



Racismo Algorítmico em Plataformas Digitais: microagressões e discriminação em código

TARCÍZIO SILVA¹

Resumo:

Ambientes como Facebook, Twitter, YouTube e *marketplaces* de aplicativos mobile são mecanismos centrais do capitalismo de vigilância. Os vieses algorítmicos e *affordances* racistas nestas plataformas digitais de publicidade estão sendo crescentemente analisados nos últimos 15 anos por pesquisadoras informadas por perspectivas variadas. Como resultado estão em desenvolvimento iniciativas de auditoria de algoritmos e plataformas, construção de *datasets* e procedimentos mais responsáveis, além de indicações regulatórias. A opacidade presente nas caixas-pretas dos grupos de sistemas automatizados e semi-automatizados baseados em algoritmos que regem visibilidade, classificação, vigilância e regras de uso nas plataformas digitais complexifica e dificulta esta investigação. Esta questão se aplica tanto ao desenho de processos internos quanto à configuração de algoritmos e é agravada pela ideologia da “cegueira racial”, tática que compõe historicamente do genocídio negro, que também é epistemológico. O trabalho propõe colaborar ao campo de estudos sobre vieses algorítmicos ao aproximar a investigação sobre vieses raciais nos algoritmos de plataformas digitais ao conceito de microagressões raciais (Pierce, 1970). Estas seriam definidas como mensagens rotineiras que comunicam insultos e desprezo racial e podem ser apresentadas de forma verbal, comportamental ou ambientalmente contra grupos racializados. A pesquisa sobre microagressões raciais propõe modos de analisar, entender e combater este tipo de violência em contextos de socialização, midiáticos ou educacionais com o objetivo de minimizar o impacto na formação e pleno desenvolvimento de populações negras, entre outras. Aplicando a tipologia de microagressões proposta por Tynes et al (2018) a partir de Sue (2007) a mapeamento realizado pelo autor, o artigo discute casos de comunicação algorítmica racista em plataformas digitais ligando aspectos das plataformas a categorias e subcategorias propostas sobre microagressões.

Palavras-chave:

Microagressões; racismo algorítmico; plataformas digitais; capitalismo de vigilância; algoritmos

Conflitos epistêmicos sobre o papel da internet na relação, intensificação ou erosão de grupos identitários e suas controvérsias estiveram presentes desde as primeiras discussões acadêmicas e

¹ Doutorando em Ciências Humanas e Sociais (UFABC), Mestre em Comunicação e Cultura Contemporâneas (UFBA) e cofundador do IBPAD. Email: eu@tarciziosilva.com.br



vernaculares sobre comunicação e cultura digital. A ideia de um *self* cambiante que poderia ser diferente a cada nova janela dos ambientes online ganhou popularidade sobretudo em um período que: a) os ambientes digitais eram ainda informacionalmente escassos, com poucas modalidades de comunicação, focando sobretudo em textualidade; b) não havia massa crítica de pesquisadores advindos de populações racializadas nos países de diáspora africana; c) a pretensão de neutralidade das plataformas e mídias, advindas de um tecnoliberalismo em consolidação, já se fazia vigente.

Entretanto, grupos de cientistas, teóricas e ativistas da comunicação e tecnologia apontaram os processos pelos quais a construção tanto das tecnologias digitais de comunicação quanto da ideologia do Vale do Silício são racializadas, a partir de uma lógica da supremacia branca (Nakamura, 2008; Daniels, 2013; Broussard, 2018).

Especificamente sobre manifestações explícitas de mensagens racistas nas mídias sociais, destacamos o trabalho recente de Trindade (2018), que traz dados especialmente relevantes sobre comportamento dos agressores, características das vítimas e posições quanto à tecnologia: 81% das vítimas de racismo no Facebook no Brasil são mulheres de classe média; 76,2% dos agressores não tinham nenhum relacionamento prévio com a vítima; e nutre-se uma crença de que o ambiente virtual se constitui num espaço totalmente desregulado (Trindade, 2018). Estudos sobre marcadores textuais específicos em torno de questões raciais apontam as controversas e tentativas de diluição do debate, como a identificação, em análise sobre o *#DiadaConsciênciaNegra* no Twitter em que a data foi "reduzida à questão do feriado, tornando irrelevante sua associação com o debate racial, com a questão histórica e com o desvelamento da dominação" (Recuero, 2014, p.307).

Defendemos que, apesar do frequente foco da pesquisa digital em ações – individuais ou coletivas – pontuais, o racismo online é um “system of anti-people of color practices that privilege and maintain political, cultural, and economic power for Whites in digital space” (Tynes, Lozada, Smith & Stewart, 2019, p.195). Isto significa dizer que o foco apenas em casos, práticas ou modalidades específicas e explícitas de discurso racista dão conta de apenas uma parte da questão. Nos ambientes digitais, entretanto, temos um desafio ainda mais profundo quanto à materialidade dos modos pelos quais o racismo se imbrica nas tecnologias digitais através de processos “invisíveis” nos recursos automatizados como recomendação de conteúdo, reconhecimento facial e processamento de imagens. É preciso entender também suas manifestações “built and expressed in the online infrastructure or back end (e.g., algorithms), or through the interface (e.g., symbol, image, voice, text, and graphic representations)” (Tynes et al, 2019 ,p.195).



