

A política das máquinas: vieses em algoritmos de relevância pública

RESUMO

As práticas da vida cotidiana estão sendo cada vez mais reproduzidas sob a mediação de plataformas digitais (JURNO e DALBEN, 2018). Neste contexto, Algoritmos de Relevância Pública (ARP) determinam qual informação é relevante para os usuários, função de extrema importância para a sua participação nos grupos sociais (GILLESPIE, 2018a). Deste modo, problemas que já ocorrem nas sociedades contemporâneas, como preconceito racial, de gênero, classe social e outros enraizados na sociedade podem repercutir neste meio. O presente estudo busca refletir sobre como os ARP podem conter vieses e reproduzir preconceitos, observando alguns dos fatores que possam contribuir para este processo, através de uma revisão crítica da literatura. Observa-se, através dos resultados das pesquisas analisadas, que a origem destes problemas perpassa a tecnologia dos algoritmos, revelando que ainda há muito a ser feito para que se alcance o fim da desigualdade social reproduzida nas plataformas digitais.

PALAVRAS-CHAVE: Algoritmos. Plataformas digitais. Viés algorítmico. Preconceito.

Luis Gustavo Bier Schuler

lg@feevale.br

Universidade Feevale – Novo Hamburgo,
Rio Grande do Sul, Brasil

Sandra Portella Montardo

sandramontardo@feevale.br

Universidade Feevale – Novo Hamburgo,
Rio Grande do Sul, Brasil

INTRODUÇÃO

Algoritmos são lógica, passos bem definidos para resolver um problema especificado de forma clara (CORMEN et. al., 2002, p.3). Não há computação sem algoritmos, sendo a programação parte essencial dos computadores da forma que são conhecidos na contemporaneidade, seja em formato móvel ou não. Deste modo, estão inseridos na vida cotidiana dos indivíduos dos mais diversos grupos sociais através de plataformas digitais que já fazem parte da rotina diária de milhões de pessoas¹, fazendo, inclusive, com que atividades antes não realizadas no ambiente digital passem a ser inseridas nessa realidade, pode-se dizer, de maneira orgânica.

As plataformas digitais interferem na vida social de diversas formas — sendo elas partes de um ecossistema que possui suas próprias regras e atua através de seus próprios mecanismos, tornaram os seus usuários dependentes de sua infraestrutura para realizar atividades como comunicação e interação online, por exemplo (VAN DIJCK, POELL e DE WAAL, 2018). Segundo o relatório Digital 2019 (KEMP, 2019), os usuários ativos de plataformas de mídias sociais já somam 45% da população mundial, 9% a mais do que no ano anterior, 2018 (KEMP, 2019). Com práticas da vida cotidiana sendo reproduzidas no meio digital e mediadas por plataformas em controle dos detentores dos códigos-fonte e únicos reguladores dos algoritmos que as definem (JURNO e DALBEN, 2018), os problemas sociais de preconceito racial, gênero, classe social, xenofobia e outros enraizados na sociedade podem repercutir neste meio, tanto a partir das práticas dos seus usuários, quanto a partir das práticas as quais as próprias plataformas, através dos seus algoritmos, foram programadas para realizarem.

Este estudo tem o enfoque nos Algoritmos de Relevância Pública - ARP (GILLESPIE, 2018a), que são aqueles utilizados pelos mecanismos de busca, plataformas de mídia social e sistemas de recomendação, e que, segundo Gillespie (2018a) “exercem uma função crescentemente importante em selecionar qual informação deve ser considerada mais relevante para nós, uma característica crucial da nossa participação na vida pública” (GILLESPIE, 2018a, p.96).

Posto isto, o presente estudo busca refletir sobre como os ARP podem conter vieses, reproduzir preconceitos e reforçar estigmas (GOFFMAN, 1982), destacando alguns dos fatores de cunho social ou tecnológico que possam contribuir para este processo, através de uma revisão crítica da literatura sobre o tema em questão.

METODOLOGIA

A metodologia do estudo consiste em uma pesquisa exploratória (PRODANOV e FREITAS, 2013) sobre a forma a qual os ARP (GILLESPIE, 2018a) agem no ciberespaço e como podem apresentar vieses, reproduzindo preconceitos e reforçando estigmas. Para tal, foi realizada uma revisão crítica da literatura publicada por autores da área de Ciência, Tecnologia e Sociedade - CTS, abordando também casos que envolveram plataformas digitais divulgados pela mídia nos últimos 5 anos.

