apesar de ter sido bastante utilizado, atualmente esta metodologia é considerada obsoleta por apenas exigir uma das várias formas que a inteligência pode assumir¹⁶. O teste de Q.I. preza pela inteligência lógico-matemática, favorecendo grandes mentes como Tesla, Newton e Einstein. O próprio Einstein teria afirmado que todos nós somos gênios, mas se você julgar um peixe pela sua capacidade de escalar uma árvore, ele passará o resto da vida acreditando não ser¹⁷.

É por conta dessas múltiplas facetas que a inteligência pode assumir que muito se fala da Teoria das múltiplas inteligências de Gardner¹⁸. Listando as diversas formas, Gardner traz a inteligência linguística, musical, espacial, corporal cenestésica, intrapessoal, interpessoal, naturalista e existencial. E pela pluralidade deste conceito, faz-se necessário delimitar as perspectivas pela qual o âmbito do conceito de inteligência pode ser observada, o qual se passa a fazer a seguir.

2. Os limites da Inteligência Artificial

Se o parâmetro da inteligência for apenas uma de suas facetas, representadas pela excelência ao jogar xadrez, pode-se responder à pergunta de Turing: sim, as máquinas já pensam e já ultrapassaram o ser humano no critério inteligência. Mas se este critério de inteligência for um grande leque de diversas representações, a resposta do vencedor da competição

¹⁵ Quociente de inteligência (QI), conforme GOULD, Stephen Jay, *The mismeasure of man*. Nova York, Norton, 1996, p. 30-35.

¹⁶ Ideia desenvolvida por Gould que na obra “a falsa medida do homem” demonstra os principais erros da investigação científica da inteligência humana escorada no quociente de inteligência (QI). Veja-se em GOULD, Stephen Jay, *ibid*.

¹⁷ COLOMBO, Jorge A. “A critical view of the quest for brain structural markers of Albert Einstein’s special talents (a pot of gold under the rainbow)”. *Brain Structure and Function*, v. 223, n. 5, 2018, p. 2515-2518.

¹⁸ Ver mais GARDNER, Howard. *A Multiplicity of Intelligences*. Scientific American, 1998.

virtual “inteligência humana *versus* inteligência artificial” não se limita a uma resposta simples.

Ademais, merece ainda levantar o potencial da IA para cada vertente de cada tipo de inteligência, o que pode levar a diferentes respostas. Por exemplo, poderá a IA compor músicas de uma forma melhor que os compositores humanos? Existem diversas soluções no mercado musical que oferecem auxílio por IA para músicos, sugerindo e complementando o trabalho de composição musical¹⁹.

Uma das várias plataformas com inteligência artificial é a Amper²⁰. Apesar de ser considerada uma IA que compõe músicas, é apenas um dos vários programas semelhantes que os músicos podem usar para compor músicas, incluindo Magenta do Google, Flow Machines da Sony e Jukedeck²¹.

O que existe de comum nestas plataformas de IA musicais é que elas precisam de escolhas humanas. As máquinas não criam músicas por si só, mas precisam de intervenção humana prévia para definir os seus fins e meios, a exemplo da escolha da tonalidade da música, instrumentos específicos, as batidas por minuto do ritmo, etc. Então, a IA oferece uma gama de ferramentas que trabalham ritmos e estilos criados e definidos pelos humanos. Se esta capacidade for considerada “inteligência musical”²², a IA vai apresentar uma das facetas descritas acima.

Se as máquinas puderem pensar, e, ainda, se apresentarem características que sejam enquadradas nas mais diversas inteligências que existem, vai ser discutido se aquilo que a IA produz corresponde a um processo de criatividade²³.

¹⁹ BARRETT, Maura; WARD, Jacob. *AI can now compose pop music and even symphonies. Here's how composers are joining in*. Disponível em: <<https://www.nbcnews.com/mach/science/ai-can-now-compose-pop-music-even-symphonies-here-s-ncna1010931>> Acedido a: 29.11.20

²⁰ BARRETT, Maura; WARD, Jacob. *Ibid.*

²¹ BRIOT, Jean-Pierre; PACHET, François. “Deep learning for music generation: challenges and directions”. *Neural Computing and Applications*, v. 32, n. 4, 2020, p. 981-993.

²² Tradução livre para *Musical Intelligence*, conforme definido por ROADS, Curtis. “Artificial intelligence and music”. *Computer Music Journal*, v. 4, n. 2, 1980, p. 13-25.

²³ Para aprofundar sobre este questionamento da criatividade na IA, veja-se BODEN, Margaret A. “Creativity and artificial intelligence”. *Artificial intelligence magazine*, v. 103, n. 1-2, 1998, p. 347-356.

Ainda sob o desígnio específico de uma IA para fins musicais, um algoritmo de aprendizagem que – ao ser treinado com todas as músicas de Beethoven – cria uma sinfonia inédita que poderia ter sido composta pelo músico e não foi, não há consenso se isto torna o algoritmo criativo. Até pelo facto de que a criatividade, assim como a inteligência, é um conceito amplo e subjetivo²⁴, não se deseja aqui debater a qualidade de uma composição de IA, mas se tal composição representa um exercício inteligente por parte da IA.

Desta forma, os algoritmos de IA oferecem ferramentas cada vez mais completas nas mais diversas áreas do conhecimento, não se limitando a cálculos computacionais, mas enveredando em searas artísticas, seja para auxílio dos indivíduos da área, seja como entidades inteligentes e criativas.

Ainda merece destacar o presente conceito de inteligência mais aprofundado, à luz da inteligência humana como parâmetro, o que difere do conceito de *smart devices* desenvolvido nos últimos anos. Apesar de em tradução literal significar aparelhos inteligentes, o grau de profundidade objetivado não pretende alcançar (ou ultrapassar) a inteligência humana. Merece ser trazida a definição de Lazar *et al*²⁵ sobre o movimento dos objetos *smart*: os dispositivos inteligentes surgiram na promessa de oferecer aos utilizadores uma “riqueza de informações que lhes permitirá se tornarem as melhores versões de si mesmos”²⁶.