Plataformização, algoritmos e raça online

O diagnóstico da “plataformização” da web tomou forma através da percepção de que, gradualmente, ambientes digitais como Facebook estenderam-se para toda a web em uma espécie de integração que priorizou a concentração de dados e valor – inclusive financeiro – em poucas empresas. A plataformização transformou *sites* de redes sociais em *plataformas* de mídias sociais, nos termos de Helmond, que adiciona que como “infrastructural model, social media platforms provide a technological framework for others to build on, which I argue, is geared towards their expansion into the rest of the web” (Helmond, 2015, p.3). Entre infraestrutura e economia, o impulso do capital financeiro permitiu que grupos como Alphabet (detentora do Google), Facebook, Amazon, Microsoft e Apple investissem no desenvolvimento estratégico de suas plataformas, incluindo a construção de conglomerados de cientistas de variadas áreas para contínua análise e otimização dos fluxos de monetização da audiência, publicidade e comportamento de seus usuários, realizando uma verdadeira redistribuição dos métodos e hierarquias de capacidade de análise e representação da sociedade (Marres, 2012).

Para além das plataformas de mídias sociais, a tendência se expandiu para infraestruturas industriais e tecnológicas de diversos segmentos. Srnicek (2017) aponta diversos tipos além das plataformas de publicidade (como Facebook): plataformas de nuvem e computação (Amazon Web Services, Microsoft Azure etc), plataformas de produto (como Zipcar etc), plataformas *lean* (Uber, AirBnB) e plataformas de gestão industrial. O autor propõe que as plataformas são “digital infrastructures that enable two or more groups to interact. They therefore position themselves as intermediaries that bring together different users” (Srnicek, 2017, pos.497) como audiências, publicitários, fornecedores de serviço, objetos físicos e empresas.

Temos, portanto, uma crescente dataficação (van Dijck, 2014) e mediação das atividades humanas, da interação interpessoal a negociação de serviços e comércio. Apesar de manter discursivamente ideais de liberdade e horizontalização das relações, a plataformização da comunicação e economia significa concentração internacional de fluxos de dados e capital. Ao mesmo tempo, está no cerne do bom resultado financeiro e mercadológico destas empresas o uso de algoritmos que promovem a eficiência na busca das métricas definidas como objetivos de negócio, que vão de indicadores de consumo de mídia a número de pontos de dados processados.



Apesar de não ser um conceito novo, remontando à história da matemática e computação, algoritmos poderiam ser definidos a princípio como “a finite sequence of precise instructions that are implementable on computing systems” (Osoba & Welser IV, 2017, p.5). Na era computacional digital, em sistemas de *big data*, a lógica algorítmica foi expandida para processos de inteligência artificial estreita, presente nos sistemas informacionais do cotidiano.

Os sistemas algorítmicos tomam decisões por nós com frequência cada vez maior. A “authority is increasingly expressed algorithmically. Decisions that used to be based on human reflection are now made automatically. Software encodes thousands of rules and instructions computed in a fraction of a second” (Pasquale, 2015, p.4). Estas decisões trazem impactos em diferentes níveis de imediaticidade e sutileza, podendo modular o comportamento e condutas de seus usuários de forma discreta (Silveira, 2017), na maioria dos casos para reproduzir relações de poder e opressão já existentes na sociedade. Este é um dos grandes desafios e problemas da lógica do aprendizado de máquina, que se baseiam no cálculo computacional de milhares de decisões “ótimas” a partir do *input* de dados:

algorithms tend to be vulnerable to characteristics of their training data. This is a feature of these algorithms: the ability to adapt in the face of changing input. But algorithmic adaptation in response input data also presents an attack vector for malicious users. This data diet vulnerability in learning algorithms is a recurring theme (Osoba & Welser, 2017, p.7)

Os casos de identificação de racismo algorítmico somam-se na medida em que pesquisadoras, ativistas e desenvolvedoras geram relatórios, reportagens e guias de auditoria e ação sobre aspectos discriminatórios em diversos dispositivos midiáticos como análise de recomendação de conteúdo (Tufekci, 2015), anúncios (Sweeney, 2013), reconhecimento facial e visão computacional (Buolamwini 2017; Buolamwini & Gebru, 2018), buscadores (Noble, 2018) e outros².

Junto a outros indicativos sobre economia, violência, (necro)política e representação midiática, estes casos lembram que racismo “should not be understood as the exceptional behaviour of individuals deviating from a non-racist social norm, but rather, as a global socio-political system” (Ali, 2013, p.99) que inclui historicamente formatações dos campos produtivos da tecnologia que favorecem o treinamento enviesado de sistemas que intensificam discriminações e opressões. Os algoritmos “designed by people, and people embed their unconscious biases in algorithms. It’s rarely

² O autor mantém mapeamento de casos em seu log, acessível em <https://tarciziosilva.com.br/blog/posts/racismo-algoritmico-linha-do-tempo/>



intentional—but this doesn't mean we should let data scientists off the hook. It means we should be critical about and vigilant for the things we know can go wrong" (Broussard, 2018, pos.2891).

Levando em conta a necessidade de estar vigilante sobre as opressões algorítmicas sutis, propomos então um desdobramento dos estudos sobre "microagressões" digitais para o estudo do racismo algorítmico.

Microagressões

Como vimos anteriormente, os exemplos na literatura sobre racismo online tratam sobretudo do racismo explícito no formato discursivo, através de textos ou materiais multimídia. Entretanto, esta abordagem não é suficiente para dar conta dos modos pelos quais indexação, busca e descoberta de mensagens online - sobretudo com a crescente automatização - podem ter impacto na vida e saúde mental de grupos minorizados. Acreditamos que o uso do conceito de microagressões e a consequente taxonomia tentativa de casos documentados pode trazer benefícios à pesquisa sobre racismo algorítmico, através de alguns pontos: realizar a própria construção de taxonomia como exercício reflexivo; servir como ferramenta mediadora de marcação das trajetórias dos debates e casos; e permitir desenvolver categorias tentativas para articulação com o campo.