O primeiro tópico aborda o conceito de Algoritmo de Relevância Pública (GILLESPIE, 2018a) a fim de estabelecer uma base para a análise crítica que segue no próximo tópico, atentando para os aspectos Tecnoculturais e Socioeconômicos

(VAN DIJCK, 2013) das plataformas digitais que podem auxiliar no — e/ou serem os responsáveis pelo — processo de reprodução de preconceitos observado nos ARP.

O segundo tópico analisa os estudos científicos e casos revelados pela mídia, os quais tentam demonstrar a dimensão dos problemas sociais causados pelo viés algorítmico. São abordados casos de racismo, misoginia, discriminação por classe social, influência de algoritmos em suicídio e até uma possível intervenção em eleições políticas. Então, busca-se refletir sobre o modo o qual as plataformas digitais lidam com esse problema, atentando para as intenções demonstradas em discursos das próprias plataformas e a comparação com a realidade observada nos casos revelados pela mídia e por estudos científicos.

Por fim, são apresentadas as considerações finais, visando apresentar algumas possibilidades de ação frente ao problema observado.

ALGORITMOS RELEVÂNCIA PÚBLICA

Este tipo de algoritmo é aquele cuja função é determinar a relevância das informações entregues aos usuários de plataformas de mídia social, em resultados de mecanismos de busca na web e em sistemas de recomendação (GILLESPIE, 2018a). Algoritmos como o de *Trending Topics (TT)* do Twitter, o qual promete entregar aos usuários de uma determinada localidade geográfica os assuntos que são “tendência” no momento, utilizam diversos parâmetros pré-determinados pela plataforma — não somente a quantidade de *tweets* sobre o mesmo tema, mas também se o assunto já esteve nos *TT* anteriormente, se o volume identificado realmente se trata de novas postagens, ou são apenas *retweets*, e vários outros — para definir a relevância desses temas para o público o qual eles se destinam. Esses parâmetros, além de serem desconhecidos pelos usuários que recebem esses conteúdos, mudam com grande frequência, criando uma discrepância entre o que os usuários esperam receber, ou mesmo creem estarem recebendo, e o que de fato os algoritmos entregam (GILLESPIE, 2011).

O que Gillespie (2018a) ressaltou como mais importante ao definir os ARP, é o fato de que eles estão “produzindo e certificando conhecimento” (GILLESPIE, 2018a). Isto é, os ARP definem, através de uma lógica própria, o que deve ser considerado relevante ao gerar conhecimento sobre determinado assunto, e as pessoas recorrem a eles para buscar esse conhecimento da mesma maneira que o fazem com especialistas (GILLESPIE, 2018a). A partir disto, Gillespie (2018a) sinaliza que os ARP podem apresentar ramificações políticas ao serem inseridos nas práticas de conhecimento humano (GILLESPIE, 2018a).

Primeiramente, os padrões de inclusão e exclusão de dados que serão utilizados posteriormente pelos ARP devem ser considerados em uma análise crítica destes agentes (GILLESPIE, 2018a). Este procedimento é realizado, pois os algoritmos são projetados para trabalharem com dados previamente categorizados e, quando um dado não se enquadra em nenhuma das categorias pré-determinadas pelo seu criador, este é excluído com a finalidade de não interferir no resultado esperado. Então, a categorização — mais precisamente a escolha das categorias — e a decisão do que vai ser excluído e incluído possuem considerável valor político (GILLESPIE, 2018a). O segundo aspecto a ser considerado está relacionado à incessante tentativa das plataformas em antecipar

para os usuários os conteúdos que eles “vão querer acessar” no futuro, buscando entender ao máximo os usuários e prevendo-os (GILLESPIE, 2018a). Para isso, coletam um grande volume de dados sobre cada usuário, como localização, dados pessoais, *likes*, cada prática realizada dentro dessas plataformas gera informações que são utilizadas pelos ARP. As questões políticas desse aspecto são a privacidade das pessoas que estão utilizando dessas plataformas, que passam a ter seus dados pessoais comercializados para outras empresas sem mesmo tomar conhecimento sobre isso, e o fato de que “os hábitos de percepção ou interpretação de alguns usuários são tomados como modelo universal, hábitos contemporâneos são imaginados como atemporais, objetivos computacionais particulares são tratados como auto evidentes” (GILLESPIE, 2018a).

