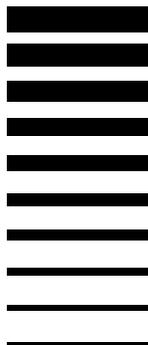


2024



∴
conferência
tecnologia/> e sociedade



coletivo

quebradev



Palavra Chave

Muitas vezes
A prisão era um martírio

Seu calabouço
Não dispunha de saídas

Havia uma porta
Porta que dava acesso à outros calabouços

Ela ficava fechada

Havia uma pequena fresta embaixo
De apenas um milímetro

Eis que ele teve uma brilhante ideia

Todas as manhãs
Através de palavras em papéis
Passava sua cabeça por debaixo da porta

Suas palavras libertavam outros presos

Outras palavras de cabeças que passavam
debaixo da mesma fresta, libertava-no.

Kaio Teixeira



Prefácio

O conhecimento tem sido transmitido por diversas gerações e de várias formas, inicialmente por meio da comunicação e gesticulação, e, principalmente, pelo aprendizado prático Cura et al. (2008). Posteriormente, surgiu a necessidade de registrar esse saber, dando início à escrita e à arte, historicamente representadas pelas figuras rupestres, que podem ser vistas até os dias atuais em sítios arqueológicos Gerber (2009).

A importância da produção e do registro do conhecimento é evidente. O saber transmitido de geração em geração nos permite avançar e construir com base em experiências anteriores. Cada passo é construído sobre o outro: não teríamos a teoria da relatividade geral de Einstein sem antes a lei da gravitação de Isaac Newton; não haveria vacinação em massa sem o sequenciamento do DNA por Craig Venter; e o avanço das teorias sociais, como o socialismo científico, não seria possível sem a produção e o registro do conhecimento em grandes bibliotecas por pensadores como Lenin, Althusser, Marx, Rosa Luxemburgo, entre outros. No passado, a grande dificuldade em compartilhar conhecimento era o acesso.

Quem produzia saber na Ásia tinha dificuldades em compartilhar suas experiências com pessoas na América, e vice-versa. Embora parte do conhecimento social seja aplicável apenas a contextos específicos, ele pode servir de base para a análise e síntese em outras sociedades. Por outro lado, o conhecimento das ciências exatas não enfrenta essa barreira de isolamento, pois as mesmas condições experimentais produzidas na Ásia, por exemplo, podem ser reproduzidas na América do Sul, desde que o conhecimento sobre como essas condições foram criadas e seus resultados seja compartilhado.

Para difundir o conhecimento científico, surgiram diversas formas de publicação, como revistas científicas, jornais acadêmicos, congressos e entrevistas em rádio e TV. Muitas dessas publicações demoravam a ser produzidas e distribuídas para outras universidades e centros de pesquisa, limitando o alcance do conhecimento. A demora era ainda maior quando se dependia de cópias xerox de livros de bibliotecas, exceto em casos de grandes descobertas, que, mesmo assim, ainda enfrentam dificuldades até hoje.

Atualmente, os meios de comunicação científica envolvem custos operacionais, tanto para a revisão quanto para a publicação de artigos. Embora muitas revisões sejam feitas de forma voluntária, há despesas desde a chamada para trabalhos até a publicação final. O problema é que esses meios de divulgação são geralmente controlados por grandes conglomerados do conhecimento, como ACM, IEEE e Scielo, que visam o lucro. Isso cria um oligopólio que atrasa o avanço científico e social. Esse oligopólio é altamente lucrativo, superando gigantes como Walmart, Toyota e Bank of America. Em 2020 de acordo com Oliveira et al. (2022), o mercado editorial científico foi identificado como um oligopólio sem qualquer tipo de regulação.

A informação científica é um insumo essencial para qualquer estrutura de pesquisa, e o mercado editorial de ciência se tornou um dos mais rentáveis. Em 2010, o grupo Elsevier registrou um lucro de R\$ 5,1 bilhões, com uma receita de R\$ 14 bilhões. Com o avanço da internet em 2002, o movimento de acesso aberto ganhou força, reivindicando o direito ao acesso público e gratuito ao conhecimento científico. A internet permitiu a ampla divulgação desse conhecimento, possibilitando que pessoas sem vínculo formal com instituições de pesquisa, estudantes e universidades ao redor do mundo tivessem acesso a novos resultados e avanços, além de colaborar

no avanço ou refutação de pesquisas infundadas.

A produção de conhecimento e o avanço científico, ainda concentrados em centros universitários e empresas privadas, sofrem com a falta de perspectivas periféricas e de pessoas que vivem nos territórios marginalizados. Com isso em mente, esta revista tem como objetivo expandir a capacidade crítica e o desenvolvimento científico, amplificando as vozes desses cientistas.

Nosso propósito é ir além das barreiras acadêmicas, propondo que a produção técnico-científica sirva como ferramenta de confronto ao status quo, seja revolucionária e construída pelos trabalhadores de todas as áreas. Aqui, você encontrará produções que desafiam, dialogam e inovam, conectando o saber técnico com as realidades das nossas periferias.

O QuebraDev

Um coletivo de tecnologia de origem periférica com destino voltado à origem. A conferência Tecnologia e Sociedade (CTS) Temos como objetivo debater a tecnologia em uma perspectiva geral do uso das tecnologias criadas, tais como, inteligência artificial, reconhecimento facial, algoritmos entre outros e o impacto destas na sociedade trazendo essa discussão sobre um olhar de raça e classe.

Na primeira edição no ano de 2023, fizemos a pergunta: Tecnologia para Quem? Com isso, trouxemos diversas pessoas para desafiar a ideia de que a tecnologia deve servir apenas a propósitos mercadológicos, trazendo o debate para a esfera do uso da tecnologia em benefício das comunidades

periféricas e apresentando reflexões sobre como isso pode ser alcançado.

Na segunda edição 2024, a que você vai encontrar os temas debatidos nesta revista, abordamos o tema: Tecnologia para Todos! Queremos que a produção técnico-científica da tecnologia seja pensada para todas as pessoas e que possa alcançar a todos. A ideia central é projetar e refletir sobre as possibilidades dessas tecnologias em nossos territórios, construídos pelos coletivos populares, cientistas de grupos sociais marginalizados e trabalhadores.

Revista Tecno-Científica do Coletivo QuebraDev

A revista tecno-científica do coletivo surge como um instrumento de luta e produção de novos horizontes a partir da classe trabalhadora e do povo periférico.

Acreditamos que a tecnologia deve ser construída e debatida sob essa perspectiva, e para isso é fundamental criar espaços que transcendem as fronteiras das universidades e empresas. A técnica e o conhecimento científico precisam ser produzidos pelo povo e para o povo.

Por isso, vemos essa iniciativa como essencial para ampliar o impacto da tecnologia na sociedade atual.

Contamos com 19 submissões, das quais duas apresentações e quatro artigos foram aceitos, levando isso em consideração, a possibilidade de aceitação é de 31,58%, agradecemos todos que submeteram as excelentes ideias e propostas e colaboraram com o progresso do avanço científico.