A mesma obra mostra o quão descartáveis estes aparelhos podem se tornar ao coletarem demasiada informação e não atingirem o objetivo que os torna *smart*. Para citar como exemplo, muitos rastreadores “inteligentes” de atividades físicas fornecem aos utilizadores o número de medidas tomadas naquele dia, na pretensão de que o conhecimento dessas informações causará mudanças no estilo de vida, tais como o aumento da atividade física. Se a coleta de informações de cada passo

²⁴ Segundo BODEN, criatividade é definível e há uma possibilidade da IA atingir os três pilares que a sustentam. “*Creativity isn’t magical. It’s an aspect of normal human intelligence, not a special faculty granted to a tiny elite. There are three forms: combinational, exploratory, and transformational. All three can be modeled by AI*”. Cf. BODEN, Margaret A. “Computer models of creativity”. *AI Magazine*, v. 30, n. 3, 2009, p. 23.

²⁵ LAZAR, Amanda et al. “Why we use and abandon smart devices”. *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*, 2015.

²⁶ LAZAR, Amanda et al. *Ibid.* p. 635. (Tradução livre)

dados, sob a permissão de localização geográfica, não impactar e gerar mudança efetiva na vida do utilizador, o aparelho vai representar uma forma de coletar dados em excesso, por opção do próprio titular dos dados²⁷.

O movimento em voga de adicionar o selo *smart* em televisão, frigorífico ou em sensores diversos, sem que o utilizador explore em plenitude aquela “inteligência” oferecida, representa um uso que, além de não ser inteligente sob o ponto de vista da teoria de Gartner²⁸, o aparelho apenas vai ser resumido em coleta excessiva de dados pessoais, sem uma contraprestação a longo prazo ao seu utilizador. Se um objeto que anteriormente não acedia à *internet*, passa a aceder, não deveria, necessariamente, ser considerado inteligência artificial, sequer ganhar o selo de aparelho inteligente²⁹.

Ante o exposto, conforme discutido acima sobre os conceitos de inteligência, é aconselhável não perceber o fenómeno dos *smart devices* de forma literal, porque muitos não oferecem sequer recursos de Inteligência Artificial de forma predominante³⁰. Via de regra, tais aparelhos apenas coletam dados para retribuir em novas funções de interconexão a outros aparelhos. O selo *smart* é muito mais um propósito comercial desvinculado a qualquer vontade de conferir inteligência ao aparelho, mas sim agregar valor a produtos comercializáveis.

E, portanto, o conceito apresentado de IA acaba por representar um grande escopo. O âmbito, para fins de investigação, segue o proposto por Fetzer que limitou o conceito de IA a apenas sistemas dotados da capacidade de aprender ou entender a partir da experiência³¹. E dentro deste âmbito, no qual a aprendizagem é requisito, a IA pode ser

²⁷ Ver mais em: LAZAR, Amanda, et al. *Ibid.*

²⁸ GARDNER, Howard. *Ibid.*

²⁹ A diferença técnica entre *smart devices* e algoritmos de IA vem definida em SCHALKOFF, Robert J. *Artificial intelligence: an engineering approach*. McGraw-Hill, 1990.

³⁰ A ausência de inteligência propriamente dita em aparelhos *smart* é bem observada em LEESA-NGUANSUK, Suchit. *Smart device market not so smart*. Disponível em: <<https://www.bangkokpost.com/business/1075440/smart-device-market-not-so-smart>> Acedido a: 02.12.2020

³¹ FETZER, James H. *Artificial intelligence: Its scope and limits*. Springer Science & Business Media, 2012.

subdividida em áreas como planejamento automatizado, aprendizagem de máquina, processamento de linguagem natural, robótica, dentre outras denominações³².

Ao contrário desta concepção, a IA pode ser erroneamente associada a utilizações que não correspondem tecnicamente a uma inteligência propriamente dita pela ausência da mencionada capacidade de aprender. Mas esta associação pode ser proveniente da dificuldade de enquadramento do âmbito da IA pela complexidade dos conceitos abstratos trazidos pela própria expressão. O conceito de inteligência humana já oferece um significativo teor de subjetividade e de complexa limitação objetiva, por isso a inteligência da máquina, computador ou qualquer outra entidade que esteja abrangida pela ideia de “artificial” vai passar pela mesma dificuldade³³.

A dificuldade conceitual não pode ser utilizada como um artifício na perigosa generalização do conceito de IA, muitas vezes ligado a usos que não apresentam aprendizagem algorítmica, ou qualquer outra forma de criatividade e inteligência por parte da entidade computacional. Dito isto, parte-se a uma problemática que emerge do processo específico a respeito da aprendizagem de máquina³⁴, cujos resultados apresentam interpretações estatísticas que podem ser discriminatórias aos indivíduos a quem a decisão automatizada foi destinada.

3. A origem da discriminação

Se o presente objetivo é entender como a IA pode oferecer resultados discriminatórios, e uma vez introduzidos os conceitos de inteligência e de inteligência artificial, precisa-se apresentar previamente o conceito da discriminação. Esta pode ser percebida a partir de uma conduta

³² Neste sentido, ver em RUSSELL, Stuart; NORVIG, Peter. *Artificial intelligence: a modern approach*. GEN LTC, 2002.

³³ Na dicotomia entre inteligência humana e inteligência artificial, cf. KURZWEIL, Ray. *Ibid.*

³⁴ Tradução de *machine learning* cf. ONGSULEE, Pariwat. “Artificial intelligence, machine learning and deep learning”. In: *2017 15th International Conference on ICT and Knowledge Engineering (ICT&KE)*. IEEE, 2017. p. 1-6.

de transgredir direitos humanos³⁵, mas na essência, a discriminação parte de outro princípio fundamental que é a igualdade entre os seres humanos.

A construção da discriminação deve ser à luz da prévia compreensão acerca do paradigma mundial que define a igualdade entre os povos³⁶. Esta noção de igualdade entre povos, etnias e indivíduos de maneira isolada, foi adotada internacionalmente aquando da criação da Declaração Universal dos Direitos Humanos³⁷, que estabeleceu o princípio ético universal da igualdade.

