



VISÃO COMPUTACIONAL E RACISMO ALGORÍTMICO: BRANQUITUDE E OPACIDADE NO APRENDIZADO DE MÁQUINA

Tarcízio da Silva¹

Resumo: Recursos de inteligência artificial estão cada vez mais acessíveis a empresas e governos de todos os tipos, sendo aplicados em sistemas de vigilância, plataformas de mídias sociais e outros tipos de agentes artificiais. Uma de suas aplicações emergente em popularidade é o da visão computacional, que permite reconhecer automaticamente objetos, entidades, conceitos ou características extraídas de imagens, inclusive sobre pessoas. O presente artigo apresenta mapeamento de casos públicos de manifestações racistas de aplicações da visão computacional nos últimos anos lidos sob as lentes da Teoria Racial Crítica e Estudos da Branquitude. Analisando práticas de visibilidade e invisibilidade da pessoa negra nas bases de dados e processamento de recursos de visão computacional, o trabalho reflete como categorias e tipologias são manifestações de como os registros informacionais configuram e reproduzem poder, com impacto real nas possibilidades de representação e expressão equitárias entre diferentes segmentos da sociedade.

Palavras-chave: visão computacional; raça; algoritmos; APIs; representação

COMPUTER VISION AND ALGORITHMIC RACISM: WHITENESS AND OPACITY IN MACHINE LEARNING

Abstract: Artificial intelligence services are increasingly accessible to businesses and governments of all kinds, being applied to surveillance systems, social media platforms and other types of artificial agents. One of its emerging applications in popularity is computer vision, which allows the automatic recognition of objects, entities, concepts or features extracted from images, including people. The article presents mapping of public cases of racist manifestations of computer vision applications in the last years, read under the lens of Critical Racial Theory and Whiteness Studies. Analyzing practices of visibility and invisibility of black people in the databases and processing of computer vision resources, the article discusses how categories and typologies are manifestations of how informational

¹ Tarcízio Silva é graduado em Comunicação - Produção Cultural e mestre em Comunicação e Cultura Contemporâneas pelo Programa de Pós-Graduação em Comunicação e Cultura Contemporâneas da Universidade Federal da Bahia. Atualmente realiza doutorado pelo Programa de Pós-Graduação Interdisciplinar em Ciências Humanas e Sociais da Universidade Federal do ABC. Email: tarushijio@gmail.com / Site: tarciziosilva.com.br



records configure and reproduce power, with real impact on the possibilities of fair representation and expression between different segments of society.

Palavras-chave: computer vision; race; algorithms; APIs; representation

VISIÓN COMPUTACIONAL Y RACISMO DE ALGORITMO: BLANQUITUD Y OPACIDAD EN EL APRENDIZAJE DE MÁQUINAS

Resumen: Las capacidades de inteligencia artificial son cada vez más accesibles para empresas y gobiernos de todo tipo, y se aplican a sistemas de vigilancia, plataformas de redes sociales y otros tipos de agentes artificiales. Una de sus aplicaciones emergentes en popularidad es la visión por computadora, que permite el reconocimiento automático de objetos, entidades, conceptos o características extraídas de imágenes, incluidas las personas. El presente presenta el mapeo de casos públicos de manifestaciones racistas de aplicaciones de visión computacional en la última lectura bajo la lente de los Estudios de Teoría Crítica de Raza y Blanquitud. Analizando las prácticas de visibilidad e invisibilidad de las personas negras en las bases de datos y el procesamiento de los recursos de visión por computadora, el trabajo refleja cómo las categorías y tipologías son manifestaciones de cómo los registros informativos configuran y reproducen el poder, con un impacto real en las posibilidades de representación y expresión equitativas. entre diferentes segmentos de la sociedad.

Palavras-chave: visión computacional; raza; algoritmos; APIs; representación

VISION INFORMATIQUE ET RACISME D'ALGORITHMES: BLANCHEUR ET OPACITÉ DANS L'APPRENTISSAGE AUTOMATIQUE

Resumé: Les capacités d'intelligence artificielle sont de plus en plus accessibles aux entreprises et aux gouvernements, qu'elles soient appliquées à des systèmes de surveillance, à des plateformes de médias sociaux ou à d'autres types d'agents artificiels. La vision par ordinateur, qui permet la reconnaissance automatique d'objets, d'entités, de concepts ou de caractéristiques extraits d'images, y compris des personnes, est l'une de ses applications en plein essor. Cet article présente la cartographie des cas publics de manifestations racistes d'applications de vision computationnelle au cours des dernières années, lues à la loupe des études critiques sur la théorie raciale et la blancheur. Analysant les pratiques de visibilité et d'invisibilité des Noirs dans les bases de données et le traitement des ressources de vision par ordinateur, les travaux montrent comment les catégories et les typologies sont des manifestations de la manière dont les enregistrements informationnels configurent et reproduisent le pouvoir, avec un impact réel sur les possibilités de représentation et d'expression équitables. entre différents segments de la société.

Palavras-chave: vision par ordinateur; race; algorithmes; APIs; représentation

TECNOLOGIAS DIGITAIS DE COMUNICAÇÃO: BRANQUITUDE COMO PADRÃO E REPRESENTAÇÃO



As reflexões sobre representação e grupos minorizados em tecnologias digitais ganharam fôlego desde a década de 1990 inicialmente a partir de desdobramentos sobre os estudos em torno do conceito de fosso digital (“digital divide”) e desigualdade de acesso. Durante os primeiros anos da discussão sobre a web social, como fóruns e blogs textuais, o discurso hegemônico em disciplinas que estudaram a comunicação mediada por computador cogitava uma “descorporificação” dos indivíduos e suas identidades.

Uma vez que a comunicação não-simultânea limitada por texto permitia experimentar a simulação de diferentes marcadores sociais e identitários, parte da pesquisa sobre raça na internet explorava como alteridades raciais e de gênero poderiam ser “experimentadas” em fóruns e posteriormente como avatares visuais em ambientes tais como o *Second Life*. Como aponta Daniels, porém, este olhar evocava uma “linguagem da exploração e descoberta (por exemplo: ‘descobrir identidade racial em um grupo *Usenet*’). Esta retórica foi consistente com as então correntes descrições da Internet como uma ‘fronteira eletrônica’ e evoca o olhar do colonizador” (Daniels, 2013, p. 708, trad. livre). Porém, a autora (2009, 2013) resgata como supremacistas brancos viram a internet como um campo de atuação desde meados dos anos 1990, com sites de desinformação contra figuras históricas tais como Martin Luther King ou ainda a construção de portais com o objetivo de conectar internacionalmente afiliados aos grupos mais extremistas de supremacia branca (como o *Stormfront*).