Cada algoritmo tem como premissa tanto um pressuposto sobre a análise adequada de relevância, quanto em uma representação desse pressuposto em uma técnica de avaliação (computacional). Podem existir premissas implícitas incorporadas na ideia de relevância de um site, podem existir atalhos incorporados em sua representação técnica dessa ideia, e podem existir fricções entre elas duas (GILLESPIE, 2018a, p. 106).

A avaliação dos ARP para determinar o que é o ou não relevante para os usuários faz parte de um cálculo complexo que implica uma quantidade não conhecida de parâmetros e intervenções das redes sociotécnicas por trás deles, parâmetros e intervenções que também são desconhecidos pelos usuários (GILLESPIE, 2018a). Ademais, o termo “relevância” nesse contexto depende da interpretação de cada plataforma, já que seus significados podem ser variados, fazendo com que a criação dos parâmetros consista apenas da subjetividade de seus criadores (GILLESPIE, 2018a). Esse tipo de interpretação dúbia de termos pelas plataformas digitais já foi sinalizada por Van Dijck (2013), quando demonstrou a evolução do termo “compartilhamento” no Facebook ao longo do tempo. Por tratar-se de um termo ambíguo, pode referir-se à distribuição de informações pessoais entre os usuários da plataforma, mas também implica a divulgação dessas informações pessoais a terceiros (VAN DIJCK, 2013). Além disso, através das conexões possíveis com outros usuários da plataforma e as interações “likes”, os termos “amizade” e “gosto” foram imputados com significados tecnoculturais e socioeconômicos específicos (VAN DIJCK, 2013). À relevância também estão implicados aspectos culturais e econômicos quando as plataformas digitais inserem objetivos comerciais nestes algoritmos, incluindo anúncios pagos, por exemplo, entre os resultados considerados mais relevantes para um usuário no *Feed* de Notícias do Facebook (GILLESPIE, 2018a). Apesar de todos estes fatos que comprovam o contrário, as plataformas mantêm um discurso que afirma que seus algoritmos são imparciais e objetivos, fundados em descrições vagas e amplas que não explicitam que seus parâmetros foram pré-programados por pessoas (GILLESPIE, 2018a; JURNO e DALBEN, 2018). O fato é que esses algoritmos são moldados por seus criadores e mantenedores, mas também são definidos pelas suas próprias ações a partir das informações que processam e do que seus usuários fazem deles no seu uso cotidiano (GILLESPIE, 2018a). Porém, as plataformas continuam sendo as únicas responsáveis por criá-los e mantê-los, obtendo assim todo o poder sobre eles e, principalmente, sobre o entendimento que os usuários possuem acerca deles, possibilitando a manutenção de uma incerteza sobre seus critérios (GILLESPIE, 2018a).

Por fim, as plataformas criam o que Gillespie (2018a) chama de “públicos calculados”, que são uma espécie de apresentação algorítmica, ou categorização, do público em certos grupos, moldando a própria percepção de si desse público dentro das plataformas (GILLESPIE, 2018a). Esse tipo de categorização tem como base preferências e previsões algorítmicas estabelecidas por interações prévias nesses ambientes. Os problemas dessa criação seriam a extrapolação, ou seja, a generalização de algum fato para diante de uma população inteira com base em uma parcela desta, visto que o ARP só tem acesso aos dados dos usuários daquela plataforma a qual ele pertence e, mais importante, as intenções pelas quais as plataformas criam essas representações e se elas passam a se tornar legítimas por seus usuários ou pelas autoridades (GILLESPIE, 2018a).

VIESES EM ALGORITMOS DE RELEVÂNCIA PÚBLICA

O processo de criação dos algoritmos se dá através da transcrição de ideias teóricas para códigos computacionais em forma de instruções objetivas, uma espécie de tradução da linguagem humana para a linguagem da máquina – códigos e operações matemáticas – que possui consideráveis limitações tecnológicas na reprodução de algumas dessas ideias (FINN, 2017). Nesse contexto, é importante atentar-se para o fato de que linguagens são construções socioculturais onde estão inseridas questões subjetivas e ideológicas (FINN, 2017). O que é afirmado por Bucher (2018), quando diz que “algoritmos nunca são neutros, mas sim refletem os valores e a cultura das pessoas que os escrevem” (BUCHER, 2018, p. 90, tradução nossa).