A igualdade é um princípio que não funciona sob a ótica de igualdade absoluta entre indivíduos. A igualdade foi construída para ser interpretada à luz da isonomia material, uma vez que o mandamento constitucional não estabelece o tratamento de todos os indivíduos da mesma forma. A igualdade material orienta para um tratamento desigual na medida em que haja desigualdade. Em outras palavras, segundo o conceito aristotélico: “Entre semelhantes, a honestidade e a justiça consistem em que cada um tenha a sua vez. Apenas isto conserva a igualdade. A desigualdade entre iguais e as distinções entre semelhantes são contra a natureza e, por conseguinte, contra a honestidade”³⁸.

Portanto, desde Aristóteles, o tratamento diferenciado para que seja diferente não representa discriminação de conotação negativa pela noção de isonomia material. Conforme defende Sen³⁹, o objetivo da igualdade é preservar a individualidade, sem ignorar as diferenças culturais e materiais de cada indivíduo, a fim de uma busca pela equiparação de oportunidades. E é nesta análise de oportunidade que deve pender a balança imaginária

³⁵ Neste sentido, veja-se VIERDAG, Egbert Willem. *The concept of discrimination in international law: with special reference to human rights*. Springer Science & Business Media, 2012.

³⁶ LINDOSO, Maria Cristine Branco. *Discriminação de gênero em processos decisórios automatizados*. Repositório Universidade de Brasília, 2019.

³⁷ Organização das Nações Unidas. *Declaração Universal dos Direitos Humanos*, 10 de dezembro de 1948. Disponível em: <<http://www.un.org/en/universal-declaration-human-rights/>> Acedido a: 29.11.20.

³⁸ Conforme Aristóteles. *Política*. Coleção a obra prima de cada autor. Le livros. Disponível em <<file:///C:/Users/ADM/Downloads/Livro%20A%20Politica%20-%20Aristoteles.pdf>> Acedido a: 02.02.2020; p. 40.

³⁹ SEN, Amartia. *The idea of justice*. Belknap Press, 2011, p. 293-295.

que avalia a conduta como um tratamento diferenciado isonômico para um tratamento discriminatório.

Apesar de ter sido adotada na seara mais específica da discriminação de gênero, a definição adotada na Convenção Internacional sobre a Eliminação de Todas as Formas de Discriminação Contra a Mulher, conforme explicado por Rios *et al*⁴⁰, ajuda na conceção da discriminação como qualquer distinção, exclusão, restrição ou preferência que tenha o propósito ou o feitiço de anular ou prejudicar o reconhecimento, gozo ou exercício em pé de igualdade de direitos humanos e liberdades fundamentais nos campos econômico, social, cultural ou em qualquer campo da vida pública⁴¹.

É imperioso destacar que esta definição de discriminação é abrangente do ponto de vista de atitudes (ao utilizar o termo “qualquer” antes de várias condutas), mas faz uma restrição relevante no que concerne aos efeitos destas condutas. Deve existir – intrinsecamente ao conceito discriminatório – uma violação jurídica de caráter de direitos humanos e liberdades fundamentais. Esta jusfundamentalidade é o parâmetro para separar um tratamento diferenciado isonômico⁴² do tratamento discriminatório.

Por isso é que, quando a igualdade dos povos passou a ser um direito humano reconhecido internacionalmente, ao Estado passou a ser exigido a adoção de uma verdadeira atuação positiva para a equiparação de oportunidades na promoção do tratamento diferenciado isonômico, em nome da valorização de todas as vidas humanas, conforme foi desenvolvido por Dworkin⁴³. E esta mencionada atuação positiva do Estado

⁴⁰ Veja-se em RIOS, Roger Raupp; SILVA, Rodrigo da. “Democracia e Direito da Antidiscriminação: interseccionalidade e discriminação múltipla no direito brasileiro”. *Revista de Ciência e Cultura*, São Paulo, v. 69, n. 1, 2017, p. 44-46.

⁴¹ Conceito adaptado da previsão do artigo 1º da Convenção sobre a Eliminação de Todas as Formas de Discriminação Contra as Mulheres Aprovada pela Assembleia Geral das Nações Unidas em 18 de dezembro de 1979. Disponível em <<https://www.unicef.org/brazil/convencao-sobre-eliminacao-de-todas-formas-de-discriminacao-contra-mulheres>> Acedido a 02.02.2020.

⁴² Justificado pela inclusão e prática da isonomia material como desígnio da justiça.

⁴³ No conceito prático de “equality” a igualdade de oportunidades foi levada em consideração. Veja-se em DWORKIN, Ronald. *Sovereign Virtue. The Theory and Practice of Equality*. London, England: Harvard University Press, 2002, p. 390.

não diz respeito apenas ao sentido de atuação jurisdicional mediante provocação do Judiciário por parte daqueles que sofreram alguma discriminação.

A atuação jurisdicional obedece ao princípio da inércia do Judiciário, que precisa de tal provocação do interessado para uma possível composição da lide⁴⁴. A não discriminação passa a ser um direito com reflexos constitucionais reconhecidos que demanda a atuação positiva estatal independente de provocação, como por exemplo, iniciativa do Legislativo para combater a discriminação, ou programas do Executivo para coibir práticas discriminatórias, sendo a necessidade de combater a discriminação um ponto de convergência entre os três Poderes estatais, por sua elevação a direito humano reconhecido.

Portanto, a discriminação deve ser interpretada como uma violação de direitos humanos, com devidos efeitos jurídicos⁴⁵. A juridicidade como requisito para a discriminação passa a funcionar como um elemento que confere segurança jurídica à própria proteção contra a discriminação, uma vez que evita a banalização do termo⁴⁶. Não se deve generalizar a interpretação para abranger qualquer tratamento diferenciado sendo discriminatório, porque diversas vezes a diferença de tratamento vai ser em decorrência da isonomia material e no combate a uma desigualdade existente.

É sob este prisma que a discriminação praticada por sistemas de inteligência artificial vai ser analisada. Não se busca analisar cada efeito que uma decisão automatizada pode gerar, mas sim os casos em que o resultado de um sistema de IA pode gerar efeitos jurídicos indevidos na vida de algum indivíduo.

⁴⁴ Para aprofundar sobre a inércia e outros princípios da teoria geral do processo civil, veja-se DIDIER JR., Fredie. *Sobre a Teoria Geral do Processo*. v. II. Salvador: Editora Jus Podivm, 2013.