A evolução de características sociotécnicas da internet como multiplicação de plataformas digitais de interação tais como mídias sociais, conexão pervasiva e *mobile*, alta velocidade de transferência de dados, centralidade da imagem, plataformas de vídeo e aumento da população conectada trouxe mais debates sobre raça e comunicação digital. Os casos midiáticos de racismo online discursivo multiplicaram-se, como os mapeados por Trindade (2018), vitimando, sobretudo mulheres negras em alguma posição de destaque ou desafio ao *status quo*.

É preciso, porém, sublinhar que o racismo online não trata apenas de casos aberracionais de injúrias isoladas, mas sim um “sistema de práticas contra grupos racializados que privilegiam e mantêm poder político, cultural e econômico para os brancos no espaço digital” (TYNES *et al*, 2019, p.195). Estes privilégios são construídos, mantidos e transformados em novas manifestações geralmente elusivas que vão além da materialidade dos discursos e



imagens, indo da contratação privilegiada de grupos raciais hegemônicos (NELSON, TU & HINES, 2001) à aplicação diferencial de punições ligadas aos Termos de Uso das plataformas². Dados da pesquisa EEO-1 nos Estados Unidos, por exemplo, mostram que as empresas do Vale do Silício estão muito longe de representar a diversidade da população do país³ - e devido a concentração global das tecnologias criadas nos EUA, a situação se agrava quando se trata da complexidade e diversidade mundiais.

Considerando que as manifestações do racismo são “construídas e expressas na infraestrutura online ou *back end* (por exemplo: algoritmos) ou através das interfaces (por exemplo: símbolos, imagem, voz, texto e representações gráficas)” (TYNES et al, 2019, p.195, trad. livre) é necessário olhar também para estes materiais relacionados às práticas econômicas e produtivas nos contextos socio-históricos onde emergem. As interfaces e sistemas podem materializar, de forma análoga ao que acontece nos fluxos semânticos da sociedade os "preconceitos não manifestos, presentes invisivelmente na cabeça dos indivíduos, e as consequências dos efeitos da discriminação na estrutura psíquica das pessoas" (CARONE & BENTO, 2017, pós. 74).

Crescentemente, os algoritmos e inteligência artificial ganham destaque no debate sobre comunicação, sociedade e tecnologia por regirem cada vez mais esferas da vida. Recomendação de conteúdo nas *timelines* de mídias sociais, segurança digital, biometria, processamento de linguagem natural e reconhecimento facial são algumas das aplicações já frequentes no cotidiano em nossos computadores e *smartphones*. Mas os problemas destes agentes artificiais que tomam decisões de visibilidade, acesso, classificação e processamento de materiais digitais também são frequentes, muitas vezes ligados a vieses de raça, gênero, classe, localidade, neuroatipicidade e outros. Entender a “racialização tecnológica como uma forma de opressão algorítmica nos permite usá-la como um importante *framework* para criticar discursos sobre a Internet como um ambiente democrático” (NOBLE, 2018, pos. 1341, trad. livre).

Defendemos que os estudos sobre a branquitude são uma chave importante para entender os modos pelos quais as tecnologias automatizadas demonstram continuamente vieses racistas, mesmo provenientes de empresas globais bilionárias com todo aparato tecnológico-

² <https://forward.com/fast-forward/423238/twitter-white-nationalists-republicans/>

³ <https://www.eeoc.gov/employers/eo1survey/about.cfm>



financeiro disponível. Interessada nos mecanismos que compõe o que chamou de “pacto narcísico da branquitude” em contextos organizacionais de empresas e setor público, Bento descreve a branquitude como

um lugar de privilégio racial, econômico e político, no qual a racialidade, não nomeada como tal, carregada de valores, de experiências, de identificações afetivas, acaba por definir a sociedade. Branquitude como preservação de hierarquias raciais, como pacto entre iguais, encontra um território particularmente fecundo nas Organizações, as quais são essencialmente reprodutoras e conservadoras (BENTO, 2002, p.7).

A manutenção e reprodução dos privilégios da branquitude partindo de uma centralidade evocativa à Europa se ligaram histórica e economicamente à dominação colonial e neocolonial, com desdobramentos da ciência à tecnologia, mas sempre através da evitação ao debate sobre raça. Mills chama este esforço coletivo de inverter a epistemologia de uma “epistemologia da ignorância”, um “padrão particular de disfunções globais e locais (que são funcionais psicológica e socialmente), produzindo o resultado irônico que os brancos em geral serão incapazes de entender o mundo que eles próprios formataram” (MILLS, pos. 327, trad. nossa).

Estas disfunções são também registradas nos mitos de racial-*blindness*, pos-racialidade (BONILLA-SILVA, 2006) e democracia racial (NASCIMENTO, 2016[1978]; SOUZA, 2017) como vetores do silenciamento do debate sobre raça em todas as esferas, que acreditamos que se liga plenamente *technochauvinismo*. Este seria “a crença que a tecnologia é sempre a solução”, acompanhada de noções de “computadores são mais “objetivos” ou “enviesados” porque eles destilam questões e respostas em avaliação matemática” (BROUSSARD, 2018, pos. 166, trad. livre).

Propomos que isto implica em uma “dupla opacidade”, que definimos como o modo pelo qual os discursos hegemônicos invisibilizam tanto os aspectos sociais da tecnologia quanto os debates sobre a primazia de questões raciais nas diversas esferas da sociedade – incluindo a tecnologia, recursivamente. A inspiração para estas críticas parte dos pilares da Teoria Racial Crítica (TRC), *framework* teórico-metodológico e ativista desenvolvido na década de 1980 nos EUA, inicialmente para transformar os Estudos Legais no Direito, promovendo a compreensão da sociedade em um panorama da supremacia branca (DELGADO &



STEFANCIC, 2017). Questionando a invisibilidade legal do racismo em uma sociedade baseada no mesmo, a TRC propõe uso de contra-narrativas, conhecimento interdisciplinar experiencial e agência efetiva no combate à opressão racial. A atuação e temas da TRC trazem muitos pontos em comum com o pensamento social negro brasileiro (FERREIRA & QUEIROZ, 2018; SILVA, 2019a) e tem sido crescentemente usada internacionalmente para entender a tecnologia (ALI, 2014; BROCK, 2016; NOBLE, 2018; BENJAMIN, 2019).

A ideia de “caixa preta”, que se tornou recurso explicativo na área de Estudos da Ciência e Tecnologia, se refere a dispositivos que funcionam em termos de *entrada* de informações e decisões e *saída* de resultados e operações que ocultam os modos pelos quais ciência e tecnologia são construídas. Falando de dispositivos tecnológicos e softwares, “abrir a caixa preta” significa observar “não somente que qualquer ordem social é impactada por desenvolvimento tecnológico [...] mas também que normas, ideologias e práticas sociais são parte constitutivo de desenhos técnicos” (BENJAMIN, 2019, p. 72).