Tarleton Gillespie (2018a) afirma que:

[...] à medida que adotamos ferramentas computacionais como nossos principais meios de expressão e passamos a fazer não só da matemática mas de toda a informação “digital”, passamos a sujeitar o discurso e o conhecimento humano a essas lógicas procedimentais que sustentam toda a computação. E há implicações específicas quando usamos algoritmos para selecionar o que é mais relevante a partir de um corpus de dados composto por rastros das nossas atividades, preferências e expressões (GILLESPIE, 2018a, p. 97).

Sendo assim, vieses e preconceitos já presentes no contexto social podem passar a ser reproduzidos e perpetuados através dos algoritmos no ciberespaço, seja através da subjetividade inerente à sua criação; através dos dados interpretados por eles — e das interferências humanas nesses dados, necessárias para que os algoritmos possam trabalhar com eles; por limitações tecnológicas; por serem realizações instáveis; ou porque sua introdução nas práticas de conhecimento humano podem conter ramificações políticas (GILLESPIE, 2018a; JURNO e DALBEN, 2018). Todavia, é extremamente importante não se observar os algoritmos como agentes únicos e responsáveis por toda ação e influência dos outros atores das redes sociotécnicas as quais pertencem, pois eles também são agenciados e influenciados pelas ações destes outros atores (JURNO e DALBEN, 2018).

Casos de escolha algorítmica que demonstram a veracidade dessas afirmações têm sido expostos pela mídia e por pesquisadores nos últimos anos, reforçando a importância dos problemas sociais que emergem na internet por meio da operação

destes agentes técnicos autônomos, ditos imparciais e objetivos por seus mantenedores (GILLESPIE, 2018a, p. 106).

Ali et al (2019) realizaram testes com a ferramenta de anúncios da plataforma Facebook a fim de identificar a que público estes seriam entregues, já que a empresa mantém em segredo as regras de seus algoritmos, que são protegidas por acordos de confidencialidade e leis de sigilo comercial (ALI et al, 2019). A ferramenta permite que os anunciantes selecionem o tipo de público que desejam atingir — ou excluir —, baseando-se em parâmetros como localização geográfica, faixa etária, gênero, idiomas, dados demográficos, interesses, comportamentos e outros². Então, o Facebook determina, através de seus algoritmos e da grande massa de dados que possuem sobre seus usuários, quais pessoas verão esses anúncios por um período de tempo pré-determinado. Os pesquisadores observaram uma significativa tendência de entrega de anúncios a partir de linhas raciais e de gênero, considerando os testes efetuados com oportunidades de emprego e moradia, apesar dos parâmetros neutros de segmentação escolhidos por eles em uma tentativa de garantir um elevado nível de inclusão (ALI et al, 2019). Anúncios de emprego de várias áreas ofertadas no estado da Carolina do Norte – EUA foram criados e, sem nenhum parâmetro além deste, o Facebook exibiu vagas de emprego de caixa de supermercado para um público de 85% mulheres; empregos em empresas de táxi para uma maioria de usuários negros (75%); e na indústria madeireira foram 72% brancos e 90% homens (ALI et al, 2019). Analisando os resultados, pode-se observar que o algoritmo de entrega de anúncios do Facebook agiu exatamente como descrito por Gillespie (2018a), quando apresentou as vagas de emprego para “públicos calculados” com base em uma avaliação de relevância cujos parâmetros estão ocultos, mas a decorrência de suas ações expressa um fato inegável sobre ele: está enviesado, reproduzindo preconceito racial e de gênero.

Apesar da própria discriminação observada no estudo, ao utilizar a ferramenta de Gerenciamento de Anúncios do Facebook, é solicitado ao usuário que aceite uma norma da plataforma intitulada Política de Não Discriminação³, que diz:

As Políticas de Publicidade do Facebook proíbem que os anunciantes usem nossos produtos de anúncios para discriminar indivíduos ou grupos de pessoas. Os anúncios são discriminatórios quando negam oportunidades para indivíduos ou grupos de pessoas com base em determinados atributos como raça, etnia, nacionalidade, religião, idade, gênero, orientação sexual, identidade de gênero, estado civil/status familiar, deficiência ou condição médica ou genética (FACEBOOK, 2019).