⁴⁵ Nos termos do Artigo 7 da Declaração Universal dos Direitos Humanos. Disponível em <<https://www.unicef.org/brazil/declaracao-universal-dos-direitos-humanos>> Acedido a 19.04.21

⁴⁶ Sobre os riscos da banalização no contexto da diversidade étnico-cultural, veja-se ROSHWALD, Aviel. “Between Balkanization and Banalization: Dilemmas of Ethno-cultural Diversity”. *Ethnopolitics*, v. 6, n. 3, 2007, p. 365-378.

4. A discriminação científica precedente à discriminação algorítmica

Antes de debater os algoritmos e seus efeitos, precisa-se trazer um problema que antecede a própria criação das mais diversas tecnologias. Apesar da tecnologia, por si só, ser um ponto final neutro, precede-lhe uma jornada feita por indivíduos, técnicos de alguma área do conhecimento e cientistas, que movidos pelas mais diversas motivações, buscam novos resultados. Ocorre que, segundo Cowan⁴⁷, desde os primórdios do desenvolvimento tecnológico não existe igualdade de participação na construção do conhecimento. A autora defende que um dos primeiros ambientes em que a tecnologia da informação revolucionou a vida cível foi o ambiente doméstico; mas ocorreu de forma incongruente.

Para ilustrar a origem da desigualdade tecnológica, Cowan cita os primeiros eletrodomésticos no contexto da sociedade patriarcal: os aparelhos iriam ser utilizados pelas mulheres, excluídas do mercado de trabalho e presas no âmbito doméstico, mas inexistiam mulheres participando da produção destes aparelhos. Havia uma inadequação do uso da tecnologia em relação à realidade das mulheres em decorrência da falta de participação feminina na vida científica⁴⁸.

Não é distante a correlação entre sub-representação de mulheres (ou qualquer outro grupo social) na produção do conhecimento e um desenvolvimento científico baseado na desigualdade. A pluralidade social na produção científica auxilia na consecução de usos tecnológicos positivos na esfera individual e coletiva, uma vez que os produtos da ciência são institutos que modificam a sociedade e atingem a esfera jurídica humana. Por isso é que, a fim de reduzir futuros efeitos discriminatórios, segundo a obra de Bray⁴⁹, há de se encontrar soluções para que os efeitos da desigualdade possam ser minimizados no (e através do) campo da ciência e da tecnologia.

Conforme é visto a seguir, a discriminação algorítmica pode atingir diversas minorias. Os grupos de notáveis cientistas da computação

⁴⁷ COWAN, Ruth. *More Work for Mother: The Ironies of Household Technology from the Open Hearth of the Microwave*, Basic Books, New York, 1983.

⁴⁸ COWAN, Ruth. *Ibid.*

⁴⁹ BRAY, Francesca. "Gender and Technology". *The Annual Review of Anthropology*, 2007.

que desenvolvem os algoritmos deveriam precaver-se do tratamento discriminatório que resulta em danos jurídicos a serem reparados. Uma das formas de evitar danos discriminatórios seria antecipadamente promover, ainda na etapa de criação do sistema inteligente, a formação de grupos de desenvolvimento mais plurais. Há pesquisas indicando que mais representatividade pode liderar a um maior cuidado e atenção, não apenas no desenvolvimento, mas também na auditoria contínua do algoritmo⁵⁰.

A citar como exemplo, se algum sistema de IA for desenvolvido por um grupo de desenvolvedores que contenha mulheres, a representatividade feminina desde a concepção ajuda numa maior diligência prévia em relação ao algoritmo. Não significa que um algoritmo desenvolvido também por mulheres não possa adotar uma decisão automatizada discriminatória, mas a pretensão é que uma supervisão feminina em relação às bases de dados usadas nos treinamentos e nos resultados do algoritmo seja mais equilibrada.

Por isso, a pluralidade científica deve ser tomada como prioridade no momento antecedente ao desenvolvimento de alguma tecnologia, porque a diversidade vai auxiliar em precaução acerca dos sistemas discriminatórios, ou se mesmo assim o uso específico da tecnologia gere efeitos discriminatórios relacionados a alguma minoria, este grupo vai estar representado dentre os especialistas que podem reparar tal dano.

Sob esta perspectiva, a IA como conceito tecnológico amplo é neutra, mas o seu uso em específico pode gerar efeitos jurídicos discriminatórios. Conforme foi mencionado, a prevenção dos efeitos discriminatórios não se resume a uma fase do ciclo de vida do algoritmo, uma vez que a prevenção deve ocorrer desde o momento da concepção até ao acompanhamento constante no momento de utilização do algoritmo. E sobre os efeitos dessa possível discriminação passa-se a discorrer.

⁵⁰ Noção desenvolvida por no estudo da PRETALAB acerca do urgente debate sobre representatividade no universo da inovação. Veja-se em PRETALAB. Um levantamento sobre a necessidade e a pertinência de incluir mais mulheres negras na inovação e na tecnologia. 2018. Disponível em: <<https://www.pretalab.com/>>. Acedido a: 02.02.2020.

5. A Inteligência Artificial e a possível Discriminação Algorítmica

A participação e a representatividade na ciência, ao longo do tempo, tem sido uma das causas para a formação de novos cientistas e especialistas mais plurais⁵¹. E na seara da IA, a necessidade parte do mesmo pressuposto de pluralidade, sobretudo para evitar que uma decisão seja ao mesmo tempo injusta e automatizada.

Segundo Barocas et al⁵², qualquer sistema de tomada de decisão é capaz de errar. Sejam decisões baseadas em métodos convencionais com envolvimento humano significativo ou completo, mas também os mais sofisticados algoritmos de aprendizagem que podem assumir as diferentes formas de tomada de decisão. Sobre estas várias possibilidades, a discriminação algorítmica tem definição dificultada por não ser um termo de arte, mas um termo geral que pode se referir a um viés específico ou uma coleção de vários deles⁵³.

Uma definição sobre discriminação algorítmica que leva esta complexidade em consideração é a que se refere tecnicamente a entradas ou saídas de um sistema de IA, onde tais resultados apresentam um resultado errôneo ou injustificado sobre dois grupos. Deve haver, neste sentido, tratamento diferenciado ilícito proveniente de erros ou problemas no sistema de IA, ou pelos dados que são utilizados, ou pelos efeitos por si, ou por todas as opções anteriores⁵⁴.