Desta forma, o referencial de estudos sobre a “branquitude” é elaborado a seguir como ferramenta para abrir as caixas-pretas da visão computacional de forma crítica ao racismo algorítmico.

VISÃO COMPUTACIONAL: INVISIBILIDADE, VISIBILIDADE E INTERPRETAÇÃO RACISTAS

A computação visual (*visual computing*) é um termo que engloba áreas da computação e iniciativas interdisciplinares dedicadas a construir sistemas para o entendimento automatizado de imagens e vídeos. Ou seja, busca-se fazer softwares e hardwares computacionais capazes de entender objetos, entidades e situações a partir de informação visual, como imagens ou *frames* de vídeo. Suas aplicações são amplas, da indústria e mecatrônica até o marketing e publicidade.

Especificamente a visão computacional (*computer vision*) “se refere à coleta, análise e síntese de dados visuais através de computadores, com objetivos diversos como a identificação de rostos e biometria, a análise de representações de objetos, entidades, conceitos e contextos em imagens, entre outros” (Wang, Zhang & Martin, 2015, trad. livre). Foi popularizada com os buscadores, mídias sociais e aplicativos *mobile*. Por exemplo, a

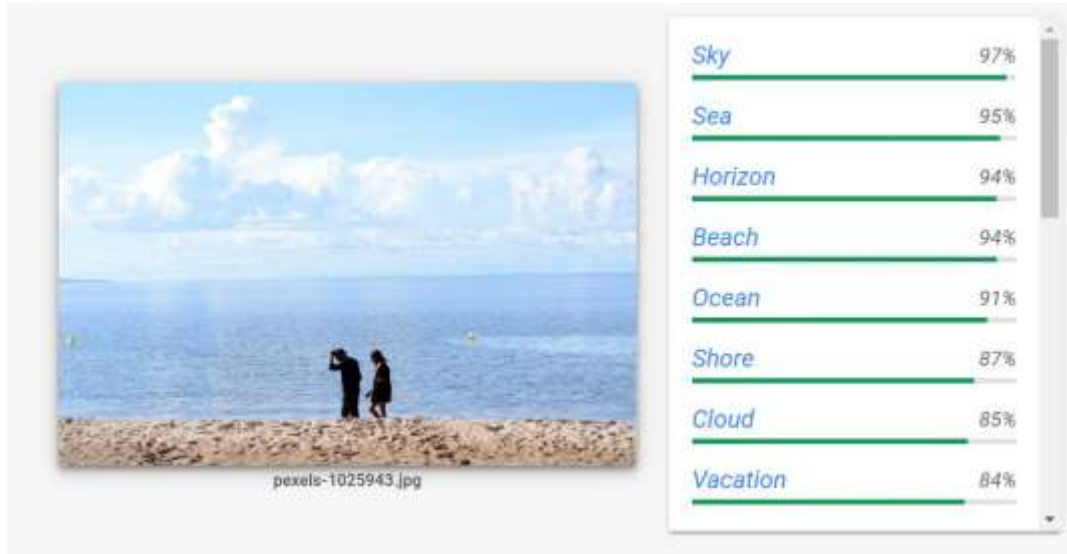
busca Google que identifica imagens similares e quais objetos estão fotografados; a marcação automática de rostos de amigos em fotos no Facebook; e os filtros lúdicos do Snapchat e Instagram são todas aplicações que entraram no dia a dia de grande parte dos usuários contemporâneos de tecnologias digitais.

Os sistemas passam a saber como identificar rostos, objetos e contextos nas imagens através do que é chamado de “aprendizado de máquina” (*machine learning*), um campo da inteligência artificial restrita que trata do reconhecimento de padrões através de uma base de dados e posterior aplicação do aprendizado no reconhecimento das variáveis em outras unidades ou conjuntos de dados (Oliveira, 2018). No caso da visão computacional, trata-se de “treinar” o sistema alimentando-o com um número relevante de imagens já marcadas (por ex: centenas de imagens da classe “cachorro”; centenas de imagens de uma raça específica e assim por diante) para que o software “entenda” novas imagens que não foram anteriormente vistas.

A ampla disponibilidade destes recursos hoje é um fato não só para empresas altamente especializadas em inteligência artificial, mas para todo um rol de empresas de mídia e tecnologia que podem usar os recursos oferecidos pelas grandes corporações da área. Empresas como IBM, Google, Microsoft e Amazon oferecem serviços de inteligência artificial e computação cognitiva através do modelo *API* (*application programming interfaces*) que permite que qualquer desenvolvedor com o mínimo de conhecimento contrate recursos de inteligência artificial tais como processamento de linguagem natural, descoberta de padrões em dados tabulares, modelos de aprendizado de máquina e visão computacional. Na *Figura 1* podemos ver uma captura de tela da demonstração do recurso de visão computacional da Google, chamado de Google Vision:

Figura 1: Captura de tela da Google Vision⁴

⁴ O recurso da Google Vision pode ser testado no endereço web <https://cloud.google.com/vision/>. A foto utilizada é de autoria de Dazzle Jam e está disponível através da licença *creative Commons* em <https://www.pexels.com/@dazzle-jam-374075>.



Como é possível ver acima, o recurso identifica elementos, cena, objetos e conceitos. Na fotografia de exemplo, o recurso identificou céu, mar, horizonte e o possível contexto de férias, entre outros. Recursos como a Google Cloud Vision e outros citados também prometem identificar expressão facial de rostos, padrões de cores, elementos sensíveis (como pornografia ou violência), personalidades famosas, logotipos e marcas, identificar imagens similares na web e outros grupos de entidades em constante desenvolvimento (ETLINGER, 2016; MINTZ *et al*, 2019),

Entretanto, a aplicação da visão computacional traz problemáticas e controvérsias específicas a diversos campos. Neste trabalho o objetivo é focar no uso de tecnologias públicas como mídias sociais, computadores pessoais, aplicativos *mobile* e outros recursos disponíveis para a maior população no cotidiano. Na outra ponta estão tecnologias usadas por corporações e governos, especialmente reconhecimento facial para fins de vigilância ou policiamento. Entender estas tecnologias de forma múltipla é relevante pois a experiência do afrodiaspórico foi construída pelo colonialismo sob uma lógica "vigilância racializada como tecnologia de controle social onde práticas, políticas e performances de vigilância remetem a produção de normas pertinentes a raça" (BROWNE, 2015, p. 32, trad. livre) e se relacionam de diversos modos a tecnologias e ideologias carcerais (Benjamin, 2019) que vão muito além da prisão, mas definem o lugar e o não-lugar do negro, passível sempre a violência física ou simbólica.