Mesmo com essa ação da plataforma em atentar os seus usuários para que não utilizem parâmetros discriminatórios, os resultados das exibições de anúncios provam que os algoritmos não seguem a mesma política. Esse fato deixa o questionamento da real preocupação da plataforma em atentar para a busca do fim da discriminação em seu ambiente *online*.

Singer (2018) sinalizou, em matéria para o jornal *The New York Times*, que o *software* de reconhecimento facial Rekognition⁴, desenvolvido pela Amazon e utilizado nos EUA por autoridades para identificação de criminosos, apresentou erros graves ao ser submetido em um teste com parlamentares norte-americanos (SINGER, 2018). A experimentação foi realizada utilizando imagens de todos os

legisladores federais, as quais foram comparadas com um conjunto de 25.000 fotografias de criminosos e, como resultado, o *software* identificou 28 membros do Congresso erroneamente, em sua maioria negros, segundo um relatório da *American Civil Liberties Union – ACLU* (SINGER, 2018). Buolamwini e Gebru (2018) demonstraram, ainda, que algoritmos da mesma empresa apresentaram um erro de mais de 30% na classificação de gênero em mulheres negras a partir do reconhecimento facial em imagens (BUOLAMWINI e GEBRU, 2018). Esses algoritmos utilizados no *software* de reconhecimento facial apresentam um perigoso viés, reforçando preconceitos de raça e gênero, e pondo em risco a liberdade de pessoas inocentes no momento em que são implantados por serviços de segurança pública para identificar criminosos.

Além dos casos apresentados envolvendo preconceito racial e de gênero sendo reproduzidos e reforçados pelos ARP, outro problema social de grande relevância, o suicídio, foi demonstrado ter sofrido influência desses agentes no caso da adolescente britânica Molly Russell em 2017 (GERRARD e GILLESPIE, 2019). A adolescente buscou por imagens de automutilação e suicídio *online*, o que foi descoberto por autoridades, mas também recebia imagens deste mesmo assunto das plataformas as quais utilizava para consumir conteúdo digital (GERRARD e GILLESPIE, 2019). Os algoritmos dos sistemas de recomendação indicavam para ela mais do conteúdo o qual havia buscado, pois a encaixou dentro de um “público calculado” e indiscriminado, apenas preocupando-se em entregar mais dos seus interesses e o fizeram, mesmo que isso violasse as diretrizes das próprias comunidades das plataformas (GERRARD e GILLESPIE, 2019). Sistemas de recomendação são os principais agentes tecnológicos das plataformas de mídia social, responsáveis por manter os usuários no site por mais tempo, visualizando e clicando em anúncios que geram renda para as empresas que as criam e mantêm. Esses algoritmos são desenvolvidos para analisar as preferências de um usuário e recomendar para ele o conteúdo considerado mais relevante, não estando preparados para atingir as expectativas, ou trabalhar em benefício coletivo (GERRARD e GILLESPIE, 2019).

Diante dos problemas sociais disseminados pelos ARP já apresentados, também são observados outros casos de racismo (SKEEM e LOWENKAMP, 2015), misoginia (CALISKAN, BRYSON e NARAYANAN, 2017; SALAS, 2017), discriminação por classe social (SALAS, 2017) e até uma possível intervenção algorítmica em eleições políticas (MONTEIRO, 2017). Ainda, Cathy O’Neil (2016), afirma que, além de problemas gerados por falhas na seleção dos dados interpretados pelos algoritmos e o preconceito subjacente na sociedade que acabam influenciando as escolhas algorítmicas, o maior problema estaria no modelo econômico adotado por algumas empresas que os mantêm pois, para elas, o dinheiro prova que seus modelos estão funcionando, que o *software* está “fazendo o seu trabalho” e o lucro acaba servindo como um substituto, ou procurador, da verdade (O’NEIL, 2016, p.12).

Já para os usuários das plataformas, os algoritmos estão invisíveis, suas regras ofuscadas e os dados e metadados utilizados em suas escolhas operam em uma escala difícil de compreender (GILLESPIE, 2018a, p. 117).

[...] nós queremos nos livrar do dever de sermos céticos sobre informações que nunca poderemos garantir com certeza. Através desses mecanismos, chegamos a um acordo sobre esse problema (se não o resolvemos) e, assim, eles são soluções com as quais não

apenas podemos contar, mas em que devemos acreditar (GILLESPIE, 2018a, p. 117).

