



VISÃO COMPUTACIONAL E VIESES RACIALIZADOS: BRANQUITUDE COMO PADRÃO NO APRENDIZADO DE MÁQUINA

Tarcízio Roberto da Silva (UFABC)¹

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. Um dos campos mais inovadores é o da visão computacional, que permite reconhecer automaticamente objetos, entidades, conceitos ou características extraídas de imagens, inclusive sobre pessoas. A partir da exploração de como a visão maquínica dos provedores de inteligência artificial reconhecem objetos, conceitos e dimensões em redes semânticas, investigamos como o reconhecimento automatizado de imagens e posterior modulação algorítmica de comportamento parte frequentemente de um olhar eurocêntrico quanto a categorias raciais, sobretudo vinculado a inclusão/exclusão e visibilidade/invisibilidade do outro racializado, com impacto efetivo para os usuários de agentes artificiais, plataformas de mídias sociais e bancos de dados. Esta investigação parte da bibliografia de estudos sobre a branquitude e o seu papel enquanto definidora da sociedade e tecnologias de produção e controle. O trabalho apresenta criticamente casos mapeados por ativistas e desenvolvedoras em torno do mundo sobre problemas especificamente ligados à visão computacional. Tratam-se de casos de equivalência de conceitos como visibilidade, humanidade e beleza à experiência de usuários brancos em sistemas como robôs interagentes, buscadores e mídias sociais. Uma vez que 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, o estudo busca jogar luz sobre estas questões relevantes ao debate sobre algoritmos e suas.

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

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

¹ Tarcízio Silva é estudante de Doutorado em Ciências Humanas e Sociais da Universidade Federal do ABC, membro do NEAB e LabLivre da mesma universidade e co-fundador do Instituto Brasileiro de Pesquisa e Análise de Dados. Email: eu@tarciziosilva.com.br



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 ex: ‘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, Daniels (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.

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 se trata 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, Lozada, Smith & Stewart, 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.

Considerando que as manifestações do racismo são “construídas e expressas na infraestrutura online ou *back end* (por ex: algoritmos) ou através das interfaces (por ex: 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

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



produtivas nos contextos sócio-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, pos. 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 em nossos cotidianos em 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. 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-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 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. livre).

Suas manifestações 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 (SILVA, 2019). Trataremos, em seguida, como a visão computacional representa esta dupla opacidade irmanada com práticas da branquitude.

Visão Computacional: raça, identificação e representação

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 do usuário contemporâneo de tecnologias digitais. Os sistemas “sabem” 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.

Entretanto, a aplicação da visão computacional tem seus problemas quanto a vieses possíveis. A Tabela 1 mostra alguns dos casos mais famosos de problemas identificados por usuários ou jornalistas³ 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 Selecionados de Manifestação de Viés/Racismo Algorítmico em Visão Computacional

Caso	Problema/Viés Algorítmico	Causa/problema “técnico”
Google marca pessoas negras como gorilas	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
Faceapp embranquece pele para deixar “mais bonita” a selfie	Representação eurocêntrica de beleza; desumanização	<ul style="list-style-type: none">• Base de dados insuficiente• Ausência de testes

³ Mantemos uma linha do tempo com mais casos, em atualização contínua, no endereço <https://tarciziosilva.com.br/blog/posts/racismo-algoritmico-linha-do-tempo/>



APIs de análise de expressões faciais associam emoções negativas a negros	Percepção eurocêntrica; estereotipização	<ul style="list-style-type: none">• Base de dados insuficiente• Ausência de testes
APIs não reconhecem gênero e idade de mulheres negras	Representação eurocêntrica de gênero e idade	<ul style="list-style-type: none">• Base de dados insuficiente• Ausência de testes
Google Vision confunde cabelo negro com peruca	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	Desumanização; risco físico direto	<ul style="list-style-type: none">• Base de dados insuficiente• Ausência de testes

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 Twitter⁴, 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 recurso⁵.

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, identificamos 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. De modo geral, o estudo identificou que “arquiteturas algorítmicas favorecem a performance de neutralidade cultural, mas com categorias genéricas com muito mais destaque do que termos específicos” (MINTZ, SILVA et al,

⁴ <https://www.theverge.com/2015/7/1/8880363/google-apologizes-photos-app-tags-two-black-people-gorillas>

⁵ <https://www.wired.com/story/when-it-comes-to-gorillas-google-photos-remains-blind/>



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).

Como quarto caso ligado à identificação, podemos destacar os experimentos e projeto *GenderShades.org* de Buolamwini e colaboradoras. Em experimento, Buolamwini e Gebru (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 1 mostra dois 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.

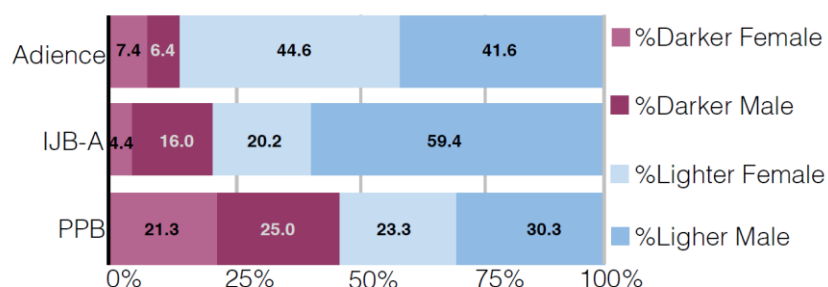


Figura 1: Comparação de datasets de treinamento (BUOLAMWINI & GEBRU, 2018)

Em 2019, as pesquisadoras retornaram aos dados 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 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 diminuem” (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 resultados⁶. 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 Unidos⁷.

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

⁶ IBM <http://gendershades.org/docs/ibm.pdf> e Microsoft: <http://gendershades.org/docs/msft.pdf>

⁷ <https://gizmodo.com/ibm-sells-face-recognition-surveillance-to-a-dictatorsh-1835101881>



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.

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 último 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” (2019, p.9, trad. livre).

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, 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 10x 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

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



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 máquinas (BROUSSARD, 2018; NOBLE, 2018).

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).



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, mostra 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 (ROBERTS, 2012), a favor da segmentação dos corpos quanto a genética, biomedicina e vigilância (MBEMBE, 2001, 2016; HARARI, 2015; 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” (BENTHAL & 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.



Bibliografia:

- ALI, Syed Mustafa. Race: The difference that makes a difference. **tripleC: Cognition, Communication, Cooperation**, 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.
- 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.
- EPSTEIN, Ziv et al. Closing the AI Knowledge Gap. **arXiv preprint arXiv:1803.07233**, 2018.
- GILLESPIE, Tarleton. The relevance of algorithms. **Media technologies: Essays on communication, materiality, and society**, v. 167, 2014.
- 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; TYNES, Brendesha M. **The intersectional internet: Race, sex, class, and culture online**. Peter Lang International Academic Publishers, 2016.

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.

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, 2019.

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.