Referências

Cura, S., Cura, P. and Oosterbeek, L. (2008). Projecto andakatu: Didáctica da pré-história através da experimentação, *Arqueologia Experimental* p. 54.

Gerber, I. (2009). Figuras rupestres: arte e/ou escrita?, *Ide* **32**(48): 26–34.

Oliveira, M. F., Kowaltowski, A. and Silber, A. M. (2022).

No lucrativo mercado da ciência aberta, quem paga a conta?, *Ciência Hoje* **387**. Acesso em: 18 jun. 2024. **URL:**

<https://cienciahoje.org.br/artigo/no-lucrativo-mercado-da-ciencia-aberta-quem-paga-a-conta>

Conteúdo

1. Palestras	9
1.1 Encruzilhadas da educação tecnológica no Brasil: desafios e caminhos para uma educação tecnológica popular e contra colonial.....	10
1.2 Como eu ensino programação para meninas na Brasilândia.....	14
2. Artigos	16
2.1 Motivações.....	16
2.2 Impacto da Divulgação Científica.....	16
2.3 Notas gerais sobre o capítulo.....	17
2.4 Subvertendo Algoritmos e Sistemas de Controle.....	19
2.5 Assinatura dogenocídio nos dados de mortes em Gaza.....	28
2.6 Observatório do Santo Amaro - Uma Iniciativa Crucial para a Comunicação Social Inclusão Digital nas Favelas.....	36
2.7 TBC-DTW: Medição automática de viés midiático baseada em modelagem de tópicos.....	44

1. Palestras

As palestras são ferramentas essenciais para a ampla divulgação de ideias e técnicas, sendo amplamente utilizadas em eventos de tecnologia, acadêmicos e empresariais.

Elas têm como objetivo apresentar novas perspectivas, compartilhar histórias pessoais e estimular a reflexão. Neste capítulo, apresentamos dois resumos de palestras ministradas na Conferência Tecnologia e Sociedade: Tecnologia para Todos!

Convidamos você a conhecer os resumos e a biografia dos palestrantes.

1.1 Encruzilhadas da educação tecnológica no Brasil: desafios e caminhos para uma educação tecnológica popular e contra colonial



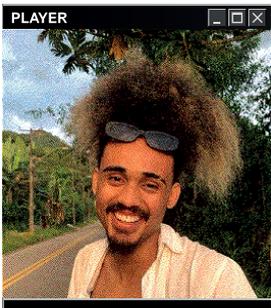
Autora

Marcela da Silva
Canto Bastos

marcela.canto98@gmail.com

Mini-Bio

Marcela foi bolsista do programa Youth do CGI.br em 2024, é graduanda em Ciências Sociais pela Universidade Federal Fluminense, e realiza pesquisas nas áreas de Sociologia Digital e Letramento Digital. Integrante do grupo de pesquisa Pensamento Social em Dados (Data_PS) e educadora popular de Sociologia, cidadania digital e letramento digital em São Gonçalo- RJ.



Autor

Rafael Ramires
Baptista

educador.ramires@proton.me

Mini-Bio

Ramires é sulmatogrossense e morador intermunicipal na Baixada Fluminense. Atua como educador popular em tecnologias digitais, trabalha como especialista em segurança digital e da informação na Escola de Ativismo e pesquisa exclusão digital e tecnológica.

Formado em Tecnologias Educacionais pela Universidade Estácio de Sá, atualmente cursa o Mestrado em Humanidades Digitais na Universidade Federal Rural do Rio de Janeiro. É integrante da InfoCria, o primeiro coletivo de educação tecnológica popular da Baixada Fluminense.

Integra o coletivo InfoLambe, que trabalha a Geração Cidadã de Dados na Baixada Fluminense, misturando arte e tecnologia para comunicar e produzir dados. É cineclubista no Cineclubes Imbariê nos Trilhos, e colabora ativamente com movimentos sociais da Região Metropolitana do Rio de Janeiro, somando forças na facilitação do uso de equipamentos tecnológicos para o ativismo no território.

Resumo da Palestra

A apresentação teve como objetivo abordar as atuais legislações e documentos normativos sobre o currículo em educação digital como a Base Nacional Comum Curricular, o Plano Nacional de Educação Digital e a Lei de Diretrizes e Bases da Educação e suas insuficiências para o contexto brasileiro; e buscou dialogar sobre as diferentes vivências concretas dentro e fora do chão da escola pública e apontará caminhos para uma educação digital que seja emancipadora e forme cidadãos digitais éticos e participativos.

Ademais, elucidou sobre o que se constitui a cidadania digital, que não se restringe ao uso da tecnologia, mas sim o uso ético, responsável, seguro e participativo da Internet. A apresentação buscou demonstrar que a cidadania envolve inclusão, direitos, deveres e obrigações; e que para tal é necessária a integração e o reconhecimento de cada indivíduo enquanto sujeito digno de direitos.

Para isso ocorrer, práticas discriminatórias no ambiente digital devem ser combatidas não só através da criminalização como também através de práticas pedagógicas que almejam a formação para cidadania dentro e fora do digital. Cada participante trouxe suas experiências para dialogar de forma interseccional abordando a necessidade de um currículo que leve em conta tanto questões de raça, gênero, classe e sexualidade; entendendo que todas as questões formam um único nó dinâmico sendo inseparáveis de uma sociedade que tem um real compromisso com os direitos humanos.

A apresentação teve também como objetivo apresentar possíveis caminhos e medidas a serem adotadas no currículo e nas políticas públicas voltadas para educação digital.

Dando enfoque à realidade concreta das escolas e buscando caminhar para uma educação emancipadora que busque não só o desenvolvimento da autonomia, mas também o desenvolvimento do pensamento crítico, antirracista, feminista, anti LGBTQIA+fóbico, visando a construção de uma sociedade mais justa e participativa.



1.2 Como eu ensino programação para meninas na brasilândia



Autora

Larissa
Vitoriano

larissavitoriano@gmail.com

Mini-Bio

Mestra em Tecnologias da Informação, Comunicação e Multimédia pelo Instituto Universitário da Maia, no Porto, Portugal. Jornalista formada pela Faculdade Paulus de Tecnologia e Informação (FAPCOM).

Atuo principalmente como professora front-end para crianças e adolescentes e tech community manager. Idealizadora do projeto Mais Meninas na Tecnologia, que visa a educação tecnológica para estudantes entre 8 e 10 anos de escolas públicas periféricas na zona norte de São Paulo. Autora do livro infantil "Minidev: a menina que deseja programar".

Resumo da Palestra

O futuro (e o presente!) está sendo escrito por meio de linhas de código. Mas quem coda essa features?