O conceito de *microagressão* foi proposto pelo psiquiatra Chester Pierce (1969; 1970) ao desenvolver trabalho sobre a necessidade de estudar também os "mecanismos ofensivos" dos grupos opressores em medida similar ao que as práticas psiquiátricas já realizavam sobre os "mecanismos defensivos". Pierce explica que aparatos da educação e mídia criam um ambiente onde "most offensive actions are not gross and crippling. They are subtle and stunning. The enormity of the complications they cause can be appreciated only when one considers that these subtle blows are delivered incessantly" (1970, pp.265-266). As situações cotidianas de microagressões racistas apresentadas no trabalho seminal de Pierce são acrescidas de outras na bibliografia de psicologia social e análise de mídia documentadas através de experimentos e posteriormente organizadas em tipologias.

As microagressões raciais são "ofensas verbais, comportamentais e ambientais comuns, sejam intencionais ou não intencionais, que comunicam desrespeito e insultos hostis, depreciativos ou negativos contra pessoas de cor" (Sue, 2010a, p. 29), aplicados consciente e inconscientemente



como uma “form of systemic, everyday racism used to keep those at the racial margins in their place” (Huber & Solorzano, 2014, p.6).

O trabalho de Sue (2007; 2010a; 2010b) é um dos mais prolíficos sobre microagressões, abrangendo suas manifestações quanto à discriminação de raça, gênero e orientação sexual. Ao longo de sua produção, inicialmente o autor propôs uma taxonomia reunindo microagressões raciais que poderiam ser verbais, não-verbais ou ambientais. Os seus três tipos são respectivamente Microinsultos, Microinvalidações e Microataques. Microinsultos seriam “mensagens que conotam rudeza e insensibilidade e aviltam a herança racial de um indivíduo” (SUE, 2010a, p.29); Microinvalidações são “mensagens que excluem, negam ou nulificam as reflexões psicológicas, sentimentos ou realidades experienciais” (SUE, 2010a, p.29); e, por fim, Microataques seriam “mensagens derogatórias explícitas caracterizadas por um ataque violento verbal, não-verbal ou violento com intenção de machucar a vítima através de xingamentos, comportamento de evitação ou ações discriminatórias propositais” (SUE, 2010a, p.29).

O uso do termo “micro” não se refere necessariamente ao grau de virulência, mas antes a pervasividade e o fato que “a agressão incide em um nível individual e/ou local, ou mesmo em situações “privadas” ou limitadas, que permitem certo grau de anonimato por parte do agressor” (Silva & Powell, 2016, p.46) ou ainda permitem evasão, através de disputa sobre intencionalidade ou modalidade (“era só uma piada”).

Entretanto, propostas como a de Levchak (2018) buscam criar uma distinção mais explícita entre macro e microagressões, como podemos ver na Figura 01. A diferença entre o caráter explícito e implícito do racismo e entre a intencionalidade são úteis para delimitar a tipologia, uma vez que diferentes ações nos campos educacionais ou comunicacionais são possibilitadas ou restringidas de acordo com cada tipo.

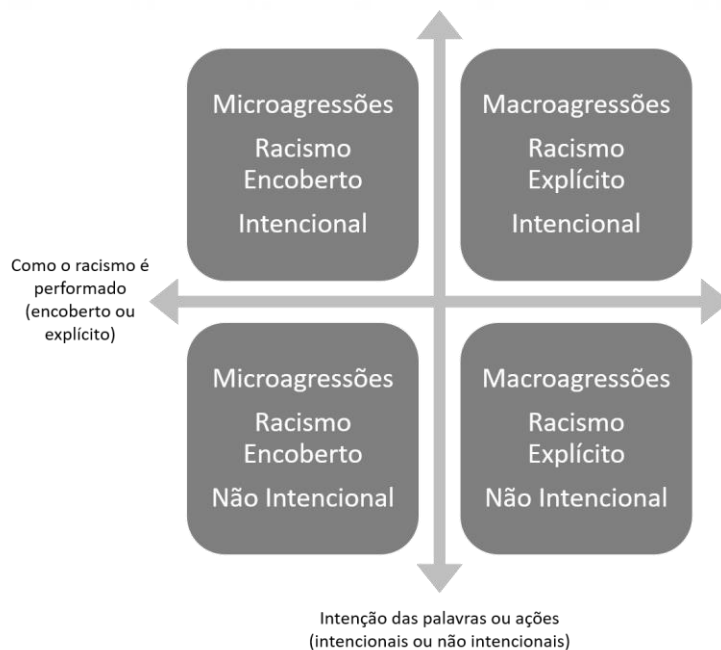


Figura 1: Continuum entre microagressões e macroagressões (Levchak, 2018; tradução nossa)

Pensando especificamente no racismo online, Tynes e colaboradores (2019) incluem outras duas categorias especialmente úteis para o estudo também das manifestações automatizadas: *Deseducação* e *Desinformação*. Quanto a primeira, define deseducação como “creation of online learning materials that in most cases unintentionally demeans or omits People of Color” (Tynes et al, 2019, p. 201). Apesar das autoras tratarem também de materiais pedagógicos nos espaços formais de educação, listam ambientes e produtos como websites, livros digitais e blogs.

Já desinformação pode ser de dois tipos – deliberada ou não-deliberada. Esta última, não-deliberada, se refere ao ato de desinformar intencional ou não-intencionalmente devido ao fato do emissor da mensagem estar também desinformado, baseado em informação incorreta ou errônea. Já desinformação deliberada se trata especificamente de gerar ou repassar informação falsa com a intenção de gerar mais desinformação nos receptores (Tynes et al, 2019). Entretanto, como a intencionalidade dos atores é difícil de ser comprovada – ou, mais, é distribuída em plataformas e algoritmos – pesquisadores do campo costumam focar apenas desinformação de modo amplo³.