Em aplicações como as que veremos, entretanto, três pontos são especialmente relevantes na relação com as tecnologias do cotidiano: são recursos que normalizam a identificação algorítmica e registro em *big data* dos indivíduos; são dispositivos usados pelas corporações para o treinamento de suas bases de dados para diversos fins; e se tratam de dinâmicas de comunicação e interação com impactos cumulativos na autoestima e saúde mental dos indivíduos (PIERCE, 1970; TYNES *et al*, 2018; Silva, 2019b).

A *Tabela 1* mostra alguns dos casos mais famosos de problemas identificados por usuários ou jornalistas com resultados nocivos a indivíduos negros ou população negra de modo geral. Procederemos à descrição breve destes casos, para demonstrar o paralelismo com características da branquitude nas epistemes da tecnologia das plataformas de comunicação e inteligência artificial.

Tabela 1: Casos Seleccionados de Manifestação de Viés/Racismo Algorítmico em Visão Computacional

Caso / Reação	Data	Problema/Viés Algorítmico	Causa/problema “técnico”
Reconhecimento de faces de computador da HP não reconhece usuário negro	12/2009	Desumanização; Invisibilidade	<ul style="list-style-type: none"> • Base de dados insuficiente • Ausência de testes
Google marca pessoas negras como gorilas	07/2015	Representação e associação racista; desumanização	<ul style="list-style-type: none"> • Base de dados insuficiente • Base de dados com associações racistas intencionais • Ausência de testes
Robôs interagentes não encontram rosto de mulher negra	03/2017	Desumanização; Invisibilidade	<ul style="list-style-type: none"> • Base de dados insuficiente • Ausência de testes
Faceapp embranquece pele para deixar “mais bonita” a selfie	04/2017	Representação eurocêntrica de beleza; desumanização	<ul style="list-style-type: none"> • Base de dados insuficiente • Ausência de testes
APIs não reconhecem gênero e idade de mulheres negras	02/2018	Representação eurocêntrica de gênero e idade	<ul style="list-style-type: none"> • Base de dados insuficiente • Ausência de testes
Kairos retira do ar aplicativo de “diversidade”	06/2018	Tipologia racial essencialista	<ul style="list-style-type: none"> • Tipologia centrada nos EUA
APIs de análise de expressões faciais associam emoções negativas a negros	01/2019	Percepção eurocêntrica; estereotipização	<ul style="list-style-type: none"> • Base de dados insuficiente • Ausência de testes



Google Vision confunde cabelo negro com peruca	02/2019	Reforço de apropriação cultural; desumanização	<ul style="list-style-type: none">• Base de dados insuficiente• Base de dados com exemplos de apropriação estético-cultural• Ausência de testes
Carros autônomos tem mais chance de atropelar pessoas negras	03/2019	Desumanização; risco físico direto	<ul style="list-style-type: none">• Base de dados insuficiente• Ausência de testes

Um dos primeiros casos de problemas quanto de visão computacional racista a viralizar aconteceu em 2009 com computadores da Hewlett-Packard (HP). Em vídeo publicado no YouTube, um homem negro e uma mulher branca testam um computador da HP em uma loja de eletrônicos. O recurso MediaSmart de rastreamento de movimento de rostos conseguia identificar o rosto da mulher branca, mas não o rosto do homem negro. Com rápida viralização (400 mil visualizações em poucos dias, hoje o vídeo tem mais de 3 milhões de visualizações), o caso ganhou cobertura jornalística e um porta-voz da empresa alegou problemas que ocorrem quando “iluminação de fundo não é suficiente”⁵.

Em 2015, ao lançar o recurso de etiquetagem automática de fotos em seu aplicativo *Google Photos*, o programador Jacky Alcine teve fotos suas e de sua namorada marcadas com a tag “gorilas”. Ao divulgar o insulto no Twitter⁶, um engenheiro da Google pediu desculpas, não prometeu ações concretas, mas divulgou publicação no portal da *Fortune* que dizia que o sistema de reconhecimento de faces da Google é considerado o melhor do mercado. Em 2018, jornalistas da *Wired* realizaram testes e mostraram que a solução da empresa foi simplesmente eliminar a tag “gorila” do rol possível na ontologia do recurso⁷.

Outro exemplo de confusão de “tags” ligadas a indivíduos e população negras foi realizado em projeto de interrogação de APIs de visão computacional e bancos de imagens. Ao comparar como sites de bancos de imagens como Shutterstock representam fotos de diferentes países, estudo identificou que diversas fotos de mulheres com cabelos crespos volumosos foram marcadas com a tag “wig”, que significa “peruca”. Este erro possivelmente é resultante de uma base de treinamento mais rica em fotos de pessoas brancas em contexto de apropriação cultural-estética – como a prática de usar perucas afro como fantasias em festas e carnaval. Os autores concluem que “arquiteturas algorítmicas favorecem a

⁵ <https://www.wired.com/2009/12/hp-notebooks-racist/>

⁶ <https://www.theverge.com/2015/7/1/8880363/google-apologizes-photos-app-tags-two-black-people-gorillas>

⁷ <https://www.wired.com/story/when-it-comes-to-gorillas-google-photos-remains-blind/>



performance de neutralidade cultural, mas com categorias genéricas com muito mais destaque do que termos específicos” (MINTZ et al, 2019) onde esta ideia de neutralidade é marcada socialmente como a percepção dos desenvolvedores.

Ainda sobre identificação de características de imagens, o reconhecimento de expressões por visão computacional permitiria marcar rostos em tipologias de emoções humanas básicas organizadas pela Psicologia, tais como Raiva, Medo, Surpresa, Nojo e Tristeza. Experimento de Rhue (2019) com conjunto de fotos cuidadosamente selecionadas de atletas para testar como os fornecedores *Face++* e *Microsoft* identificaria imagens equivalentes de atletas brancos e negros. Os resultados consistentemente atribuíram emoções/expressões negativas aos atletas negros. Rhue conclui que “o uso de reconhecimento racial pode formalizar estereótipos preexistentes em algoritmos, automaticamente incorporando-os na vida cotidiana” (RHUE, 2018, s.p., trad. livre).