Nesta palestra foi contado a história da autora e do projeto social Mais Meninas na Tecnologia, que já impactou mais de 90 meninas na região da Brasilândia, na zona norte de São Paulo, com oficinas maker sobre tecnologia, programação, inteligência artificial e empoderamento de meninas periféricas em situação de vulnerabilidade social. Também desenvolvemos o aplicativo Minidev, que ensina por meio de atividades lúdicas programação e conteúdos front-end. Afinal, o Brasil (e o mundo!) merece ter #MaisMeninasNaTecnologia.

2. Artigos

Este capítulo agrega todos os artigos escolhidos para a publicação da primeira edição da revista CTS (Conferência Tecnologia e Sociedade) nos próximos subtópicos discutiremos sobre as motivações, o impacto da divulgação científica e também as notas gerais sobre os artigos aqui publicados.

2.1 Motivações

Nosso principal objetivo é proporcionar um espaço onde vozes diversas possam ser ouvidas. Queremos que pesquisa dores, desenvolvedores, educadores e ativistas compartilhem suas visões e experiências, contribuindo para um diálogo mais amplo sobre a tecnologia e seu papel na sociedade. Ao promover a divulgação científica, buscamos não apenas informar, mas também inspirar ação e reflexão crítica sobre o uso da tecnologia.

2.2 Impacto da Divulgação Científica

A divulgação científica desempenha um papel crucial na democratização do conhecimento. Ao tornar as informações acessíveis a todos, especialmente a grupos que historicamente foram marginalizados, contribuimos para a formação de uma sociedade mais justa e igualitária. O impacto da divulgação vai além do simples compartilhamento de informações; ela estimula a inovação, a educação e a mobilização social.

Esperamos que os artigos apresentados nesta edição da revista não apenas informem, mas também provoquem discussões significativas sobre como podemos usar a tecnologia para o bem comum. Queremos incentivar todos a pensar criticamente sobre suas próprias experiências e a

se envolver ativamente na construção de soluções que beneficiem suas comunidades.

O QuebraDev convida todos os leitores a se juntarem a nós nesta jornada de descoberta e aprendizado. Juntos, podemos transformar a maneira como pensamos sobre tecnologia e sua aplicação na sociedade. Os artigos aqui apresentados são uma contribuição para essa conversa e esperamos que inspirem mais pessoas a se envolverem e a compartilharem suas próprias histórias.

2.3 Notas gerais sobre o capítulo

A ideia de criar uma revista científica surgiu em um novo contexto dentro do nosso coletivo, motivada pelas necessidades da realidade em que vivemos. Muitos dos cientistas que participaram dos nossos podcasts e lives desafiaram o status quo a partir de uma perspectiva inovadora, abordando questões de raça, classe e propondo novos horizontes tecnológicos.

Este capítulo da revista reflete o que acreditamos enquanto coletivo, com o objetivo de produzir conhecimento científico voltado para os trabalhadores e por eles. É dentro dessa nova perspectiva que apresentamos este capítulo.

Neste capítulo, você encontrará quatro artigos. O primeiro, de Maria Alice Silva Santos Felix, explora como subverter algoritmos e sistemas de controle, fazendo um contraponto à ideologia californiana e ao estereótipo do "nerd branco". No segundo artigo, Lucas da Silva Alves Pessoa discute o Observatório de Santo Amaro, uma iniciativa fundamental para a comunicação social e inclusão digital nas favelas.

Ele aborda a coleta e análise de dados, tecnologias emergentes, redes comunitárias para promover conectividade para todos e a alfabetização digital. O terceiro artigo analisa os dados sobre as mortes em Gaza, com o objetivo de compreender o genocídio em curso no território palestino. Por fim, é apresentada uma proposta de medição automática de viés midiático com base na modelagem de tópicos.

Acreditamos que estes artigos trazem uma perspectiva nova sobre a tecnologia para além do desenvolvimento mercadológico, mas para construção de novos horizontes de luta e de direitos para aqueles trabalhadores e trabalhadoras de São Paulo, Brasil e o mundo.

2.4 Subvertendo Algoritmos e Sistemas de Controle



Autora

Maria Alice Silva
Santos Félix

larissacvitoriano@gmail.com

*Universidade Federal de Viçosa
(UFV) – Viçosa – MG – Brasil*

Mini-Bio

Bacharela em Serviço Social pela Universidade Federal de Ouro Preto (UFOP). Mestranda em Política Social pela Universidade Federal de Viçosa (UFV).

Resumo do Artigo

A Indústria 4.0, marcada pelas transformações digitais globais, resgata antigas expressões da questão social, como o racismo. O estudo analisa a integração das Tecnologias da Informação e Comunicação (TIC) e o racismo, com foco no racismo algorítmico e no hacktivismo digital.

A partir de uma revisão de textos clássicos e contemporâneos, busca-se entender como o capitalismo global, com este processo de ampla transformação econômica, social, cultural e política, reatualiza esse fenômeno, impondo novos desafios à população negra.

Ressalta-se a importância de enfrentar essas inovações tecnológicas com ações coletivas, o oposto ao fardo do nerd branco, que se revela como uma ideologia na imagem habitual do hacker, apresentado como um justiceiro solitário.

Introdução

No final do século XIX e início do século XX, eugenistas e defensores do branqueamento racial acreditavam que em 100 anos não haveria mais negros no Brasil. Um dos principais proponentes dessa ideia foi João Batista de Lacerda, médico e antropólogo brasileiro, que, em 1911, na Conferência Internacional de Raças, em Londres, apresentou o trabalho *Sur les métis au Brésil (Sobre os mestiços no Brasil)*. Nele, Lacerda defendia que, através da miscigenação e da imigração europeia, a população negra e indígena seria absorvida, desaparecendo em cerca de cem anos, resultando em uma população predominantemente branca. O ano de 2011, porém, marcou a resistência do povo negro diante desse projeto genocida, tanto no Brasil quanto no mundo. O século XXI, por sua vez, em especial pós-pandemia, marca os novos desafios frente uma soberania digital eurocêntrica.

Essa tentativa de apagamento histórico se insere em um contexto mais amplo. O Brasil se desenvolveu ao longo de aproximadamente 400 anos sob um sistema escravista colonial, e, mesmo após o fim legal desse regime, no início do século XX, o Estado brasileiro implementou novos métodos de exploração e opressão das pessoas de origem africana e de seus descendentes, anteriormente escravizadas. Se antes era frequente que figuras brancas cumprissem o papel do mito do fardo do homem branco, enquanto um herói e salvador das populações não brancas, hoje nos deparamos com uma versão reelaborada. Sob a égide do fardo do nerd branco, esse se revela como uma ideologia na imagem habitual do hacker, apresentado como um justiceiro solitário, adepto da cultura do faça-você-mesmo (do it yourself), que transforma o mundo por conta própria e detém mais capacidade de confrontar o sistema do que uma greve organizada Faustino (2023).