Em outras palavras, um viés algorítmico pode existir e levar a um tratamento diferenciado, mas isto demanda maior detalhamento sob o ponto de vista da já explicada isonomia material para fins de definir o grau tendencioso do referido algoritmo. Não obstante, deve-se analisar no caso concreto se tal decisão algorítmica representa uma discriminação ilegal, onde o resultado além de não possuir uma justificação legal, passa

⁵¹ Neste sentido REIS, Claudio Ricardo Martins dos. *Ciência e Valores: em defesa de um pluralismo sensível ao contexto*. Lumes, 2019.

⁵² BAROCAS, Solon; SELBST, Andrew D, “Big Data’s Disparate Impact”, v. 104, *California Law Review*, 2016, p. 671.

⁵³ LEE, Nicol Turner; RESNICK, Paul; BARTON, Genie. *Algorithmic Bias Detection and Mitigation: Best Practices and Policies to Reduce Consumer Harms*. Brookings, 2019.

⁵⁴ Neste sentido, veja-se em: LATTIMORE, Finn. O’CALLAGHAN, Simon; PALEOLOGOS, Zoe, et al. *Using artificial intelligence to make decisions: Addressing the problem of algorithmic bias*. Australian Human Rights Commission, Technical Paper, 2020.

a infringir direitos humanos e garantias fundamentais. Por isso, a atenção ao problema da discriminação algorítmica desde a concepção de qualquer produto que utilize IA vai ser importante para mitigar os riscos de uma violação de direito.

Para mitigar riscos de discriminação é relevante considerar que já existe uma desigualdade natural nas relações sociais⁵⁵, naturalmente extrínseca aos modelos de IA. Acima desta desigualdade existe a discriminação, fruto de decisão automatizada que pode ser apenas um reflexo da desigualdade social existente que gera dados enviesados ou resulta do próprio sistema algorítmico que foi mal desenvolvido, sob a perspectiva técnica.

Todavia, nem toda a discriminação algorítmica deve ser resultado de um erro técnico, mas pode também ser fruto de um algoritmo que funciona da forma exata para a qual foi desenvolvido, mas que foi treinado com dados enviesados. Os resultados vão depender dos dados através dos quais o algoritmo foi treinado⁵⁶. Seja problema intrínseco ao sistema, seja uma polarização em resposta a dados enviesados ou treinamento tendencioso da máquina, para fins de relevância jurídica, apenas os resultados que infringem direitos devem ser levados em consideração, de modo a limitar o presente estudo e qualquer medida de mitigação a ser adotada.

Uma das vertentes trazidas é a discriminação algorítmica proveniente dos dados⁵⁷. Os sistemas de IA são treinados a partir de significativas bases de dados, contudo a diferença é que algumas irão conter dados pessoais – sobre pessoas naturais identificadas, outras sobre pessoas naturais identificáveis⁵⁸. Outras bases de dados promovem a atenção ao

⁵⁵ Conforme JOHANSEN: A desigualdade nas relações sociais advém de uma natural distribuição de grupos populacionais com perfis socioeconômicos distintos. Ver mais em JOHANSEN, Igor Cavallini; CARMO, Roberto Luiz do; ALVES, Luciana Correia. “Desigualdade social intraurbana: implicações sobre a epidemia de dengue em Campinas, SP, em 2014”. *Cadernos Metrópole*, v. XVIII, n. 36, 2016, p. 421-440.

⁵⁶ Neste sentido, COWGILL, Bo; TUCKER, Catherine. *Algorithmic bias: A counterfactual perspective*. NSF Trustworthy Algorithms, 2017.

⁵⁷ Conforme identificado através do *data mining* e ciência dos dados, cf. HAJIAN, Sara; BONCHI, Francesco; CASTILLO, Carlos. “Algorithmic bias: From discrimination discovery to fairness-aware data mining”. *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 2016. p. 2125-2126.

⁵⁸ O que para fins do RGPD tem o mesmo efeito de proteção, cf. União Europeia. Regulamento (UE) 2016/679 do Parlamento Europeu e do Conselho de 27 de abril de 2016

princípio da proteção de dados e para fins de treinamento de algoritmos utilizam dados sintéticos provenientes de uma simulação⁵⁹. Com dados sintéticos, os desenvolvedores de algoritmos podem treinar os sistemas de decisão com dados mais plurais⁶⁰.

Além de respeitar a privacidade e evitar um tratamento excessivo de dados pessoais, os algoritmos de IA treinados com base em dados sintéticos são mais ajustáveis do que os treinados com dados reais⁶¹. Alterar a realidade representa demasiado esforço, quando a realidade reflete formas de discriminação, é por isso que criar uma realidade artificial, com diversidade sociocultural, pode ser uma forma mais eficaz de combater discriminação algorítmica. Em outros termos, a capacidade de controlar um enviesamento dos dados, é possível com um simulador de dados; e tal simulação deve considerar desde a conceção as possíveis formas de discriminação para evitá-las⁶².

Portanto, o uso de dados sintéticos vem sendo aventado como uma possível solução que ao mesmo tempo confere atenção a três importantes institutos: (i) a proteção de dados pessoais e a possível coleta excessiva para treinamento de algoritmos de IA; (ii) criação de uma realidade artificial que se sobrepõe e não reflete a acima mencionada desigualdade endêmica nas relações sociais, naturalmente extrínseca aos modelos de IA; (iii) capacidade de controlar um enviesamento dos dados que possivelmente vai gerar uma discriminação com efeitos jurídicos sobre os indivíduos.

relativo à proteção das pessoas singulares no que diz respeito ao tratamento de dados pessoais e à livre circulação desses dados.

⁵⁹ Dados sintéticos são dados para teste que oferecem maior possibilidade de uma grande formação rotulada, pelo facto de serem artificialmente gerados. Conforme DAHMEN, dados sintéticos oferecem menos limitações, porque “os conjuntos de dados do mundo real (...) são limitados em termos de disponibilidade e variedade”. Veja-se em DAHMEN, Jessamyn; COOK, Diane. *SynSys: A synthetic data generation system for healthcare applications*. *Sensors*, v. XIX, n. 5, 2019, p. 1181. Tradução livre.