³ É essencial notar que na bibliografia anglófila citada são usadas as palavras *misinformation* e *disinformation*, sendo esta última a intencional. Apesar dos termos estarem em voga na pesquisa, devido às investigações sobre *fake news* e outros tipos de desinformação, ainda não há um consenso em português para os termos.



Diversos estudos empíricos levantaram relatos de microagressões nos contextos educacionais, interpessoais e laborais, descobrindo alguns padrões das microagressões racistas nos contextos americanos e brasileiros, com algumas particularidades nos casos de racismo anti-negro, anti-asiático ou anti-indígena (Sue, Capodilupo et al, 2007; Sue, Bucceri et al, 2007; Nadal et al 2011; Nadal 2014; Silva & Powell, 2016; Levchak, 2018; Tynes, 2019). Para materializar, seguiremos na descrição de alguns tipos comuns identificados na literatura acadêmica:

Suposição de Criminalidade. No caso das populações brasileiras, este é uma das microagressões mais pervasivas. Trata da suposição que uma pessoa racializada tem mais chance de ser “dangerous, criminal, or deviant on the basis of their race” (Sue, Capodilupo et al, 2007) e suas manifestações presenciais são bem conhecidas pela população negra. Por exemplo, excessiva vigilância em um estabelecimento comercial é uma queixa frequente, mas devido a fácil negação da intencionalidade só gera repercussão aos perpetradores em casos críticos.

Negação de Realidades Raciais / Democracia Racial. Outro tipo de microagressão comum é a negação de realidades raciais ou, no caso brasileira, a defesa à equivocada – talvez estratégica, do ponto de vista da branquitude – ideia de “democracia racial” que influenciou a sociologia e discurso popular brasileiro no último século e se alastra em outros países da diáspora africana como EUA (Bonilla-Silva, 2006). Essa ideologia é usada para promover tanto a negação de atitudes racistas – pois o perpetrador “não veria cor”, quanto para deslegitimar produção de conhecimento – seja científico ou vernacular – por pensadores, pesquisadores e ativistas negros. Ao discutir genocídio epistemológico, Nascimento apontou que o negro brasileiro, embora “seja discriminado exatamente por causa de sua condição racial e da cor, negam a ele, com fundamentos na lei, o direito legal da autodefesa” (Nascimento, 2016 [1977], pos.1546).

Suposição de Inferioridade Intelectual. Este tipo de microagressão é especialmente relevante no contexto educacional, onde a distribuição de oportunidades, reprimendas ou atenção por agentes educacionais não é igualitária. Discursivamente, supor que um indivíduo possui determinada “posição na universidade ou no trabalho por meio de uma política afirmativa” (Guimarães & Silva, 2016, p.51), “surpresa” com a articulação ou ideias de grupos racializados ou, ainda, forçar estereótipos disciplinares, como no caso de descendentes de asiáticos.

Patologização de Valores Culturais. Na história dos estudos da branquitude sobre culturas e religiões africanas e afro-brasileiras esta distorção foi comum. Nascimento revisa as leituras patologizantes do candomblé na ciência social brasileira, percebendo que “as concepções metafísicas da África, seus



sistemas filosóficos, a estrutura de seus rituais e liturgias religiosos, nunca merecem o devido respeito e consideração como valores constitutivos da identidade do espírito nacional" (2017[1977], pos. 1536) – chegava-se ao cúmulo de usar referencial psiquiátrico para tratar das manifestações da religião afro-brasileira. No cotidiano, microagressões de fundo patologizante são frequentes na percepção enviesada de julgamentos sobre “agressividade”, “timidez” e outras características como se fossem próprias do grupo ao invés de contextual.

Exotização. A patologização irmana com a exotização das populações racializadas, sobretudo nas opressões interseccionais. A exotização de mulheres racializadas soma-se à misoginia e leva a mensagens e associações à hipersexualização.

Estrangeiro na Própria Terra / Negação de Cidadania. No caso de grupos minorizados quantitativamente, são enquadrados como se não fossem efetivamente do local – como acontece com os povos indígenas em países da América. Essa negação de cidadania é frequente também em ambientes de consumo, onde o grupo hegemônico recebe tratamento preferencial.

Exclusão ou Isolamento. Por fim, podemos citar comportamentos que geram exclusão ou isolamento dos indivíduos racializados, de modo a não se sentirem pertencentes de um determinado grupo nas relações interpessoais, educacionais ou laborais.

Na Figura 2 podemos ver a Taxonomia do Racismo Online proposta por Tynes e colaboradores (2019), que nos permite posicionar e distinguir as práticas do racismo online. Defendemos que as manifestações algorítmicas de racismo são microagressões frequentes de diversos tipos, que podem afetar os usuários de plataformas de forma individual ou vicária.