Portanto, este estudo, ao analisar o panorama do racismo algorítmico e estratégias hacktivistas, compreende, como bem colocou Fanon (2021), que no racismo não se expressa apenas sobre as ofensas abertamente violentas ou estereotipa das, mas, também está presente na tentativa de universalizar os valores e referências culturais europeias, como se fossem normas universais aplicáveis a toda a humanidade (ibid). Ademais, o estudo considera esse histórico de formação social e econômica, cujas raízes dão origem às ramificações do racismo estrutural que permeia a sociedade e as relações sociais. Ao analisar as mudanças ocorridas no capitalismo contemporâneo, especialmente com a ampla digitalização da economia, percebe-se que as inovações técnico científicas-digitais não possuem neutralidade, e as desigualdades sociais de classe, gênero, sexualidade e etnia/raça continuam presentes nos ambientes mediados por algoritmos e sistemas informacionais computacionais.

Racismo na Era Digital

Como coloca de Lucena (2024), “para que uma realidade setorne natural, unívoca e amplamente aceitável é primordial que seus antecedentes, processos de constituição ou alternativas sejam apagados, ocultados, silenciados, assim pode-se impedir ou dificultar comparações e críticas ao status quo”. Em outras palavras, para que o racismo se naturalize na sociedade, é necessário silenciar e ocultar os fatores históricos que o construíram, como a escravidão e as políticas de exclusão subsequentes. Esse apagamento dificulta a percepção das injustiças, tornando mais difícil a crítica e a resistência às estruturas racistas. Assim, o racismo se mantém invisível, mas ainda presente e operante. No contexto digital atual, as manifestações abertas e impunes do racismo

refletem um fenômeno global reacionário.

O racismo, fundamentado no contexto histórico e reforçado pela política de branqueamento da população promovida pelo Estado no início do século XX, garantiu a exclusão da população ex-escravizada e de seus descendentes do mercado de trabalho. Essa política favoreceu a ocupação das posições na nascente indústria e nos empregos mais prestigiados do mercado interno por pessoas brancas, marginalizando os negros dessas oportunidades Silva (2022). Hoje, essa divisão racial permanece, e o aparato tecnológico é construído por e para pessoas brancas, como mostram os dados da pesquisa Potências Negras Tec, de 2021, cujo mercado de trabalho em tecnologia é composto predominantemente pelo público masculino (73%) e a maioria dos cargos de liderança é ocupada por brancos (56%).

Essa construção tecnológica sob uma visão branca resulta no fato de que mulheres negras têm uma menor chance de identificação facial checada, além do direcionamento racista sob o qual a tecnologia tem sido fundamentada, como expõe o Gender Shades Project (2018), uma abordagem interseccional sobre a inclusão de raça e gênero na IA, realizado pelo Natural Science and Engineering Research Council of Canada (NSERC).

Outro estudo, realizado pela National Institute of Standards and Technology (NIST), no mesmo ano, expôs que os algoritmos de IA aprimoraram sua eficácia em 25 vezes no reconhecimento facial de usuários. A taxa de erro hoje gira em torno de 0,2. No entanto, esse número não é o mesmo para a população negra, e o quadro se agrava ao se tratar do perfil de mulheres negras. Ainda referente ao mesmo ano, o sistema do C6 Bank não reconheceu a face de um correntista negro, como expõe o artigo Money Times. Enfim, reconhecidos pela IA quando não deveriam ser,

e não reconhecidos pela IA, quando deveriam ser Félix (2022).

Um agravante significativo é a crescente tendência de expansão do uso das tecnologias de reconhecimento facial pelas instituições de segurança pública em todo o mundo. Dados levantados pela Rede de Observatórios de Segurança de março a outubro de 2019, indicaram 151 cidadãos presos a partir da tecnologia e reconhecimento facial nas regiões brasileiras Sul, Sudeste e Nordeste; 90,5% das pessoas abordadas eram negras.

Além da IA no reconhecimento biométrico, também o processo de *Machine Learning* pode agravar fenômenos relacionados ao racismo estrutural. Na medida em que se trata de uma tecnologia que “aprende” com o comportamento online dos usuários, se os usuários expressam um comportamento racista, o algoritmo pode aprofundá-lo ainda mais neste cenário. Assim sendo, acreditamos que, com o desenvolvimento desses algoritmos sem se ter uma intervenção oposta a práticas racistas, caminhamos para o acirramento de expressões do racismo estrutural em meio digital.

Essa seletividade racial está presente, por exemplo, no alcance das redes sociais. Testes feitos por influenciadores digitais no Instagram mostraram que o engajamento do perfil de influenciadores negras aumentou em 6.000% ao publicar fotos de mulheres brancas. Isso mostra que o que antes parecia uma oportunidade para que vozes silenciadas pudessem se expor, Na verdade demonstrou-se como terreno fértil para Numa nova marginalização. Os desvalidos só participam da era da informação enquanto mão de obra barata. Por outro lado, em maio deste ano, a plataforma Canva, sob um prompt de imagem de IA para “réu juvenil monitor de tornozelo na escola” forneceu apenas imagens de meninos negros (Fiesler 2024).

Com o que vimos, podemos dizer com segurança que os aparelhos e serviços high-tech não estão isentos de uma Machine Learning sem reverberações do racismo. Neste sentido, é preciso se estabelecer uma ética tecnológica que, além de não racista, seja anti-racista. Para isso, partimos da compreensão de que a tecnologia não é neutra, o treinamento de robôs utilizando dados humanos responde às dinâmicas econômicas, culturais e políticas, logo, no sistema capitalista, que visa o lucro e a proteção de bens, a divisão racial é reafirmada categoricamente Félix (2022). Cremos ainda que o racismo estrutural que se expressa nas tecnologias distingue negativamente o alcance digital de pessoas com o fenótipo negro, reduzindo oportunidades e acentuando desigualdades. Se o universo do on-line e digital não para de se expandir em todos os cantos do mundo, é vital recordar a importância do hacktivism enquanto estratégia de subversão ao sistema capitalista, que é, fundamentalmente, racista.

Desafios e Perspectivas do hacktivism no século XXI

A compreensão do hacktivism no século XXI deve ser analisada à luz das dinâmicas coloniais e eurocêntricas que moldaram e moldam as estruturas sociais e tecnológicas. O eurocentrismo, historicamente, ocultou produções teóricas africanas e afro-diaspóricas, perpetuando o racismo através de métodos e conceitos que reforçam as desigualdades sociais. Como apontado por Marx, as relações de poder são fundamentais para a criação de categorias como a de “escravo”, uma reflexão que Fanon complementa ao afirmar que “é o branco que cria o negro” negando a humanidade do sujeito racializado. Essa dinâmica inscreve o negro em uma posição subalterna na relação social contemporânea, especialmente no capitalismo.

Dentro desse contexto, o hacktivismo decolonial emerge como uma resistência ao individualismo típico do hacker tradicional, promovendo o agir coletivo e a descolonização dos horizontes tecnológicos. O foco deve estar em iniciativas como os laboratórios de periferia, os “perilabs”, que exercem uma engenharia reversa da tecnologia e da produção de conhecimento. No entanto, é crucial questionar, como propõe Faustino (2023), se o ativismo digital restrito às grandes plataformas monopolistas representa uma verdadeira subversão ou apenas alimenta o ciclo de exploração, prolongando o tempo de permanência dos usuários em sistemas projetados para a extração de dados.