Outra nota particularmente interessante foi emitida pelo CEO do aplicativo *FaceApp*. O aplicativo de edição de selfies possuía um filtro de “embeleazar” o rosto dos usuários. Uma das principais edições automáticas era clarear a pele, gerando resultados aberrantes em fotos de pessoas negras ou indianas, por exemplo. Depois da divulgação dos problemas, Yaroslav Goncharov, CEO do aplicativo, alegou que é “um infeliz efeito colateral da rede neural subjacente causado pelo conjunto de dados de treinamento, não comportamento esperado”⁸. Como estamos falando de aprendizado de máquina (*machine learning*), os dados de treinamento são fator essencial em como o sistema vai performar. Ao mesmo tempo em que algumas posturas alegam que o “algoritmo em si” seria neutro, alegações como a de Goncharov são comuns ao fugir da responsabilidade sobre a seleção dos dados de treinamento, assim como aconteceu no caso citado anteriormente. Podemos evocar aqui como a invisibilidade do negro é “elemento importante da identidade do branco: ele não vê o negro. Uma reflexão sobre relações raciais pode explicitar um desconforto do branco diante da paradoxal constatação que ele não vê, não lembra, nunca pensou nos negros” (BENTO, 2002, p.91). Comumente aplicativos que buscam alcance global projetam seus consumidores apenas como brancos. Ao viralizar novamente em 2019 em terras brasileiras, o problema do embranquecimento permanece pervasivo no recurso de envelhecimento do app⁹.

⁸ <https://www.mirror.co.uk/tech/faceapp-apologises-hot-selfie-filter-10293590>

⁹ <https://www.estadao.com.br/infograficos/link,alem-de-envelhecer-faceapp-embranquece-rostos-negros,1018384>



Uma vez que boa parte das empresas supracitadas também desenvolvem sistemas para vigilância e policiamento, biometria e reconhecimento facial, o potencial de impactos fatais contra a vida de grupos racializados parece já ser óbvia. Mas um caso a ser citado neste artigo pode mostrar a pervasividade do problema além das plataformas e aplicativos de comunicação. Em estudo publicado neste ano, Wilson, Hoffman e Morgenstern (2019) analisaram 8 sistemas de reconhecimento de imagens usados em carros autônomos e a acurácia na identificação de pessoas com pele escura poderia ser 5% menor, resultando em potenciais atropelamentos. Os autores concluem a necessidade de se olhar para o “real problema que pode surgir se este tipo de fonte de viés de captura não for considerado antes de distribuir estes tipos de modelos de reconhecimento” (WILSON, HOFFMAN e MORGENSTERN, 2019, p.9, trad. livre).

Os casos acima, portanto, mostram que os modos pelos quais a visibilidade ou invisibilidade de brancos e negros são contextuais de acordo com as vantagens e desvantagens de cada ponto dessa dicotomia nas mídias e tecnologias em questão. A invisibilidade negra surge como uma questão de fato enquanto a invisibilidade branca é um conceito muito diferente pois se vincula à sua padronização e referência como universal e, como relata Machado “não porque nossa cor branca passa despercebida, pelo contrário, porque nossa cor é visível, porém tal visibilidade é silenciada e representa o universal garantindo a nós brancas/os, posições de privilégios, das quais ninguém quer abrir mão” (MACHADO, 2018, p.381).

REAÇÕES E CONTRA-NARRATIVAS: CÓDIGOS CONTRA-HEGEMÔNICOS

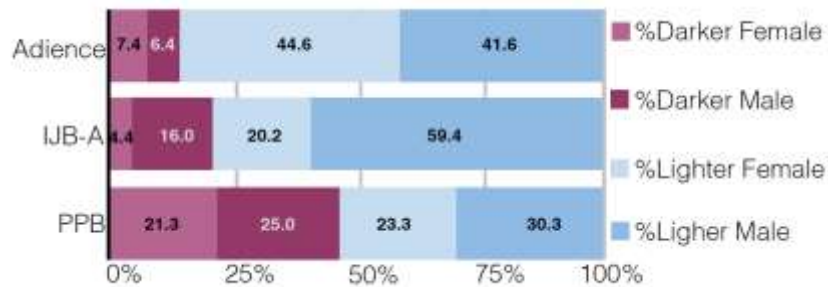
E quais as reações possíveis? Três casos de auditoria, reações e ajustes levados a cabo por desenvolvedores negros pode nos mostrar alternativas possíveis e responsáveis.

Em um dos casos ligados à identificação, podemos destacar os experimentos e projeto *GenderShades.org* de Buolamwini e colaboradoras. Em experimento inovador, Buolamwini e Gebu (2018) testaram os recursos de identificação de idade e gênero fornecidos por Microsoft, Face++ e IBM. Elas descobriram que os recursos falhavam com muito mais frequência em fotos de mulheres negras, no que chamaram de “disparidade interseccional”. Ao investigar as causas, descobriram que o conjunto de dados para “treinamento” do sistema



era muito enviesado, com muito mais fotos de homens de pele clara (sobretudo caucasianos) do que mulheres de pele escura. A *Figura 2* mostra duas destas bases enviesadas usadas pelos sistemas (Adience e IJB-A) e uma base de treinamento criada por elas, com o nome de *Pilot Parliament Benchmark*, com uma boa distribuição de gênero e fenótipo quanto à cor da pele. O trabalho teve um impacto relevante no campo pois, sozinhas, elas conseguiram fazer um sistema mais preciso neste aspecto do que o oferecido por aquelas corporações.

Figura 2: Comparação de datasets de treinamento



Fonte: BUOLAMWINI & GEBRU, 2018.

É relevante citar que uma das motivações para a trajetória de Buolamwini foi a experiência real, em primeira mão, de racismo algorítmico. Durante sua graduação na Georgia Tech, a estudante foi a diversas feiras de tecnologia e empreendedorismo. Por mais de uma vez os robôs conseguiam encontrar os rostos de todos seus colegas brancos – exceto ela. Para mostrar o absurdo dos problemas enviesados destes aplicativos e robôs, a pesquisadora usou uma máscara branca genérica, sem feições – e até esta máscara foi reconhecida, em uma reedição involuntária da metáfora de *pele negra, máscaras brancas* de Frantz Fanon (FANON, 2008). Em palestra no TEDx onde falou sobre a questão e sua jornada acadêmica, Buolamwini convoca espectadores a se conectarem pela “criação de um mundo onde a tecnologia trabalhe em favor de todos, não apenas em favor de alguns, um mundo onde valorizemos a inclusão e tenhamos como foco a mudança social”¹⁰.