A confiança cega nas informações elencadas pelos algoritmos nas plataformas é um problema cujas consequências se estendem para a convivência em sociedade, pois, segundo Recuero (2009) “[as plataformas de mídia social] ocupam um papel central nas profundas mudanças experimentadas em todos os aspectos da vida social”. Entretanto, sinaliza que “a natureza, os motivos, prováveis e possíveis desdobramentos dessas alterações são extremamente complexos, e a velocidade do processo tem sido estonteante” (RECUERO, 2009, p. 12). Deste modo, os grupos sociais se apropriam das plataformas, que possuem um considerável potencial comunicativo e possibilitam a reconstrução de práticas do dia a dia, onde são criados espaços conversacionais semelhantes aos da conversação física (RECUERO, 2012). Contudo, as informações oferecidas aos usuários dessas plataformas podem ser adaptadas especificamente às suas preferências, tanto por ele próprio, quanto pelo provedor, fazendo com que o conteúdo consumido por um usuário possa se diferenciar do outro de maneira discrepante (GILLESPIE, 2018a), os conduzindo, assim, para dentro de “filtros-bolha” (PARISER, 2011) que influenciam diretamente na diversidade do conhecimento público e do diálogo político, “onde encontramos apenas as notícias que esperamos encontrar e as perspectivas políticas que já nos são caras” (GILLESPIE, 2018a).

Cabe uma análise dos “filtros-bolha” (PARISER, 2011) sob a perspectiva das identidades sociais real e virtual de Goffman (1982). A primeira, a identidade social real, é a que interage com as outras pessoas do grupo social, conduzindo também a maneira com a qual elas respondem ao indivíduo; e a segunda, é algo não concreto, virtual, onde se encontram as impressões que as outras pessoas formam sobre esse indivíduo, inclusive escondendo todas as informações sobre ele as quais deseja que os outros não tomem conhecimento (GOFFMAN, 1982). Assim, o resultado da ação dos “filtros-bolha” (PARISER, 2011) pode causar a impressão aos usuários das plataformas de que existe o equilíbrio entre ambas as identidades nesse ambiente pois, ao receber da filtragem dos algoritmos somente informações que reafirmam suas crenças e anseios, estão escondendo, ou sendo escondidos, de tudo aquilo o que lhes é desconfortável — e fazendo também com que os outros indivíduos da bolha percebam somente o que lhes interessa sobre ele. Da mesma forma, continua ocorrendo a interação social — mediada pelas plataformas e seus algoritmos — característica da identidade social real desse indivíduo (PARISER, 2011; GOFFMAN, 1982).

Polarizando opiniões, reafirmando ideologias e conduzindo os usuários para ideais políticos enviesados, os algoritmos das plataformas podem reforçar preconceitos e proporcionar uma ideia de legitimação de discursos de ódio contra minorias, tudo isso a custos altos para os que já sofrem com seus estigmas.

CONSIDERAÇÕES FINAIS

Plataformas, sob uma visão geral, são ferramentas tecnológicas que permitem que pessoas conversem, pesquisem, comprem, vejam vídeos, organizem viagens e realizem diversas outras atividades online (VAN DIJCK, POELL e DE WAAL, 2018). Mas elas não são apenas *websites* ou aplicativos facilitadores, como são

constantemente definidas pelas próprias empresas que as criam e mantêm, e sim mediadores que moldam a performance dos atos sociais (VAN DIJCK, 2013). Elas proveem tecnologias (*software* e às vezes *hardware*) e serviços que ajudam a codificar as atividades sociais em uma arquitetura computacional, processam dados e metadados através de algoritmos antes de apresentar a sua lógica interpretada através de interfaces amigáveis no navegador ou em aplicativos, refletindo as escolhas estratégicas do seu proprietário (VAN DIJCK, 2013), onde elas não produzem o conteúdo, mas fazem escolhas importantes sobre ele (GILLESPIE, 2018b, p.19). Dito isto, pode-se afirmar que plataformas são agentes privados que influenciam diretamente em valores públicos (VAN DIJCK, POELL e DE WAAL, 2018). Entretanto, não é correto afirmar que a tecnologia é a responsável pela condução de todo esse discurso, mas sim a ação humana na criação desses agentes — os algoritmos — e nos dados os quais são interpretados através deles.