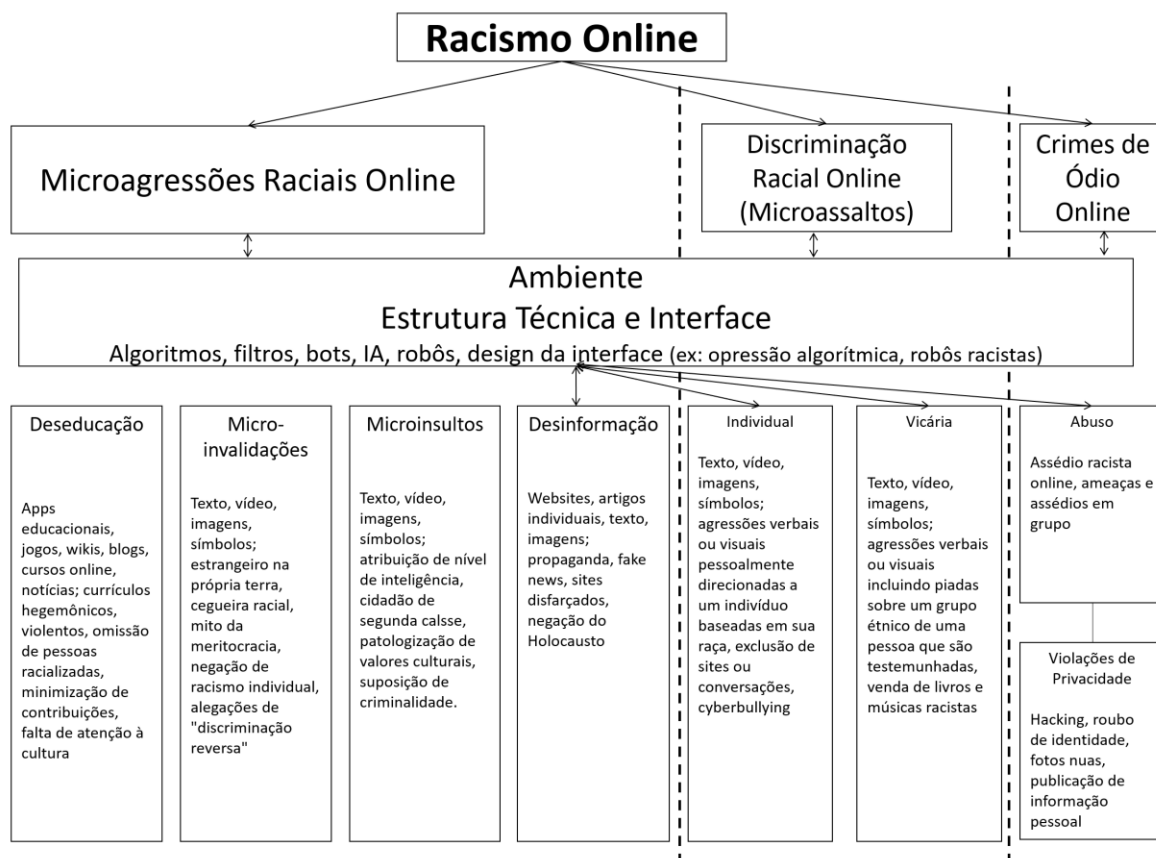


Figura 2: Taxonomia do Racismo Online (Tynes et al, 2019; tradução nossa)

A bibliografia registra outros tipos comuns de microagressões raciais e relacionadas a outros grupos minorizados quanto a gênero, sexualidade, neuroatipicidade e outras variáveis. Para os fins deste trabalho, a descrição dos tipos anteriores nos é útil para seguir na citação de casos de racismo algorítmico observados nos últimos anos por pesquisadoras, desenvolvedoras, ativistas e jornalistas.

Opressões algorítmicas como microagressões

Acreditamos, portanto, que podemos ver boa parte das discriminações raciais algorítmicas como microagressões. Lau e Williams (2010) revisaram aspectos metodológicos da literatura sobre microagressões e identificaram que as formas operacionais de mensuração de microagressões tomaram a forma sobretudo de investigação qualitativa, através de: relatos subjetivos, observações, medições de traços e registros de arquivos.



Os registros de arquivo “may be more reliable sources of information if accuracy of subjective report is a concern for future researchers” (Lau e Williams, 2010, p. 321) mas são menos usados na pesquisa em microagressões, segundo os autores. Entretanto, apesar de não abarcarem a temática dos algoritmos, falam do potencial da tecnologia digital como sites de redes sociais por permitirem uma oportunidade única de “take advantage of research on electronic archival data could uncover interesting findings not as easily revealed in interviews or more structured methods” (Lau e Williams, 2010, p.322).

Podemos enumerar uma dúzia de casos que poderiam ser abordados pela perspectiva das microagressões digitais, a partir de mapeamento em produção. A Tabela 1 cita de forma concisa 12 casos documentados em sistemas de publicação de anúncios, buscadores de imagens, processamento de linguagem natural, visão computacional e *bots* conversacionais.

Tabela 1: Lista de Casos de Racismo Algorítmico mapeados pelo autor

Caso de Racismo Algorítmico	Microagressões	Categoria
Sistema do Google permite empresas exibirem anúncios sobre crime especificamente a afro-americanos (Sweeney, 2013)	Suposição de Criminalidade	Microinsultos
Resultados no Google Imagens apresentam hiper-sexualização para buscas como “ <i>garotas negras</i> ” (Noble, 2013, 2018)	Exotização; Negação de Cidadania	Microinsultos
Facebook esconde manifestações contra violência policial racista (Tufekci, 2014)	Negação de Realidades Raciais	Microinvalidações
Google Photos marca fotos de jovens negros com a tag “Gorila” ⁴	Negação de Cidadania	Microinsultos
Chatbot da Microsoft torna-se racista em menos de um dia ⁵	Diversas	Microinsultos
Robôs conversacionais de startups não encontram face de mulher negra; sistemas de visão computacional erram gênero e idade de mulheres negras (Buolamwini, 2018)	Negação de Cidadania; Exclusão e Isolamento	Microinvalidações
Mecanismos de busca de bancos de imagens invisibilizam famílias	Negação de	Microinvalidações;

⁴ <https://www.theguardian.com/technology/2015/jul/01/google-sorry-racist-auto-tag-photo-app>

⁵ <https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist>



e pessoas negras ⁶	Realidades Raciais	Desinformação
App que transforma selfies equipara beleza à brancura ⁷	Exotização; Exclusão e Isolamento	Microinsultos; Microinvalidações
APIs de visão computacional confundem cabelo negro com perucas (Mintz, Silva et al, 2019)	Exotização	Microinvalidações; Deseducação
Ferramentas de processamento de linguagem natural possuem vieses contra linguagem e temas negros ⁸	Patologização de Valores Culturais	Deseducação
Análise facial de emoções associa categorias negativas a atletas negros ⁹	Suposição de Criminalidade	Microinsultos
Twitter decide não banir discurso de ódio nazista/supremacista branco para não afetar políticos republicanos	Negação de Realidades Raciais; Exclusão	Deseducação; Desinformação

A Tabela 1, portanto, representa um modo de pensar os casos de racismo algorítmico ligados ao conceito de microagressões e seus impactos, intensificados ou transformados pelas características das plataformas digitais.