Em 2019, as pesquisadoras retornaram aos dados do experimento citado acima para analisar se a divulgação dos problemas gerou impacto na diminuição da disparidade. De fato, a iniciativa gerou impacto e os sistemas da Microsoft e IBM diminuíram o abismo da taxa de

¹⁰ https://www.ted.com/talks/joy_buolamwini_how_i_m_fighting_bias_in_algorithms

erros na acurácia dos sistemas. As autoras lembram, na conclusão, que a importância deste tipo de auditoria vai muito além do campo das tecnologias de comunicação, uma vez que “o potencial de abuso e aplicação bélica das tecnologias de análise facial não podem ser ignorados nem as ameaças à privacidade ou infrações de liberdades civis diminuídas mesmo quando as disparidades de precisão diminuíam” (RAJI & BUOLAMWINI, 2019, p.6, trad. livre).

Depois que o primeiro estudo foi publicado, tanto a IBM quanto a Microsoft emitiram notas afirmando compromisso com a busca pela equidade nos resultados¹¹. A Microsoft emitiu nota breve, mas a IBM detalhou longamente projetos internos e experimento próprio defendendo o uso responsável de suas tecnologias. Apesar disso, relatórios mostram que a empresa vende a tecnologia para projetos repressivos de governos, como nos Emirados Árabes Unidos¹².

Iniciativa similar de transformação da tecnologia tem sido realizada por Jacky Alciné, o desenvolvedor que foi vítima da marcação racista pela *Google Photos* citada na seção anterior. Com ampla experiência em programação, o profissional lançou em 2018 o projeto *black.af* que desenvolve produtos, serviços e aplicativos para “fornecer ferramentas e plataformas a pessoas para que possam usá-las e expandir seu impacto e alcance em projetos de apoio social, promoção da justiça e comércio digital”¹³. Assim como outros desenvolvedores de grupos minorizados, Alciné busca ligar a iniciativa comercial à transformação social do mundo a partir de um ponto de vista muito particular e ainda raro nas empresas de tecnologia. Sua iniciativa evoca os projetos da Teoria Racial Crítica que propõe como um dos seus pilares o conhecimento experiencial na pesquisa e na *práxis* buscando se associar a conhecimento criado por grupos minorizados para “revelar os modos pelos quais raça, classe, gênero e outras formas de opressão interagem para mediar as experiências e realidades daqueles afetados por tais opressões” (MALAGON, HUBER & VELEZ, p.264, trad. livre).

¹¹ IBM <http://gendershades.org/docs/ibm.pdf> e Microsoft: <http://gendershades.org/docs/msft.pdf>

¹² <https://gizmodo.com/ibm-sells-face-recognition-surveillance-to-a-dictatorsh-1835101881>

¹³ <https://black.af/>



Os casos citados foram identificados tanto por cientistas e engenheiras da computação quanto por ativistas, jornalistas e pesquisadoras das ciências sociais e humanidades. Concordamos com Osoba e Welser IV ao dizer que:

a pesquisa técnica em vieses no aprendizado de máquina e inteligência artificial ainda está em sua infância. Questões sobre vieses e erros sistêmicos em algoritmos demandam um diferente tipo de sabedoria de cientistas de dados e criadores de algoritmos. Estes profissionais são comumente engenheiros e cientistas com menos exposição a questões de políticas públicas ou sociais (OSOBA & WELSER IV, 2017, p.24, trad. livre).

Deste modo, a pouca exposição à alteridade, sobretudo em áreas como Computação e Engenharias, reforça uma “invisibilidade, distância e um silenciamento sobre a existência do outro [...]. A racialidade do branco é vivida como um círculo concêntrico: a branquitude se expande, se espalha, se ramifica e direciona o olhar do branco” (BENTO, 2017, pos. 645). Junto a iniciativas tanto de experimentação e auditoria algorítmica como as citadas e promoção da ocupação de lugares de reflexão e poder na tecnologia por grupos racializados (DANIELS, NKONDE & MIR, 2019), faz-se necessário também – pela branquitude consciente – buscar desenvolvimento de competência cultural diversa (SUE, 2001; JAIME, 2014).

Um último caso que podemos citar de reação a controvérsias ligadas a racismo algorítmico e visão computacional foi a polêmica em torno do aplicativo *Diversity Recognition* da Kairos. Lançado em 2017, o aplicativo Diversity Recognition buscou divulgar a empresa Kairos através da identificação de traços fenotípicos da "herança genética" dos usuários. Ao enviar uma selfie para o aplicativo, cada usuário recebia sua selfie com um gráfico sobreposto, onde as categorias Hispânica, Negra, Branca, Asiática ou "Outros" eram representadas através de porcentagem. Usado por mais de 5 milhões de pessoas, o aplicativo logo se tornou fonte para memes humorísticos e conteúdo ofensivo. Porém também críticas muito relevantes de usuários¹⁴ que não se encaixam nos grupos étnico-raciais listados.

Em junho de 2018, o CEO da Kairos, o desenvolvedor negro Brian Brackeen publicou texto com o título “Nós aposentamos o aplicativo Diversity Recognition: estes são os

¹⁴ <https://medium.com/@BrianBrackeen/diversity-gone-viral-c7d5c132029d>



motivos”. Ao longo da publicação Brackeen cita controvérsias de casos de visão computacional aplicados a policiamento, deixando explícita sua decisão de não oferecer os serviços da Kairos para este tipo de uso. Especificamente sobre o Diversity Recognition, Brackeen “enquanto a maioria dos usuários vão receber um resultado preciso, nós reconhecemos que as categorias de etnicidade atualmente oferecidos (Negros, Brancos, Asiáticos, Hispânicos, “Outros”) não representam a diversidade rica e rapidamente cambiante da complexidade da cultura e taça”¹⁵. Além disto, ao aplicar uma tipologia específica do ponto de vista estadunidense hegemônico a um recurso usado em todo o mundo focando em características fenotípicas, podemos lembrar que a identidade branca criou um fenótipo específico e cambiante sempre a serviço de “demarcar a superioridade dos colonizadores, dos indivíduos pertencentes ao grupo que impôs sua hegemonia em prol da exploração dos povos nativos” (BASTOS, 2016, p.228).