Exemplos concretos de descolonização tecnológica já estão presentes, como o trabalho dos “wikipedistas”, inicialmente marginalizado pela academia e agora amplamente utilizado por grandes plataformas como o ChatGPT e o Bing. A iniciativa “Mais Pretas em Teoria da História na Wiki” é uma evidência desse esforço decolonial. Algoritmos, como produções humanas, refletem as tradições, valores e interesses historicamente determinados que os atravessam.

Portanto, é imperativo considerar os efeitos da Quarta Revolução Industrial (4a RI) para sociedades em posição de dependência, como o Brasil, onde a questão racial é central desde sua formação. A reparação das injustiças históricas requer um letramento digital crítico da população negra, capacitando-a para uma formação política e técnica voltada para a soberania digital popular. Afinal, o mercado digital também lucra com as desigualdades raciais, perpetuando-as no imaginário coletivo. Assim, o hacktivismo decolonial se apresenta como uma ferramenta para confrontar essas estruturas, promovendo a emancipação e a construção de uma nova ordem social e tecnológica.

Referências

de Lucena, M. d. S. (2024). Opressão da (s) mulher (es) na sociedade de classe: “trabalhadora não é o feminino de trabalhador”: Perspectivas feministas em debate, *Revista Ciências do Trabalho* (25).

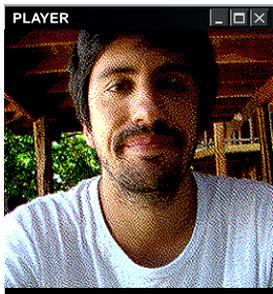
Faustino, Deivison Mendes e Lippold, W. (2023). Que humano é esse das humanidades digitais? por uma crítica hacker-fanoniana ao fardo do nerd branco, *Revista de Teoria da História* 26(1): 120–143.

Fiesler, C. (2024). Casey flesler on linkedin: Last night a tiktok creator posted a video about how on canva, https://www.linkedin.com/posts/casey-fiesler-bb3493243_last-night-a-tiktok-creator-posted-a-video-activity-7188604030170943488-PADW/. Accessed: 2024-10-23.

Félix, M. A. S. S. (2022). Um robô poderia ser racista?, <https://geptsss.ufop.br/blog/um-rob%C3%B4-poderia-ser-racista>. Accessed: 2024-10-23.

Silva, T. (2022). *Racismo algorítmico: inteligência artificial e discriminação nas redes digitais*, Edições Sesc SP.

2.5 A assinatura do genocídio nos dados de mortes em gaza



Autor

Henrique
Xavier

hsxavier@gmail.com

Núcleo de Tecnologia do MTST

Mini-Bio

Doutor em astrofísica com mais de 10 anos de experiência em estatística, programação, análise de dados e modelagem matemática. Atua na área de Aprendizado de Máquina e Ciência de Dados desde 2018.

Resumo

Modelamos a distribuição em termos de idade e gênero dos mortos por ataques israelenses em Gaza no final de 2023 para estimar a fração das mortes que decorreram de ataques focados e de ataques indiscriminados. Concluímos que $82,6 \pm 0,9\%$ das mortes foram causadas por ataques indiscriminados, reforçando as acusações de genocídio atribuídas a Israel.

Introdução

Após conquistar a Palestina ao final da 1ª Guerra Mundial e sem consultar sua população local, de maioria árabe e muçulmana, o Império Britânico definiu, através da Declaração de Balfour, que aquele território seria uma pátria para judeus então perseguidos na Europa. O Sionismo promoveu a migração em massa para a Palestina de judeus da Europa Oriental. Esse processo culminou na expulsão da população Palestina e na criação do Estado de Israel em 1948.

No dia 7 de outubro de 2023, aliados e membros das brigadas Al-Qassam, vinculadas ao Hamas, realizaram um ataque a bases militares e assentamentos israelenses, matando cerca de 1.100 pessoas. Em resposta, Israel iniciou um vasto ataque aéreo e terrestre à Faixa de Gaza, matando mais de 41 mil pessoas até setembro de 2024 (Graham-Harrison (2024)). Enquanto Israel caracterizava os ataques como focados em militantes do Hamas, diversos organismos internacionais os caracterizaram como indiscriminados e, portanto, genocidas.

Este artigo visa contribuir para o esclarecimento dessa questão: visto que processos diferentes de produzir mortes geram distribuições distintas em termos de idade e gênero das vítimas, é possível estimar a contribuição de cada processo comparando modelos dos resultados esperados com os dados observados. Este trabalho é parte do projeto "Tutorial de Dados" do Núcleo de Tecnologia do MTST 1, que visa combinar o ensino de análise de dados com debates sociais importan-

¹O núcleo é um coletivo responsável por organizar, produzir e desenvolver a parte digital do Movimento dos Trabalhadores Sem Teto (MTST). Entre as suas ações, estão: Instalar wi-fi nas cozinhas solidárias, Criar apps e jogos, Automatizar horta comunitária, Formar profissionais de tecnologia da informação. Saiba mais em: <https://quebra.dev/nucleo-de-tecnologia-do-mtst/>

tes por materiais de estudo abertos. Todas as análises realizadas ficam disponíveis no link do github 2, onde maiores detalhes a respeito deste trabalho podem ser encontrados.

Dados Utilizados

Os dados utilizados foram disponibilizados no site Tech for Palestine³ e produzidos pelo Ministério da Saúde de Gaza a partir de listas de pessoas mortas por ataques israelenses, identificadas em hospitais da região. Na época da realização da análise, os dados cobriam as mortes até 5 de janeiro de 2024 para hospitais que reportam do sul de Gaza e 2 de novembro de 2023 para os que reportam do norte. Os dados são considerados confiáveis (Sawafta and Fick; 2023; Debre; 2023; Jamaluddine et al.; 2023; Khatiba et al.; 2024), contém o nome, documento de identidade, data de nascimento, idade e gênero das pessoas e ignoram pessoas desaparecidas, soterradas em escombros e vítimas de causas indiretas decorrentes da guerra, como a propagação de doenças e falta de comida. Essas causas podem resultar em mais mortes que as diretas Jawad et al. (2020).

Como nossa análise utiliza as idades das pessoas mortas, ignoramos registros sem data de nascimento, tipicamente as associados a pessoas não identificadas. Também ignoramos pessoas com aniversário em 1 de janeiro, uma data comumente utilizada quando a idade da pessoa é estimada. Esse processo reduziu o número de registros de 14.140 para 11.795 pessoas. Este trabalho também utilizou, de maneira auxiliar, dados da distribuição da população por idade i e gênero g – isto é, a pirâmide etária – de Gaza, $I(i, g)$, extraídas do gráfico disponível no CIA World Factbook de 2023⁴.