Podemos citar como relevante a característica de editorialização dos algoritmos e interfaces das plataformas como modo de gerar ou moldar informação e desinformação. Em 2014, parte da população estadunidense, sobretudo a afro-americana, engajou-se em séries de protestos contra a violência policial direcionada a jovens negros, na maioria das vezes com uso excessivo de força ou, ainda, contra totais inocentes (Freelon, McIlwain & Clark, 2016). No período, o Facebook experimentou um recurso chamado “Top Trends” que, assim como o similar recurso no Twitter (“Trending Topics”) exibia termos-chave sobre eventos em discussão massiva. Apesar dos protestos em Ferguson estarem em destaque na imprensa e nas mídias sociais, não foi exibido entre os “Top Trends” do Facebook. Agindo através de “computational agency, Facebook's algorithm had "decided" that such stories did not meet its criteria for "relevance"- an opaque, proprietary formula

⁶ <http://desabafosocial.com.br/blog/2017/06/12/desabafo-social-interfere-no-mecanismo-de-busca-do-maior-banco-de-imagem-do-mundo/>

⁷ <https://www.theguardian.com/technology/2017/apr/25/faceapp-apologises-for-racist-filter-which-lightens-users-skintone>

⁸ <https://peopleofcolorintech.com/articles/how-automated-tools-discriminate-against-black-language/>

⁹ <https://theconversation.com/emotion-reading-tech-fails-the-racial-bias-test-108404>



that changes every week, and which can cause huge shifts in news traffic” (Tufekci, 2014, p.3), influenciando a manutenção de determinados debates na esfera pública expandida. Então podemos falar de uma opacidade algorítmica que, por sua vez, decide visibilidade e invisibilidade de temas, levando ativistas e pesquisadores a uma dificuldade maior de identificar causas e efeitos. O grupo minorizado em questão – afro-americanos – se depara com esta incongruência na cobertura e citação dos temas que os afetam, uma vez que a maioria dos usuários de mídias sociais é plataforma ao mesmo tempo que “uma das principais características das microagressões, em todos seus níveis, é o fato de sua aparente “invisibilidade”, principalmente contidas em expressões não intencionais de preconceito” (Silva & Powell, 2016, p.54). Afinal de contas, como o assassinato impune de jovens e manifestações decorrentes poderiam não ser “relevantes” no principal ambiente midiático da internet?

Também elusivo é o modo pelo qual buscadores selecionam resultados para demandas de conteúdo. Ao longo de diversos trabalhos que culminaram em importante livro chamado *Algorithms of Oppression*, Safiya Noble (2018) estudou como plataformas como Google reproduzem representações nocivas de grupos minorizados, como a hiperssexualização de garotas negras e latinas em resultados de busca de imagens. Os argumentos tecnoliberais de que os resultados apenas refletiriam a produção efetivamente disponibilizada na internet por sites pornográficos e de mídia tem sido combatidos. Noble aponta que “on the Internet and in our everyday uses of technology, discrimination is also embedded in computer code and, increasingly, in artificial intelligence technologies that we are reliant on, by choice or not” (Noble, 2018, pos. 15).

Buscadores de informação, websites e imagens são uma tecnologia essencial para o uso contemporâneo da internet por pessoas comuns e profissionais. Em grande medida, indicadores mostram que a maior parte das pessoas não navega por muitas páginas, focando nos primeiros resultados. Portanto, a ordem dos resultados – definida algorítmicamente – tem papel relevante na reprodução de representações e acesso a informações consoantes ou dissonantes de olhares hegemônicos ou contra-hegemônicos. O trabalho supracitado de Noble (2013, 2018) argumenta sobre perigos da hipervisibilidade negativa e hiper-sexualizada enquanto outros trabalhos (Aiello, 2016; Mintz, Silva et al, 2019) tratam também da invisibilidade. É o caso do projeto “Vamos conversar, bancos de imagens?” do coletivo *Desabafo Social*. Através de vídeos mostrando o procedimento de buscas em bancos de imagens como Shutterstock, Getty Images, iStock e DepositPhotos, o coletivo exhibe como o resultado para termos simples como “família” ou “bebês”



mostra praticamente apenas pessoas brancas. No caso dos bancos de imagens, o seu consumo é feito por milhares de produtores de conteúdo, o que pode gerar um efeito em cascata: publicitários, blogueiros e jornalistas sem recursos para produção própria de imagens tenderão a usar imagens não-representativas da diversidade brasileira, piorando os índices de modo geral. Pierce e colaboradores apontaram a questão das microagressões em análise quantitativa de categorias de representação em mídia, concluindo que os anúncios televisivos se tornam uma “plethora of sources which spew out microaggressions through offensive mechanisms” (Pierce et al, 1977).