⁶⁰ Sem precisar filtrar no mundo analógico situações que não tragam riscos de discriminação.

⁶¹ O maior grau de ajustamento é bem explicado em HE, Haibo et al. “ADASYN: Adaptive synthetic sampling approach for imbalanced learning”. *2008 IEEE international joint conference on neural networks*, IEEE, 2008, p. 1322-1328.

⁶² Neste sentido, veja-se em: LATTIMORE, Finn; O'CALLAGHAN, Simon; PALEOLOGOS, Zoe, et al. *Ibid.*

6. Casos de estudo sobre discriminação algorítmica

Conforme foi visto, o âmbito da IA é amplo, e é por isso que os casos concretos a serem utilizados neste estudo devem atingir os requisitos teóricos do conceito. Esta definição, recorda-se, limita a IA aos sistemas dotados da capacidade de aprender ou entender a partir da experiência⁶³. Em conformidade com este processo de aprendizagem de máquina, parte-se para analisar casos em que o sistema aprende e decide, dentre algumas das áreas que compreendem a IA.

Uma das searas que mais levanta discussões sobre os impactos do tratamento de dados pessoais sensíveis para fins de treinamento de sistemas de IA com elevado poder de interferência social é a do reconhecimento facial. Em estudo recente, Joy Buolamwini do Massachusetts Institute of Technology (MIT)⁶⁴ descobriu que três modelos de reconhecimento facial sobre o gênero, nomeadamente o sistema das empresas IBM, Microsoft e da empresa chinesa Megvii, podiam identificar corretamente o gênero de uma pessoa a partir de uma fotografia com taxa de 99% (noventa e nove por cento) de acerto. Contudo, a taxa de acerto foi demonstrada ser apenas para o reconhecimento de homens de pele mais clara. O reconhecimento de mulheres com cor de pele mais escura caía para 35% (trinta e cinco por cento) de acerto⁶⁵.

Uma queda significativa na taxa de acerto por um algoritmo, no que diz respeito a cor da pele, não quer dizer que o sistema de IA opta por alcançar estes resultados, inexistindo intenção algorítmica em ser racista. O algoritmo é apenas uma tecnologia e pode ser treinado e aprender a tratar as pessoas de diferentes cores de pele de forma isonômica. O caminho para este tratamento não discriminatório está no treinamento que gera o aprendizado de máquina. Neste interim, devem as empresas desenvolvedoras de algoritmos de IA utilizar uma quantidade igual de

⁶³ FETZER, James H. *Ibid.*

⁶⁴ Sobre este estudo, veja-se BUOLAMWINI, Joy. *Gender Shades: Intersectional Phenotypic and Demographic Evaluation of Face Datasets and Gender Classifiers*. Master of Science, Massachusetts Institute of Technology, Cambridge, MA, 2017.

⁶⁵ REVELL, Timothy. *Face-recognition software is perfect – if you're a white man*. Disponível em: <<https://www.newscientist.com/article/2161028-face-recognition-software-is-perfect-if-youre-a-white-man/>> Acedido a: 02.12.2020

imagens de mulheres com a cor da pele escura e de homens com a cor de pele clara, por exemplo. Fazendo isto, o resultado do sistema de IA provavelmente não indicará tamanha discrepância nos resultados⁶⁶.

Não apenas para possíveis erros de caracterização étnica, mas os resultados deste algoritmo de reconhecimento facial lecionam que é imperioso o treinamento equilibrado com uso de dados não enviesados. Seguindo a pesquisa do MIT, a IBM anunciou de imediato que havia retreinado o seu sistema num novo conjunto de dados⁶⁷. Este posicionamento público da empresa deixa evidente a mensagem que os resultados discriminatórios, por parte do seu sistema de reconhecimento, não se deram por erros técnicos, mas pelo conjunto de dados utilizados no treinamento do sistema. *A mea culpa* da IBM evidencia que o sistema funciona da maneira esperada, sem falhas técnicas evidentes, mas que precisa ser retreinado, porque os dados de treinamento, provavelmente, possuíam muito mais casos de indivíduos do sexo masculino com cor de pele clara, em detrimento de indivíduos do sexo feminino com cor de pele escura.

Ainda sobre a pesquisa do MIT, Ruchir Puri, cientista chefe da IBM Research, acabou por aproximar-se da pesquisadora Buolamwini após o estudo que revelou uma discriminação algorítmica de cor de pele. Em resposta ao resultado discriminatório e a outros incidentes, a IBM revelou que passou a utilizar um novo banco de dados com 1 milhão de imagens para melhor analisar a diversidade dos rostos humanos. Os sistemas anteriores têm sido excessivamente dependentes do que, segundo Buolamwini, são repositórios de imagens “masculinas pálidas”⁶⁸.

A citada falha do algoritmo de reconhecimento facial, por si só, já constitui uma violação de direitos humanos por praticar discriminação sobre cor e gênero, violando o direito fundamental da não discriminação, conforme Artigo 21 da Carta Dos Direitos Fundamentais da União

⁶⁶ Conforme defendido por PENA, Alejandro et al. “Bias in multimodal AI: testbed for fair automatic recruitment.” In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. 2020.

⁶⁷ REVELL, Timothy. *Ibid*.

⁶⁸ O'BRIEN, Matt. *MIT Researcher Exposing Bias in Facial Recognition Tech Triggers Amazon's Wrath*. Disponível em: <https://www.insurancejournal.com/news/national/2019/04/08/523153.htm> > Acedido a: 02.12.2020 (Tradução livre)

Europeia (CDFUE)⁶⁹; e o princípio ético universal da igualdade, segundo a Declaração Universal dos Direitos Humanos. Entretanto, não é apenas a igualdade que pode ser afetada por uma decisão automatizada tomada com base em dados enviesados ou incompletos. O direito fundamental à própria vida, garantido pelo Artigo 2 CDFUE e pela ordem constitucional, também pode ser afetado pela discriminação algorítmica.

Diante deste cenário, é interessante mencionar um estudo que foi divulgado sobre o descrito tipo de discriminação, nomeadamente a respeito dos sistemas de IA que guiam carros autônomos. Limitando a apenas os casos em que não há qualquer envolvimento humano, existe um perigo inerente quando o sistema de IA com poderes de autonomamente conduzir o veículo, incorre em má classificação de um indivíduo.