Estes três casos citados mostram as possibilidades de se pensar a diversidade nos ciclos de treinamento e disseminação de sistemas e algoritmos. Especificamente mostram pontos de vista paradoxalmente privilegiados de desenvolvedoras e desenvolvedores negros que, ao serem afetados pelos padrões da branquitude na tecnologia, agem por projetos alternativos. Para além do papel de especialistas em programação, entretanto, um engajamento social crítico mais amplo da sociedade com a tecnologia é necessário. Daniels, Nkonde e Mir (2019) citam os três pilares da literacia racial crítica da tecnologia que poderiam ter impacto até em grandes organizações globais:

Compreensão intelectual de como racismo estrutural opera em algoritmos, plataformas de mídias sociais e tecnologias ainda não desenvolvidas; inteligência emocional sobre como resolver situações racialmente estressantes em organizações e; compromisso na tomada de ação para reduzir danos a grupos racializados (DANIELS, NKONDE & MIR, 2019, p.2)

Investigações e auditorias como as citadas se multiplicam, entretanto o volume em que algoritmos e sistemas de inteligência artificial são publicados, divulgados e aplicados em tecnologias públicas é muito maior do que seus testes ou otimizações. Ao analisar artigos publicados em uma das principais conferências de aprendizado de máquina e inteligência artificial do mundo, a Neural Information Processing Systems Conference de 1987 a 2017,

¹⁵ <https://www.kairos.com/blog/we-ve-retired-our-diversity-recognition-app-here-s-why>

Epstein (2018) observou que o *gap* entre o número de trabalhos propondo novos modelos e trabalhos estudando modelos existentes aumenta ano a ano – em 2017 os novos modelos eram 10 vezes mais numerosos do que o estudo de modelos existentes. Como a sociedade racista e desvantagens estruturais decorrentes resultam em uma pequena quantidade relativa de desenvolvedores e gerentes de produto de grupos minorizados, é especialmente premente combater “a interiorização/internalização da branquitude e do racismo entre brancos e negros, se aprofundando na supremacia branca como um fenômeno perigoso e expansionista dentro dos processos sociais e educacionais” (ROSSETTO, 2014, p.131).

Podemos falar de uma sociedade algorítmica (GILLESPIE, 2014; PASQUALE, 2015; O’NEIL, 2016; SILVEIRA, 2017) onde procedimentos do tecnoliberalismo – como classificação, pseudo-meritocracia, mensuração e vigilância - são automatizados e consolidados em caixas-pretas que permitem poucos desvios. Os casos citados são numerosos, mas apresentam apenas uma pequena parte dos problemas que podem ser analisados ou interrogados por métodos criativos. A regulação e transparência é necessária, mas a sociedade americana construiu a ideologia do Vale do Silício e seus pólos de tecnologia de modo a normalizar branquitude e masculinidade, com decorrentes impactos nos modos e níveis de avaliação dos procedimentos de treinamento de máquina (BROUSSARD, 2018; NOBLE, 2018). E no campo da educação e pesquisa, concordamos e expandimos em escopo as prerrogativas que Rossatto aponta, vendo como adequados também aos currículos interdisciplinares de sociedade e tecnologia que incluam estudos anti-racistas que “desconstruam a interiorização/internalização da branquitude e do racismo entre brancos e negros, se aprofundando na supremacia branca como um fenômeno perigoso e expansionista” (ROSSATTO, 2014, p.131).

CONCLUSÕES

A visão computacional é um grupo de tecnologias cada vez mais relevante na sociedade contemporânea, com impacto em práticas de mercado e gestão governamental. Apresentamos alguns casos de problemas em viés algorítmico que demonstram a dificuldade de se debater o que chamamos de “dupla opacidade” – o caráter difuso tanto da tecnologia,

vista erroneamente como neutra, quanto das relações étnico-raciais na sociedade e, por consequente, na tecnologia.

Nos mercados e ambientes produtivos de tecnologia de ponta, concentrados em pólos como Vale do Silício, a diversidade é rara, o que tem impactos materiais e simbólicos nas interfaces e sistemas usados por grande parte das populações mundiais. Tanto os problemas identificados como parte da reação dos desenvolvedores quando contestados, mostram que “a hegemonia da brancura presente em todos os âmbitos sociais não colabora para que os indivíduos brancos passem a questionar seus privilégios bem como se importar com as desvantagens impostas aos demais grupos” (BASTOS, 2016, p.227).

Enquanto programa de pesquisa, a análise crítica das tecnologias como visão computacional viabilizada por inteligência artificial e aprendizado de máquina se mostra urgente para os campos das ciências sociais. Uma vez que a “geração de dados é um fenômeno social reflete de vieses humanos, aplicar algoritmos “proceduralmente corretos” a dados enviesados é um jeito de ensinar agentes artificiais a imitar qualquer viés que os dados contenham” (OSOBA & WELSER IV, 2017, p. 17).

Com o mundo imerso em ideologia hiper-neo-liberal, a ciência racial ganha nova relevância com a desregulação de ciência e investimento do capital financeiro a favor da segmentação dos corpos quanto a genética, biomedicina e vigilância (MBEMBE, 2001, 2016; ROBERTS, 2011; BROWNE, 2015). Apesar de que “categorias raciais não são simples propriedades de pessoas individuais, mas sim resultados complexos de processos sociais raramente capturados no paradigma do aprendizado de máquina” (BENTHALL & HAYNES, 2019, p.3, trad. livre), temos visto como a desigualdade brutal tanto nos dados de treinamento quanto na preocupação com procedimentos de ajuste dos sistemas vitimizam populações já em desvantagem. A construção de contra narrativas e espaços de reflexão e ação sobre inteligência artificial e dataficação do mundo por desenvolvedores, sociólogos e cientistas negros de vários campos é cada vez mais premente.

BIBLIOGRAFIA

ALI, Syed Mustafa. Race: The difference that makes a difference. *tripleC: Cognition, Communication, Co-operation*, v. 11, n. 1, p. 93-106, 2013.



BASTOS, Janaína Ribeiro Bueno. O lado branco do racismo: a gênese da identidade branca e a branquitude. *Revista da Associação Brasileira de Pesquisadores/as Negros/as (ABPN)*, v. 8, n. 19, p. 211-231, 2016.

BENJAMIN, Ruha. *Race After Technology: Abolitionist Tools for the New Jim Code*. John Wiley & Sons, 2019.

BENTHALL, Sebastian; HAYNES, Bruce D. Racial categories in machine learning. In: *Proceedings of the Conference on Fairness, Accountability, and Transparency*. ACM, 2019. p. 289-298.

BENTO, Maria Aparecida Silva. *Pactos narcísicos no racismo: branquitude e poder nas organizações empresariais e no poder público*. Universidade de São Paulo, São Paulo, 2002.

BENTO, Maria Aparecida Silva. Branqueamento e branquitude no Brasil. In: CARONE, Iray; BENTO, Maria Aparecida Silva. *Psicologia social do racismo: estudos sobre branquitude e branqueamento no Brasil*. Editora Vozes Limitada, 2017.

BONILLA-SILVA, Eduardo. *Racism without racists: Color-blind racism and the persistence of racial inequality in the United States*. Rowman & Littlefield Publishers, 2006.

BROUSSARD, Meredith. *Artificial unintelligence: How computers misunderstand the world*. MIT Press, 2018.

BUOLAMWINI, Joy; GEBRU, Timnit. Gender shades: Intersectional accuracy disparities in commercial gender classification. In: *Conference on Fairness, Accountability and Transparency*. 2018. p. 77-91.