É imprescindível que os procedimentos seletivos desses agentes, que estão invisíveis e, em grande parte, inacessíveis para os usuários das plataformas em geral, sejam questionados. Erros ou vieses nesses procedimentos podem causar estragos de grandes proporções em diversas áreas, dependendo do contexto onde estiverem inseridos, resultando inclusive em processos de exclusão social característicos das sociedades contemporâneas.

Sendo assim, é necessário que sejam criadas regulamentações para que as plataformas se responsabilizem pelas ações dos seus algoritmos e que suas lógicas se tornem menos inacessíveis ao público que consome os dados gerados através de suas escolhas. A transparência poderá permitir que os usuários criem uma visão crítica sobre elas. Ademais, a diversidade nas empresas desenvolvedoras desses algoritmos poderia diminuir, ou evitar, problemas em sua lógica, visto que a subjetividade é uma questão intrínseca nesses agentes, a atuação de um quadro de desenvolvedores e testadores de *software* diverso pode proporcionar uma variedade de visões e enriquecer o resultado final das escolhas feitas por ele. Por fim, a conscientização das mediações executadas pelos algoritmos é a ação de impacto mais imediato e eficaz na tentativa de alertar para a importância de haverem mudanças na maneira com que essas tecnologias são criadas e mantidas pois, estando tão inseridas na rotina diária de milhares de pessoas, elas estão moldando diversos aspectos da vida social.

The machine politics: biases in public relevance algorithms

ABSTRACT

The practices of everyday life are being increasingly reproduced under the mediation of digital platforms (JURNO and DALBEN, 2018). In this context, Public Relevance Algorithms (PRA) determine which information is relevant to users, an important function for their participation in social groups (GILLESPIE, 2018a). Thus, problems that already occur in contemporary societies, such as racial, gender and social class prejudice and others rooted in society may have repercussions in this environment. The present study seeks to reflect on how PRA may contain biases and reproduce prejudice, observing some of the factors that may contribute to this process through a critical review of the literature. It was observed from the results of the analysed researches that the origin of these problems goes through the technology of the algorithms, revealing that much remains to be done to reach the end of social inequality reproduced in digital platforms.

KEYWORDS: Algorithms. Digital platforms. Algorithmic bias. Prejudice.

NOTAS

¹ Somente no Facebook, cerca de 1,52 bilhão de pessoas estão ativas diariamente. Dados divulgados pela própria empresa, referentes a dezembro de 2018: <https://br.newsroom.fb.com/company-info/>.

² Dados observados pelos autores na ferramenta Gerenciador de Anúncios da plataforma Facebook. Disponível em: <https://www.facebook.com/adsmanager>. Acesso em: julho 2019.

³ Disponível em: https://www.facebook.com/policies/ads/prohibited_content/discriminatory_practices

⁴ Disponível em: <https://aws.amazon.com/rekognition/>. Acesso em: julho 2019.

REFERÊNCIAS

ALI, M. et al. **Discrimination through optimization**: How facebook's ad delivery can lead to skewed outcomes. arXiv preprint, 2019. Disponível em: <<https://arxiv.org/abs/1904.02095>>. Acesso em: 18/06/2018.

BUCHER, T. **If...Then**: Algorithmic power and politics. Oxford: Oxford University Press, 2018.

BUOLAMWINI, J.; GEBRU, T. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In: **Proceedings of the 1st Conference on Fairness, Accountability and Transparency**, PMLR 81:77-91, 2018.

CALISKAN, A.; BRYSON, J. J.; NARAYANAN, A. Semantics derived automatically from language corpora contain human-like biases. **Science**, [s.l.], v. 356, n. 6334, p.183-186, 13/04/2017. American Association for the Advancement of Science (AAAS). <http://dx.doi.org/10.1126/science.aal4230>.

CORMEN, T. E., et. al. **Algoritmos**: teoria e prática. 2 ed. Rio de Janeiro: Elsevier, 2002.

FINN, E. **What algorithms want**: imagination in the age of computing. Cambridge: MIT Press, 2017.

GERRARD, I.; GILLESPIE, T. When algorithms think you want to die. **Wired**. 21/02/2019. Disponível em: <<https://www.wired.com/story/when-algorithms-think-you-want-to-die/>>. Acesso em: 14/06/2019.