Em um estudo da Computer Vision Systems⁷⁰ foi concluído que os carros autônomos testados ofereciam uma maior dificuldade de identificar pessoas com tom de pele mais escura. Se existir este grau de dificuldade, a utilização deste sistema pode ser uma ameaça aos direitos fundamentais envolvidos nesta situação. A inexistência de qualquer tratamento diferenciado proveniente da cor da pele deve ser um requisito obrigatório para o uso de mercado deste conjunto de algoritmos de IA para fins de conduzir carros autônomos.

Ou seja, uma falha tecnológica da gravidade do algoritmo de reconhecimento não identificar pessoas de tom de pele mais escura seria um problema que ultrapassa o âmbito tecnológico. A discriminação inerente ao sistema iria ser discriminatória e por em risco a vida e a liberdade de um grupo de indivíduos, violando assim Direitos Fundamentais da CDFUE e possíveis legislações dos Estados-Membros.

Em contraponto, a especialista em IA, Kate Crawford, em reportagem do jornal *Independent*, aponta que os resultados da referida pesquisa da Computer Vision Systems são contestáveis. Ela explica com o argumento que o estudo não usou as mesmas bases de dados utilizadas pelos

⁶⁹ União Europeia. Carta Dos Direitos Fundamentais Da União Europeia (CDFUE), 2016/C, 202/02 que retoma a Carta proclamada em 7 de dezembro de 2000 e substitui-a a partir da data de entrada em vigor do Tratado de Lisboa.

⁷⁰ Associated Press. *AI researchers slam Amazon for selling 'biased' facial recognition tech to cops*. Disponível em: <<https://nypost.com/2019/04/04/ai-researchers-slam-amazon-for-selling-biased-facial-recognition-tech-to-cops/>> Acedido a 02.12.2020

desenvolvedores de carros autônomos. E isto pode não refletir a real exatidão dos sistemas do “mundo real, já que em um mundo ideal, os acadêmicos estariam testando os atuais modelos e bases de treinamento usadas pelos desenvolvedores de carros autônomos”⁷¹.

Ao ver sob esta perspectiva prática, destaca-se o facto de os primeiros acidentes envolvendo carros autônomos não terem envolvido pessoas com a cor da pele negra. Pode-se listar os primeiros casos como: (i) o chinês Gao Yuning faleceu devido o *autopilot* do seu Tesla ter chocado com um veículo, em 2016; (ii) o americano caucasiano Joshua Brown que faleceu devido ao mesmo sistema *autopilot* da Tesla ter colidido com um trator, em 2016; e (iii), a morte da também caucasiana Elaine Herzberg que foi atingida por um veículo autônomo da Uber em 2018, após ter surgido de uma sombra para a estrada⁷².

De toda a forma, por mais que nos três casos citados a cor da pele não tenha sido um ponto essencial ao acidente, a discriminação algorítmica por cor, nos moldes do estudo da Computer Vision Systems, merece atenção no desenvolvimento e posterior utilização de carros autônomos (ou qualquer outra decisão automatizada). Dito isto, se de um lado, no caso dos carros autônomos, a discriminação pode gerar um risco iminente à vida dos indivíduos, por outro lado, existem casos que a vida de terceiros não está em risco, mas que representam também discriminação algorítmica.

Traz-se como exemplo o resultado do estudo da Universidade de Washington, que delineou uma espécie de discriminação de gênero. Na pesquisa de imagens sobre o termo “CEO”, o estudo obteve como resultado que o Google apenas mostrava 11% (onze por cento) das pessoas sendo mulheres, embora um levantamento tenha mostrado que 27% (vinte e sete por cento) dos chefes-executivos no mesmo âmbito territorial do estudo seria do sexo feminino⁷³.

⁷¹ Tradução livre de “In an ideal world, academics would be testing the actual models and training sets used by autonomous car manufacturers”. Ver mais em CUTHBERTSON, Anthony. *Self-driving cars more likely to drive into black people, study claims*. Disponível em: <<https://www.independent.co.uk/life-style/gadgets-and-tech/news/self-driving-car-crash-racial-bias-black-people-study-a8810031.html>> Acedido a 17.04.2021

⁷² GRAY, Stuart. List of driverless vehicle accidents. Disponível em: <<https://www.itgs-news.com/list-of-driverless-vehicle-accidents/>> Acedido a 17.04.2021

⁷³ Ver mais em PANCH, Trishan; MATTIE, Heather; ATUN, Rifat. “Artificial intelligence and algorithmic bias: implications for health systems”. *Journal of global health*, v. 9, n. 2, 2019.

Estudar os resultados dos motores de busca acaba por ser um exercício para averiguar possíveis discriminações. Resultados enviesados de uma busca podem não apenas causar discriminação, mas também manipular grupos de pessoas. Um dos que defendem esta premissa é o autor e ativista Epstein⁷⁴. Os motores de busca possuem o poder de influenciar opiniões e até eleições, através de um fenómeno denominado “SEME” (efeito manipulador do motor de busca)⁷⁵.

Contudo, a discriminação não se limita apenas aos efeitos da IA utilizada pelos motores de busca. Conforme debatido ao longo deste texto, a IA apresenta um amplo âmbito e diversas aplicações. Sendo o motor de busca enviesado apenas exemplo de um tipo de IA que pode oferecer resultados discriminatórios, há outras áreas que podem oferecer o mesmo efeito jurídico de discriminação.

Partindo desta premissa, exemplifica-se que, ainda na esfera da discriminação de gênero, os algoritmos de IA têm sido capazes de definir o futuro profissional das mulheres. Estudos conduzidos pela Carnegie Mellon University em Pittsburgh, EUA, concluíram que o sistema de publicidade online do Google mostrava empregos de alta renda para os homens com muito mais frequência do que para as mulheres⁷⁶. Não obstante os efeitos discriminatórios inerentes, esta discriminação algorítmica pode levar a repercussões contrárias à luta por um mercado de trabalho igualitário, devido ao poder e alcance do referido motor de busca.