²<https://github.com/tecMTST/tutorial-de-dados>

³<https://data.techforpalestine.org/docs/killed-in-gaza/>

Modelo para a distribuição de mortos

Assumimos que as mortes contabilizadas nos dados decorrem de duas modalidades de ataques:

1. Os indiscriminados, cuja probabilidade de causar mortes é igual para qualquer idade e gênero. Esse tipo de ataque pode ser realizado com bombas do tipo bunker buster lançadas em áreas residenciais densas, e existem evidências da utilização dessas bombas em Gaza Qiblawi et al. (2023). Como essa modalidade atinge da mesma maneira toda a população, a distribuição de mortos resultante é proporcional à distribuição da população de Gaza em termos de idade e gênero.
2. Os focados, que se direcionam a possíveis combatentes do Hamas, com igual probabilidade quaisquer que sejam suas idades e gêneros. Nesse caso, a distribuição de mortes seria proporcional à distribuição dos combatentes.

Organizações militares costumam ser compostas, majoritariamente, por jovens adultos homens. Um levantamento dos ataques a faca realizados pelo Hamas entre 2015 e 2016 indicou que 27 dos 170 ataques (15,9%) foram realizados por mulheres Margolin (2019). Ademais, nas forças armadas estadunidenses, por exemplo, 82,7% de seus membros ativos são homens, e a distribuição dos membros por idade segue a distribuição apresentada na Figura 2.41 em Departamento de Defesa dos EUA(2021).

⁴<https://www.cia.gov/the-word-factbook/countries/gaza-strip>

Este último fato nos levou a modelar a distribuição de idade i dos combatentes do Hamas com uma distribuição log normal deslocada $L(i)$, que possui características qualitativamente semelhantes ao esperado e é controlada por 3 parâmetros λ , μ e σ :

$$L(i) = \frac{1}{\sqrt{2\pi\sigma(i + \lambda)}} \exp \left[-\frac{(\ln(i + 1) - \mu)^2}{2\sigma^2} \right] \quad (1)$$

O parâmetro λ translada a distribuição para idades maiores ou menores, enquanto os demais parâmetros controlam sua largura e grau de assimetria.

A partir das considerações acima, o modelo para o número esperado de mortes decorrentes de ataques israelenses em Gaza em função da idade i e do gênero g , $M(i, g)$, segue a fórmula:

$$M(i, g) = M_{\text{tot}} [\phi I(i, g) + (1 - \phi) [\alpha \delta_{g,m} + (1 - \alpha) \delta_{g,h}] L(i)] \quad (2)$$

Na equação acima, $M_{\text{tot}} = 11.795$ é o total de mortes ocorridas, é a fração de mortes que são decorrentes de ataques indiscriminados, α é a fração dos combatentes do Hamas que são mulheres, $\delta_{ij} = 1$ se $i = j$ e 0 caso contrário, e g pode ser m (mulher) ou h (homem). O modelo, portanto, depende de 5 parâmetros livres: ϕ , α , λ , μ e σ . O parâmetro que mais nos interessa estimar é ϕ ,

Ajuste do modelo de dados

Utilizamos o método de Monte Carlo com Cadeias de Markov (MCMC), implementado pelo pacote pyMC, para ajustar o modelo aos dados e estimar a distribuição a posteriori dos parâmetros do modelo. A Figura 1 compara a distribuição das mortes observadas (barras) com o esperado pelo modelo de melhor ajuste. As linhas tracejadas e pontilhadas representam, respectivamente, as distribuições decorrentes das modalidades de ataque indiscriminado e focado. A soma das duas contribuições resulta nas curvas sólidas. O modelo representa razoavelmente bem os dados, embora possamos notar alguns desvios mais severos, como a superestimação das mortes de mulheres entre 15 e 23 anos.

Conclusões

O resultado obtido fortalece a avaliação de que Israel está praticando um genocídio em Gaza, dado que a ampla maioria das mortes causadas diretamente por seus ataques são indiscriminadas. Estimamos que 9.736 ± 104 das 11.795 mortes registradas decorreram desse tipo de ataque.

Referências

Debre, I. (2023). What is gaza's ministry of health and how does it calculate the war's death toll?, Associated Press . URL: <https://archive.ph/4v1dG>

Departamento de Defesa dos EUA (2021). 2021 demographics: profile of the military community, <https://archive.ph/TH9Za>. Accessed: 2024-10-22.

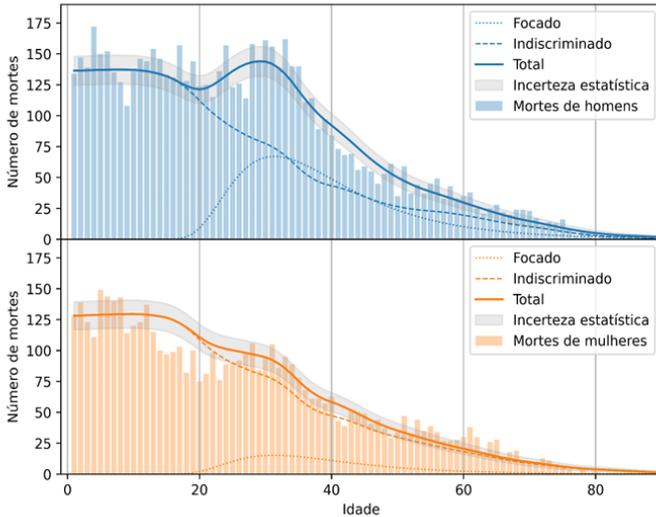


Figura 1: Comparação entre histograma das mortes observadas por idade e gênero e o esperado pelo modelo com os parâmetros de melhor ajuste.

Graham-Harrison, E. (2024). Gaza publishes identities of 34,344 palestinians killed in war with israel, The Guardian Accessed: 2024-10-22.

Jamaluddine, Z., Checchi, F. and Campbell, O. M. R. (2023). Excess mortality in gaza: Oct 7–26, 2023, The Lancet 402(10418): 2189–2190. Accessed: 2024-10-22.

Jawad, M., Hone, T., Vamos, E. P., Roderick, P., Sullivan, R. and Millett, C. (2020). Estimating indirect mortality impacts of armed conflict in civilian populations: panel regression analyses of 193 countries, 1990–2017, BMC medicine 18: 1–11.

Khatiba, R., McKee, M. and Yusuf, S. (2024). Counting the dead in gaza: difficult but essential, *The Lancet* 404(10449): 237–238. Accessed: 2024-10-22.

Margolin, D. (2019). The changing roles of women in violent islamist groups, *Perspectives on the future of women, gender, & violent extremism* pp. 40–49.

Qiblawi, T., Goodwin, A., Mezzofiore, G. and Elbagir, N. (2023). Israel lançou centenas de bombas de 907 kg em gaza, volume não visto desde a guerra do vietnã, mostra análise, <https://archive.ph/L9ByL>. Accessed: 2024-10-22.

Sawafta, A. and Fick, M. (2023). How many palestinians have died in gaza? death toll explained, <https://archive.ph/IBUNu>. Accessed: 2024-10-22.