Se os casos anteriores são exemplares de microagressões algorítmicas análogas a casos de in/visibilidade e representação problemáticas mapeados nos estudos de mídia, as capacidades de transformação algorítmica de material textual e visual trazem uma faceta nova ao problema. Aplicativos como *FaceApp* tem sido denunciados por recursos que transformam *selfies* e fotos de pessoas em mais “jovens”, “velhas”, “troca de gênero” ou “mais belas” – nesta última categoria foi percebido que características fenotípicas eurocêtricas foram valorizadas e aplicadas indiscriminadamente, chegando ao paroxismo de clarear fotos de usuários de ascendência africana e indiana. Ao ser questionado pelo *The Guardian*, o criador do app alegou que “It is an unfortunate side-effect of the underlying neural network caused by the training set bias, not intended behaviour” ignorando seu papel nas escolhas dos dados de treinamento e design do algoritmo, já amplamente documentados em guias de boas práticas (WEF, 2018).

É preciso sublinhar que tratamos aqui de casos ligados à algoritmização de plataformas de comunicação como mídias sociais, fornecedores de visão computacional, processamento de linguagem natural e aplicativos *mobile*, onde aspectos do capitalismo de vigilância se engendram em meios onde até recentemente os principais discursos festejavam a descentralização das hierarquias comunicacionais. Estas práticas se aproximam cada vez mais da vigilância estatal e policial desenvolvida a partir de tecnologias de controle de corpos negros (Browne, 2015) em manifestações problemáticas como policiamento “preditivo” (Osoba & Welser IV, 2017) e análise de reincidência criminal (Angwin et al, 2016).

Portanto, tratando-se da pervasividade de dispositivos tecnológicos de comunicação com recursos e funcionalidades baseadas em inteligência artificial e aprendizado de máquina, acreditamos que a perspectiva de estudos sobre microagressões pode ser útil para o campo dos estudos de comunicação e mídia digital. De forma programática, Levchak (2018) desenvolve reflexão sobre microagressões e racismo moderno levando em conta estratégias de enfrentamento possíveis no



que chama dos “3 Ps” - Preparação, Processamento e Proteção. Para a pesquisadora, os benefícios de se discutir racismo em diferentes grupos - tanto os hegemônicos quanto os minorizados - envolve a promoção de conversações construtivas em prol de melhores relações étnico-raciais.

Conclusões e desdobramentos

Vemos, à guisa de conclusão e indicação de trabalhos futuros, a importância da discussão e tipologia das microagressões aos estudos de vieses, discriminações e racismo algorítmico, ao nos apontar que as auditorias algorítmicas podem também “measure experiences directed at the individual as well as those that are vicarious or witnessed by a respondent. Like individual experiences, those witnessed may also impact psychological outcomes” (Tynes, Rose & Markoe, 2013). A maioria dos casos citados aqui tratam de iniciativas que habilmente exploraram ou casos críticos ou as materialidades, interfaces e APIs dos sistemas para interrogá-los. Mas quais as vozes e percepções dos usuários no dia a dia? Quais instrumentos críticos possuem em seu arsenal para observar e defender-se destes problemas?

Aproximar o referencial metodológico das microagressões e da *Teoria Racial Crítica*, usada por maior parte dos autores citados parece um programa essencial. Ainda mais que no caso do racismo explícito, “protections against microaggressions or covert racism are even more elusive, since it is often difficult to detect and address microaggressions due to their intricate and insidious nature” (Levchak, 2018, p.15). Então acreditamos que a educação sobre microagressões e suas manifestações digitalizadas e automatizadas é importante, uma vez que “the more people are aware of the term and concept, the less likely they will be defensive when confronted about their behaviors” (Nadal, 2014, p.74) – inclusive desenvolvedores e gerentes de produto, de *startups* à grandes plataformas. Promover competência cultural sobre as diferenças relevantes em um mundo racializado é essencial.

Do ponto de vista de grupos minorizados, as microagressões e suas taxonomias podem ajudar populações racializadas “to ‘name their pain,’ as Paulo Freire describes, to engage anti-oppressive strategies for liberation” (Huber & Solorzano, 2014, p.18), com especial atenção aos desenvolvedores e ativistas em ambientes comumente hostis. Daniels, Nkonde e Mir (2019) apontam três fundamentos essenciais para a literacia midiática sobre a relação entre raça e tecnologia: compreensão intelectual de como o racismo opera em algoritmos, plataformas e afins;



inteligência emocional para agir em situações racialmente problemáticas em organizações; e compromisso de tomada de ações para deduzir o dano a comunidades. Acreditamos então que uma perspectiva tanto interdisciplinar quanto interinstitucional de atuação e reflexão sobre os algoritmos é urgente para as sociedades contemporâneas.