Tal estudo da Carnegie Mellon University traz em seu título – *A tale of opacity, choice, and discrimination*⁷⁷ – três institutos basilares na percepção da discriminação algorítmica: opacidade, escolha e discriminação. O encarregado de auditar um algoritmo de IA busca romper a barreira da opacidade, a fim de perceber como ocorreu o processo de escolha,

⁷⁴ EPSTEIN, Robert. *Manipulating minds: The power of search engines to influence votes and opinions*. in MOORE et al, *Digital dominance: The power of Google, Amazon, Facebook, and Apple*. Oxford University Press.

⁷⁵ Significado da sigla: “SEME – the Search Engine Manipulation Effect”. Ver mais em EPSTEIN, R.; ROBERTSON, R. E. *The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections*. Proceedings of the National Academy of Sciences USA, 2015.

⁷⁶ DATTA, Amit; TSCHANTZ, Carl Michael; DATTA, Anupam. “Automated experiments on ad privacy settings: A tale of opacity, choice, and discrimination”. *Proceedings on privacy enhancing technologies*, n. 1, 2015, p. 92-112.

⁷⁷ Ibid.

a tomada de decisão automatizada. Em seguimento à análise, um dos resultados vai ser justamente a resposta se houve ou não discriminação. Conforme esta metodologia, o referido estudo analisou um conjunto de algoritmos publicitários do Google.

Pode-se dizer que os pesquisadores encontraram evidências de que os anunciantes conseguem direcionar seus respectivos anúncios com base no gênero, e ainda que o algoritmo de IA poderia tomar decisões automatizadas alheias às definições de gênero definidas pelos anunciantes. Datta *et al* apontaram que o Google poderia ter determinado que os homens são mais adequados para cargos executivos por conta própria⁷⁸. Facto é que tal decisão, por conta própria, configura uma tomada de decisão automatizada com efeitos jurídicos relevantes, uma vez que ao mostrar anúncios de empregos bem remunerados apenas a homens, haverá discriminação algorítmica de gênero com efeitos possíveis de enfraquecer a participação das mulheres no mercado de trabalho.

Apesar desses resultados apontarem para uma decisão automatizada obrigatória, os próprios pesquisadores expressaram que não obtiveram todas as informações, porque as mesmas foram ocultadas pela falta de transparência do algoritmo⁷⁹. Isso mostra que a análise de discriminação pode ser dificultada pela opacidade algorítmica, quando ela apresenta uma “*black box*” intransponível a ser considerada⁸⁰.

E esta opacidade acaba por ser um obstáculo para que a ciência obtenha uma análise das aplicações práticas da IA. Pelo acima exposto, a transparência algorítmica, aliada à clareza dos respectivos dados de treinamento e do processo de tomada da decisão automatizada, são institutos basilares no combate ao viés discriminatório que a IA pode adotar.

Conclusão

Ante o exposto, uma vez percebido que independentemente dos limites da máquina, o atual estado da arte é que existem diversos tipos de

⁷⁸ DATTA, Amit; TSCHANTZ, Carl Michael; DATTA, Anupam. *Ibid.*

⁷⁹ DATTA, Amit; TSCHANTZ, Carl Michael; DATTA, Anupam. *Ibid.*

⁸⁰ CASTELVECCHI, Davide. “Can we open the black box of AI?”. *Nature News*, v. 538, n. 7623, 2016, p. 20.

inteligência não-humana. Ainda é cedo para afirmar se um dia as novas manifestações de inteligências alheias ao ser humano podem ultrapassar as diversas formas de inteligência humana, mas já se faz imperioso discutir mais a fundo os efeitos das inteligências já existentes.

A prevista *singularity*⁸¹, em que a IA vai ser a própria responsável e definir seus fins e meios, ainda é utopia. O palpável no momento é identificar os efeitos jurídicos da IA existente, que já define relações sociais e interfere juridicamente na vida dos indivíduos. E uma destas formas de interferência é a discriminação algorítmica que, conforme foi visto, é fruto de uma decisão automatizada por parte de um sistema de inteligência artificial.

Muitas são as razões que podem levar um sistema, neutro por definição, a tomar uma decisão discriminatória, e uma das mais aparentes foi o treinamento dos sistemas de IA com dados enviesados. A problemática levantada foi que tais dados enviesados podem expor estatísticas que serão base de uma decisão discriminatória, com poderes de até manipular grupos de pessoas e os rumos de eleições democráticas, conforme o caso debatido sobre a IA que filtra o resultado dos motores de busca.

Ainda sobre dados, segundo Barocas, o *Big Data* diz-se neutro, mas não é⁸². Pelo facto desta diferenciação poder ocorrer de forma natural, é necessário o exame dos dados que estão sendo usados para treinar sistemas de IA. Não suficiente, um trabalho constante de melhoria na análise acerca de como o sistema de IA está interpretando os dados também deve ser tomado como prioridade pelas entidades desenvolvedoras. Ou seja, um cuidado sobre o algoritmo que deve acontecer em 3 momentos: preventivo (ainda no seu desenvolvimento), durante o tratamento dos dados e após a tomada da decisão automatizada.

Desta forma, a IA pode aprender e chegar a conclusões sozinhas, mas se esta decisão for discriminatória cabe a quem a supervisiona detetar tal resultado. Em outras palavras, a discriminação algorítmica é um problema a ser concebido pelo ser humano, ao mesmo tempo que requer uma solução também concebida pelo ser humano.

⁸¹ UPCHURCH, Martin. "Robots and AI at work: the prospects for singularity". *New Technology, Work and Employment*, v. 33, n. 3, 2018, p. 205-218.

⁸² BAROCAS, Solon and SELBST, Andrew D. "Big datas disparate impact". *California Law Review*, v. 104, 2016.

Neste sentido, não existe (ainda) um problema de cognição de IA, cujo problema aguardaria uma solução de sensibilidade da própria inteligência artificial. A IA é maleável e pode ser ensinada a não apenas evitar a discriminação, bem como identificá-la e auxiliar numa rápida reparação, tendo em vista os seus efeitos jurídicos de natureza fundamental. Mas para isso, quem analisa a possível discriminação e quem ensina o algoritmo deve analisar o sistema de tomada de decisão e os dados de treinamento de forma mais transparente. Menos opacidade e mais transparência apontam para o caminho de uma IA menos discriminatória.