2.6 Observatório do Santo Amaro- Uma Iniciativa Crucial para a Comunicação Social e Inclusão Digital nas Favelas



Autor

Lucas da Silva
Alves Pessoa

lucasdasilvaalvespessoa@gmail.com

Observatório do Santo Amaro

Mini-Bio

Lucas da Silva Alves Pessoa, 30 anos, é cria do Morro Santo Amaro, mercadólogo, bacharel em administração e analista de dados. Presidente da organização Ame o Santo Amaro, também atua como Advocacy internacional de direitos humanos, dedicando-se à melhoria das condições de vida em sua comunidade e além.

Resumo do Artigo

Este artigo discute a relevância do Observatório do Morro Santo Amaro, uma iniciativa voltada para a coleta e análise de dados sociais e econômicos sobre a comunidade local.

Criado em 2023 e formalizado como um dos Núcleos do projeto Ame o Santo Amaro, o Observatório utiliza abordagens qualitativas e quantitativas para mapear a realidade do território.

Além de levantar essas informações, o projeto busca sensibilizar os moradores sobre a importância desses dados para o desenvolvimento local e a construção de soluções coletivas.

O artigo explora como o observatório incentiva o engajamento comunitário e fortalece as capacidades locais para a geração de conhecimento.

Introdução

O Observatório do Morro Santo Amaro foi criado em 2023 como uma resposta à escassez de informações sobre a comunidade do Morro Santo Amaro, localizado no Rio de Janeiro. Ele se institucionalizou como um dos Núcleos do projeto Ame o Santo Amaro, com o objetivo de levantar dados sobre a realidade social, econômica e cultural da favela. O incentivo para sua criação veio da falta de dados de todas as naturezas sobre a comunidade, o que dificultava a implementação de políticas públicas e ações sociais eficientes. Este artigo explora o papel do observatório na inclusão digital e no fortalecimento da comunicação social nas favelas.

Tecnologias Emergentes e o Acesso nas Favelas

Apesar do crescimento das tecnologias emergentes, a falta de acesso nas favelas ainda é uma realidade marcante. O Observatório do Santo Amaro trabalha para mitigar esse problema ao implementar estratégias de inclusão digital. Além da coleta de dados, a iniciativa busca aproximar os moradores da tecnologia, facilitando o acesso a dispositivos e promovendo a alfabetização digital. Dessa forma, os residentes têm a oportunidade de se inserir de maneira mais ativa no mundo tecnológico, utilizando essas ferramentas tanto para fins pessoais quanto profissionais.

Utilizando métodos considerados obsoletos, como flyers, colagens em postes e o famoso boca a boca, buscamos integrar os moradores com as novas tecnologias sem dispensar as antigas. Essa abordagem reconhece que, embora as tecnologias modernas sejam essenciais, as formas tradicionais de comunicação ainda desempenham um papel vital na conscientização da comunidade.

As tecnologias não apenas demoram a chegar, mas também são dificultadas pelo baixo poder aquisitivo que nossa comunidade possui para acessá-las, além da falta de fomento para a educação tecnológica nas favelas. Essa combinação de fatores inviabiliza aprendizados e, conseqüentemente, a inclusão social. É fundamental que iniciativas como o Observatório continuem a trabalhar para superar essas barreiras e promover um ambiente mais inclusivo e equitativo para todos os moradores.

Redes Comunitárias: Conectividade para todos

Para superar as barreiras de acesso à internet nas favelas, o Observatório também incentiva a criação de redes comunitárias, que permitam uma conectividade mais acessível e de qualidade. Essas redes, geridas pela própria comunidade, proporcionam a infraestrutura necessária para que mais moradores possam se conectar, especialmente em áreas que não são cobertas por grandes provedores de internet. Isso não só garante o acesso às informações, mas também fortalece a colaboração entre os próprios moradores.

Além disso, liberamos nosso acesso à internet para todos os moradores, garantindo que, mesmo aqueles que não têm conexão em casa, possam se conectar de alguma maneira. Essa iniciativa busca criar um ambiente mais inclusivo, onde a falta de acesso não seja um obstáculo para a participação na vida comunitária e no uso das tecnologias disponíveis. Dessa forma, promovemos a autonomia dos moradores e ampliamos as oportunidades de aprendizado e interação digital, fundamentais para o desenvolvimento social e econômico da comunidade.

Conectando-nos com moradores de outras favelas que também trabalham com dados,

pesquisas e tecnologia, esperamos criar uma rede de colaboração, promovendo trocas de tecnologia territorializadas e parcerias. Essa articulação permitirá que possamos implementar nosso trabalho em outras comunidades e vice-versa, fortalecendo a capacidade coletiva de enfrentar desafios e buscar soluções inovadoras que beneficiem todos os envolvidos.

Inclusão Digital e a Aproximação tecnológica

O Observatório do Santo Amaro desempenha um papel crucial na inclusão digital ao proporcionar treinamento e conscientização sobre o uso de tecnologias.

Ao incluir os moradores nos processos de coleta e análise de dados, o observatório cria um ambiente de participação ativa, onde as pessoas se tornam protagonistas da transformação digital de suas comunidades.

Oficinas de capacitação e cursos sobre tecnologia fazem parte das iniciativas do projeto, permitindo que os moradores desenvolvam habilidades cada vez mais exigidas no mercado de trabalho.

A Importância da publicização dos dados e da conscientização comunitária

Além de levantar dados (do Santo Amaro (2024a); do Santo Amaro (2024b); Observatório do Santo Amaro (2024), o Observatório se compromete a tornar essas informações acessíveis a diversos públicos, desde pesquisadores e gestores públicos até os próprios moradores.

A publicização desses dados é feita de forma transparente, buscando sempre conscientizar a comunidade sobre a importância dessas informações para o planejamento de melhorias locais.

Os dados reunidos não apenas auxiliam na implementação de políticas públicas, mas também permitem que a própria comunidade entenda melhor suas necessidades e desafios, facilitando a busca por soluções coletivas.

Alfabetização Digital

A alfabetização digital é um dos pilares fundamentais promovidos pelo Observatório do Morro Santo Amaro. Compreendendo que o domínio das ferramentas tecnológicas é essencial para o desenvolvimento pessoal e comunitário, o

Observatório investe na capacitação dos moradores para que possam utilizar tecnologias de forma eficiente e crítica.

Por meio de oficinas, cursos e treinamentos, o objetivo é garantir que os residentes adquiram as competências necessárias para navegar no ambiente digital, realizar tarefas cotidianas e participar de processos que envolvem coleta e análise de dados.

Esse esforço de alfabetização também visa preparar os moradores para o mercado de trabalho, onde o uso da tecnologia se torna cada vez mais indispensável.

Além disso, ao fomentar o conhecimento sobre o uso dessas ferramentas, o Observatório estimula uma participação mais ativa dos indivíduos nas decisões e discussões sobre o futuro da própria comunidade.