Referências Bibliográficas

- Aiello, G. (2016). Taking stock. *Ethnography Matters*. Online, disponível em <http://eprints.whiterose.ac.uk/103476/>
- Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine bias. *ProPublica*, May, 23..
- Bonilla-Silva, E. (2006). *Racism without racists: Color-blind racism and the persistence of racial inequality in the United States*. Rowman & Littlefield Publishers.
- Broussard, M. (2018). *Artificial unintelligence: How computers misunderstand the world*. MIT Press.
- Browne, S. (2015). *Dark matters: On the surveillance of blackness*. Duke University Press.
- Buolamwini, J., & Gebru, T. (2018, January). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on Fairness, Accountability and Transparency* (pp. 77-91).
- Daniels, J. (2009). *Cyber racism: White supremacy online and the new attack on civil rights*. Rowman & Littlefield Publishers.
- Freelon, D., McIlwain, C. D., & Clark, M. (2016). Beyond the hashtags: #Ferguson, #Blacklivesmatter, and the online struggle for offline justice. *Center for Media & Social Impact, American University, Forthcoming*.
- Huber, L. P., & Solorzano, D. G. (2015). Racial microaggressions as a tool for critical race research. *Race Ethnicity and Education*, 18(3), 297-320.
- Lau, M. Y., & Williams, C. D. (2010). *Microaggression Research: Methodological Review and Recommendations*. Sue, D. W. (ed.), *Microaggressions and marginality: Manifestation, dynamics, and impact*. New Jersey (EUA): John Wiley & Sons, 313-336.
- Levchak, C. C. (2018). *Microaggressions and Modern Racism: Endurance and Evolution*. New Britain (USA): Springer.
- Marres, N. (2012). The redistribution of methods: on intervention in digital social research, broadly conceived. *The sociological review*, 60, 139-165.
- Mintz A; Silva, T.; Gobbo, B.; Pilipets, E.; Azhar, H.; Takamitsu, H.; Omena, J.J.; Oliveira, T. (2019) *Interrogating Vision APIs*. Relatório do Smart Data Sprint disponível em < <https://smart.inovamedialab.org/smart-2019/project-reports/interrogating-vision-apis/> >
- Nascimento, A. (2016). *O genocídio do negro brasileiro: processo de um racismo mascarado*. Editora Perspectiva SA.
- Nadal, K. L., Wong, Y., Griffin, K., Sriken, J., Vargas, V., Wideman, M., & Kolawole, A. (2011). Microaggressions and the multiracial experience. *International Journal of Humanities and Social Sciences*, 1(7), 36-44.
- Nadal, K. L. (2014). A guide to responding to microaggressions. In *Cuny Forum* (Vol. 2, No. 1, pp. 71-6).
- Nakamura, L. (2008). *Digitizing race: Visual cultures of the Internet* (Vol. 23). U of Minnesota Press.



- Noble, S. U. (2013). Google search: Hyper-visibility as a means of rendering black women and girls invisible. *InVisible Culture*, (19).
- Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. Nova Iorque (USA): NYU Press.
- Osoba, O. A., & Welsler IV, W. (2017). *An intelligence in our image: The risks of bias and errors in artificial intelligence*. Rand Corporation.
- Pasquale, F. (2015). *The black box society*. Harvard University Press.
- Pierce, C. M. (1969). *Is bigotry the basis of the medical problems of the ghetto?*.
- Pierce, C. (1970). Offensive mechanisms. In: BARBOUR, Floyd. (org.) *The black seventies*. Porter Sargent Pub, p. 265-282.
- Pierce, C. M., Carew, J. V., Pierce-Gonzalez, D., & Wills, D. (1977). An experiment in racism: TV commercials. *Education and Urban Society*, 10(1), 61-87.
- Recuero, R. (2014). Discutindo análise de conteúdo como método: o #DiadaConsciênciaNegra no Twitter. *Cadernos de Estudos Lingüísticos*, 56(2), 289-309.
- Sandvig, C., Hamilton, K., Karahalios, K., & Langbort, C. (2014). Auditing algorithms: Research methods for detecting discrimination on internet platforms. *Data and discrimination: converting critical concerns into productive inquiry*, 22.
- Silva, G. H. G., & Powell, A. B.. (2016). Microagressões no ensino superior nas vias da Educação Matemática. *Revista Latinoamericana de Etnomatemática*, 9(3).
- Silveira, Sergio A. (2017). *Tudo sobre tod@s: Redes digitais, privacidade e venda de dados pessoais*. Edições Sesc.
- Sobande, F., Fearfull, A., & Brownlie, D. (2019). Resisting media marginalisation: Black women's digital content and collectivity. *Consumption Markets & Culture*, 1-16.
- Srnicek, N. (2017). *Platform capitalism*. John Wiley & Sons.
- Sue, D. W., Capodilupo, C. M., Torino, G. C., Bucceri, J. M., Holder, A., Nadal, K. L., & Esquilin, M. (2007). Racial microaggressions in everyday life: implications for clinical practice. *American psychologist*, 62(4), 271.
- Sue, D. W., Bucceri, J., Lin, A. I., Nadal, K. L., & Torino, G. C. (2007). Racial microaggressions and the Asian American experience. *Cultural Diversity and Ethnic Minority*, Vol. 13, No. 1, 72– 81.
- Sue, D. W. (2010a). *Microaggressions in everyday life: Race, gender, and sexual orientation*. John Wiley & Sons.
- Sue, D. W. (ed.) (2010b) *Microaggressions and marginality: Manifestation, dynamics, and impact*. New Jersey (EUA): John Wiley & Sons.
- Sweeney, L. (2013). Discrimination in online ad delivery. *arXiv preprint arXiv:1301.6822*.
- Trindade, L. V. P. (2018). *It is not that funny. Critical analysis of racial ideologies embedded in racialized humour discourses on social media in Brazil* (Doctoral dissertation, University of Southampton), UK.
- Tynes, B. M., Rose, C. A., & Markoe, S. L. (2013). Extending campus life to the Internet: Social media, discrimination, and perceptions of racial climate. *Journal of Diversity in Higher Education*, 6(2), 102.
- Tynes, B. M., Lozada, F. T., Smith, N. A., & Stewart, A. M. (2018). 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*, 194-212.
- Tufekci, Z. (2015). Algorithmic harms beyond Facebook and Google: Emergent challenges of computational agency. *Colo. Tech. LJ*, 13, 203.

VI SIMPÓSIO INTERNACIONAL LAVITS >>>> SALVADOR
ASSIMETRIAS E (IN)VISIBILIDADES: VIGILÂNCIA, GÊNERO E RAÇA



Van Dijck, J. (2014). Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance & Society*, 12(2), 197-208.

WEF – World Economic Forum (2018). *How to Prevent Discriminatory Outcomes in Machine Learning*. Guia de boas práticas disponível em <<https://www.weforum.org/whitepapers/how-to-prevent-discriminatory-outcomes-in-machine-learning> >