GILLESPIE, T. A relevância dos algoritmos. **Parágrafo**, São Paulo, v. 6, n. 1, p.95-121, jan.-abr. 2018a. Tradução de: Amanda Jurno.

GILLESPIE, T. Can an algorithm be wrong? Twitter Trends, the specter of censorship, and our faith in the algorithms around us. **Culture Digitally**. 19/10/2011. Disponível em: <<http://culturedigitally.org/2011/10/can-an-algorithm-be-wrong/>>. Acesso em: 20/06/2019.

GILLESPIE, T. **Custodians of the internet**: platforms, content moderation, and the hidden decisions that shape social media. New Haven: Yale University Press, 2018b.

GOFFMAN, E. **Estigma**: notas sobre a manipulação da identidade deteriorada. Rio de Janeiro: Zahar, 1982.

JURNO, A. C.; DALBEN, S. Questões e apontamentos para o estudo de algoritmos. In: **PARÁGRAFO**: REVISTA CIENTÍFICA DE COMUNICAÇÃO SOCIAL DA FIAM-FAAM, v. 6, p. 17-29, 2018.

KEMP, S. **DATAREPORTAL**. Digital 2019: Brazil. Disponível em: <<https://datareportal.com/reports/digital-2019-brazil>>. Acesso em: 22/06/2019.

MONTEIRO, R. L. A perigosa caixa preta dos algoritmos e a campanha eleitoral de 2018. **El País**. 11/10/2017. Disponível em: <https://brasil.elpais.com/brasil/2017/10/11/opinion/1507749770_561225.html>. Acesso em: 14/06/2019.

O'NEIL, C. **Weapons of math destruction**: how big data increases inequality and threatens democracy. Nova Iorque: Broadway Books, 2016.

PARISER, E. **The Filter Bubble**: How the new personalized web is changing what we read and how we think. Nova Iorque: The Pinguim Press, 2011.

PRODANOV, C. C.; FREITAS, E. C. **Metodologia do trabalho científico**: métodos e técnicas da pesquisa e do trabalho acadêmico. 2. ed. Novo Hamburgo: Feevale, 2013.

RECUERO, R. **A conversação em rede**: Comunicação mediada pelo computador e redes sociais na Internet. Porto Alegre: Sulina, 2012. 238p.

RECUERO, R. **Redes Sociais na Internet**. Porto Alegre, RS: Sulina, 2009.

SALAS, J. Se está na cozinha, é uma mulher: como os algoritmos reforçam preconceitos. **El País**. 19/09/2017. Disponível em: <https://brasil.elpais.com/brasil/2017/09/19/ciencia/1505818015_847097.html> . Acesso em: 15/06/2019.

SINGER, N. Amazon's Facial Recognition Wrongly Identifies 28 Lawmakers, A.C.L.U. Says. **The New York Times**. 26/07/2018. Disponível em: <<https://www.nytimes.com/2018/07/26/technology/amazon-aclu-facial-recognition-congress.html?smtyp=cur&smid=tw-nytnational%E2%80%9CThis>>. Acesso em: 15/06/2019.

SKEEM, J. L.; LOWENKAMP, C. T. **Risk, Race, & Recidivism**: Predictive Bias and Disparate Impact. *Ssrn Electronic Journal*, [s.l.], p.1-55, 2015. Elsevier BV. <http://dx.doi.org/10.2139/ssrn.2687339>.

VAN DIJCK, J. **The Culture of Connectivity**: A Critical History of Social Media. Oxford: Oxford University Press, 2013.

VAN DIJCK, J.; POELL, T.; DE WAAL, M. **The Platform Society**: Public Values in a Connective World. Oxford: Oxford University Press, 2018.

Recebido: 23/07/2019
Aprovado: 13/06/2020
DOI: 10.3895/rts.v16n45.10380

Como citar: SCHULER, L.G.B.; MONTARDO, S.P. A política das máquinas: vieses em algoritmos de relevância pública. *Rev. Technol. Soc.*, Curitiba, v. 16, n. 45, p. 300-312, out./dez., 2020. Disponível em: <https://periodicos.utfr.edu.br/rts/article/view/10380>. Acesso em: XXX.

Correspondência:

Direito autoral: Este artigo está licenciado sob os termos da Licença Creative Commons-Atribuição 4.0 Internacional.