CARONE, Iray; BENTO, Maria Aparecida Silva. *Psicologia social do racismo: estudos sobre branquitude e branqueamento no Brasil*. Editora Vozes Limitada, 2017.

COTTOM, Tressie McMillan. *Black Cyberfeminism: Ways Forward for Classification Situations, Intersectionality and Digital Sociology*. 2016.

DANIELS, Jessie. *Cyber racism: White supremacy online and the new attack on civil rights*. Rowman & Littlefield Publishers, 2009.

DANIELS, Jessie. Race and racism in Internet studies: A review and critique. *New Media & Society*, v. 15, n. 5, p. 695-719, 2013.

DANIELS, J.; NKONDE, M.; MIR, D. *Advancing Racial Literacy in Tech*. Relatório do Data & Society Fellowship Program. 2019.

DELGADO, Richard; STEFANCIC, Jean. *Critical race theory: An introduction*. NYU Press, 2017.
EPSTEIN, Ziv et al. Closing the AI Knowledge Gap. *arXiv preprint arXiv:1803.07233*, 2018.

ETLINGER, Susan. Image Intelligence: Making Visual Content Predictive. Altimeter Group. Online, 2016. Disponível em: <<https://www.prophet.com/2016/07/image-intelligence-making-visual-content-predictive>>

FANON, Frantz. *Pele negra, máscaras brancas*. SciELO-EDUFBA, 2008.



GILLESPIE, Tarleton. The relevance of algorithms. *Media technologies: Essays on communication, materiality, and society*, v. 167, 2014.

JAIME, Pedro. Interculturalidade e comunicação nas organizações: Reflexões antropológicas e notas para uma agenda de pesquisas. *Organicom*, v. 11, n. 21, p. 77-89, 2014.

MACHADO, Eliana Sambo. Visibilidade não marcada da Branquitude: discursos de mulheres brancas acadêmicas. *Revista da Associação Brasileira de Pesquisadores/as Negros/as (ABPN)*, v. 10, p. 375-398, 2018.

MALAGON, Maria C.; HUBER, Lindsay Perez; VELEZ, Veronica N. Our experiences, our methods: Using grounded theory to inform a critical race theory methodology. *Seattle J. Soc. Just.*, v. 8, p. 253, 2009.

MBEMBE, Achille. *Políticas da Inimizade*. Lisboa (Portugal: Antígona, 2017).

MILLS, Charles W. *The racial contract*. Cornell University Press, 2014.

MINTZ, André. Máquinas que veem: visão computacional e agenciamentos do visível. In: MENOTTI, Gabriel; BASTOS, Marcus; MORAN, Patrícia (orgs.). *Cinema apesar da imagem*. São Paulo: Intermeios, 2016.

MINTZ, A.; SILVA, T.; GOBBO, B.; PILIPETS, E.; AZHAR, H.; TAKAMITSU, H.; OMENA, J.; OLIVEIRA, T. *Interrogating Vision APIs*. Data Sprint Report. Smart Data Sprint, Lisboa – Portugal, 2019. Disponível em <https://smart.inovamedialab.org/smart-2019/project-reports/interrogating-vision-apis/>

NASCIMENTO, Abdias. *O genocídio do negro brasileiro: processo de um racismo mascarado*. São Paulo: Editora Perspectiva SA, 2016.

NELSON, Alondra; TU, Thuy Linh Nguyen; HINES, Alicia Headlam (Ed.). *Technicolor: Race, technology, and everyday life*. NYU Press, 2001.

NOBLE, Safiya Umoja. *Algorithms of oppression: How search engines reinforce racism*. NYU Press, 2018.

OLIVEIRA, Carla. Aprendizado de máquina e modulação do comportamento humano. In: SOUZA, Joyce; AVELINO, Rodolfo; SILVEIRA, Sérgio Amadeu. *A Sociedade de Controle: manipulação e modulação nas redes digitais*. São Paulo: Hedra, 2018

OSOBA, Osonde A.; WELSER IV, William. *An intelligence in our image: The risks of bias and errors in artificial intelligence*. Rand Corporation, 2017.

PASQUALE, Frank. *The black box society*. Harvard University Press, 2015.

RAJI, I.; BUOLAMWINI, J.. *Actionable auditing: Investigating the impact of publicly naming biased performance results of commercial ai products*. In: AAAI/ACM Conf. on AI Ethics and Society. 2019.

ROBERTS, Dorothy. *Fatal invention: How science, politics, and big business re-create race in the twenty-first century*. New Press/ORIM, 2011.

ROSSATTO, César Augusto. A transgressão do racismo cruzando fronteiras: estudos críticos da branquitude: Brasil e Estados Unidos na luta pela justiça racial. *Revista da Associação Brasileira de Pesquisadores/as Negros/as (ABPN)*, v. 6, n. 13, p. 120-133, 2014.

RHUE, Lauren. *Emotion-reading tech fails the racial bias test*. The Conversation, 2019. Disponível em < <https://theconversation.com/emotion-reading-tech-fails-the-racial-bias-test-108404> >

SILVA, Tarcízio. *Teoria Racial Crítica e Comunicação Digital: conexões contra a dupla opacidade*. Anais do 42º Congresso Brasileiro de Ciências da Comunicação. Belém, Pará, 2019a.

SILVA, Tarcízio. *Racismo Algorítmico em Plataformas Digitais: microagressões e discriminação em código*. Anais do VI Simpósio Internacional LAVITS. Salvador (Bahia), Brasil, 2019b.

SILVEIRA, Sérgio Amadeu da. *Tudo sobre Tod@s: redes digitais, privacidade e venda de dados pessoais*. São Paulo: Edições Sesc, 2017.

SOUZA, Jessé. *A elite do atraso*. Rio de Janeiro: Leya, 2017.

SUE, Derald Wing. Multidimensional facets of cultural competence. *The counseling psychologist*, v. 29, n. 6, p. 790-821, 2001.

TYNES, Brendesha M. et al. From Racial Microaggressions to Hate Crimes: A Model of Online Racism Based on the Lived Experiences of Adolescents of Color. *Microaggression Theory: Influence and Implications*, p. 194-212, 2018.

WANG, JuHong; ZHANG, SongHai; MARTIN, Ralph R. New advances in visual computing for intelligent processing of visual media and augmented reality. *Science China Technological Sciences*, v. 58, n. 12, p. 2210-2211, 2015.

WILSON, Benjamin; HOFFMAN, Judy; MORGENSTERN, Jamie. Predictive inequity in object detection. *arXiv preprint arXiv:1902.11097*, 2019.

Recebido 30/11/2019

Aprovado em: 30/01/2020