Conclusão

O Observatório do Morro Santo Amaro se destaca como uma iniciativa fundamental para o fortalecimento da comunicação social e inclusão digital nas favelas. Ao levantar dados precisos sobre a comunidade, fomentar o uso consciente da tecnologia e promover a participação ativa dos moradores, o observatório contribui para a criação de uma sociedade mais conectada e informada.

Além disso, realizamos pesquisas com os beneficiados da organização para implementar ações assertivas que tenham impacto de curto, médio e longo prazo.

Essa abordagem nos permite entender melhor o perfil do morador assistido pela organização e identificar suas necessidades específicas, garantindo que a ajuda oferecida seja efetiva e direcionada.

A replicação desse modelo em outras favelas pode ajudar a amplificar a voz dessas comunidades, promovendo um futuro mais inclusivo e digitalmente integrado.

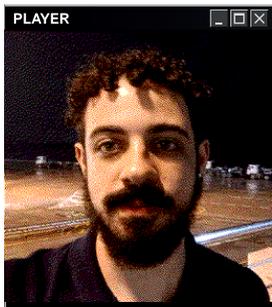
Referências

do Santo Amaro, O. (2024a). Relatório anual núcleo ame elas: Organização ame o santo amaro, <https://github.com/QuebraDev/cts2/blob/main/articles/Relat%C3%B3rio%20Ame%20Elas%202024.pdf>. Accessed: 2024-10-26.

do Santo Amaro, O. (2024b). Sistematização anual de dados: Coletados pelo formulário de cadastro do projeto socieal ame o santo amaro, <https://github.com/QuebraDev/cts2/blob/main/articles/Dados%20Santo%20Amaro%202023.pdf>. Accessed: 2024-10-26.

Observatório do Santo Amaro (2024). Acesso, qualidade e consumo de água no morro santo amaro: Sistematização de dados, <https://github.com/QuebraDev/cts2/blob/main/articles/Dados%20da%20A%CC%81gua%20Santo%20Amaro%202024.pdf>. Accessed: 2024-10-26.

2.7 TBC-DTW: Medição automática de viés midiático baseada em modelagem de tópicos



Autor

Daniel Pereira
Zitei

daniel.zitei@gmail.com
*Laboratório de inteligência
Computacional- Instituto de
Ciências Matemática e de
Computação Universidade de São
Paulo (ICMC-USP)*

Mini-Bio

Formado em Sistemas para Internet na FATEC, estudante de PHD no ICMC-USP, militante da UJC a juventude do Partido Comunista Brasileiro, cientista de dados, tendo atuado em projetos no porto de Santos, B3 e atualmente funcionário da BRQ prestando serviço para a RD Saúde atuando na área de plataforma de ciência de dados. Estudo em meu doutorado os movimentos ideológicos propagados pelos jornais digitais e impressos, com o objetivo de compreender a dinâmica do capitalismo brasileiro exposto nos jornais.

Resumo

Acrescente presença de notícias falsas, incompletas ou incorretas levanta preocupações no meio acadêmico. Esta pesquisa tem como objetivo desenvolver e avaliar uma abordagem não supervisionada para identificar viés na mídia em língua portuguesa. Para isso, são abordadas duas questões-chave.

Primeiro, como organizar artigos de notícias em tópicos usando métodos não supervisionados, incorporando a taxonomia do International Press Telecommunications Council (IPTC) para criar uma nova abordagem de modelagem de tópicos.

Em segundo lugar, como quantificar o viés entre diferentes veículos de comunicação, com base em características extraídas do conteúdo de artigos de notícias.

O estudo utiliza técnicas avançadas como Sentence BERT, BERTopic e Dynamic Topic Modeling para atingir esses objetivos.

A métrica TBC-DTW (Dynamic Time Warping) é introduzida como uma abordagem inovadora para medir o viés na mídia ao longo do tempo. Os resultados indicam o potencial dessa abordagem para compreender e comparar o viés na mídia em diferentes fontes.

Introdução

Notícias falsas, incompletas ou incorretas vêm cada vez mais tomando o discurso sob diversas frentes dentro da academia. Realizando pesquisas no Google Scholar com as seguintes palavras-chave: notícias falsas jornalismo (FALSAS), notícias incompletas jornalismo (INCOMPLETAS), notícias incorretas jornalismo (INCORRETAS) temos os resultados apresentados na Tabela 1.

Tabela 1: Quantidade de artigos catalogados no Google Scholar

Entrada	QTD FAQM*	QTD FD2019**	QTD FD2019 / QTD FAQM
FALSAS	21.100	15.400	72,99%
INCOMPLETAS	15.700	8.320	52,99%
INCORRETAS	11.400	3.960	34,73%
TOTAL	48.200	27.680	57,42%

Pode-se observar que 57,42% de artigos científicos publicados sobre os três assuntos são dos últimos 4 anos. Essas publicações vêm aumentando de produção junto com ideias de que a disseminação de notícias falsas.

Um dos grandes estudos da área foi o livro *Comparing Media Systems* Hallin (2004), no qual eles trazem 4 dimensões de análises dos veículos de comunicação em 18 países do Ocidente, sendo elas: as estruturas do mercado de notícias, o nível de profissionalização do jornalismo, o papel do Estado na intervenção da mídia e o paralelismo político.

Os autores descrevem que cada uma dessas dimensões possui formas de quantificar atributos, as quais são úteis no processo de entendimento de como a mídia se comporta em cada país. O livro nos aponta direções importantes para o entendimento do comportamento da mídia, mas pode possuir viés humano nas quantificações, pois

todas as análises foram coletadas e inferidas pelos próprios pesquisadores.

A proposta geral deste artigo é a investigação de técnicas de aprendizado de máquina não supervisionadas para identificar tópicos em diferentes meios de comunicação e, em seguida, quantificar o viés e a diferença em relação aos meios de comunicação.

O viés é quantificado de diferentes formas, por exemplo, identificando diferenças de conteúdo, vocabulário e expressões de vários meios para um mesmo tópico ou evento.

As abordagens existentes exploram especialistas de domínio ou modelos de aprendizado de máquina supervisionados que, por sua vez, também são criticados por incorporarem viés na análise.

Objetivos e Questões de Pesquisa

O objetivo geral deste projeto é desenvolver e avaliar uma abordagem não supervisionada para a identificação de viés em meios de comunicação na língua portuguesa. Para atingir o objetivo geral, são levantadas as seguintes questões de pesquisa:

- **Como organizar de forma não supervisionada um conjunto de notícias em tópicos?**

Embora existam diferentes métodos de modelagem de tópicos que atuam de forma não supervisionada, é reconhecido que as variações de seus parâmetros podem ocasionar grandes diferenças na geração dos tópicos.

Para tal, é proposto restringir o modelo de geração de tópicos a partir de um modelo de referência conhecido como IPTC (International Press Telecommunications Council), que apresenta uma taxonomia de aproximadamente 1300 temas para notícias ao redor do mundo.

Além de fornecer um vocabulário controlado para os tópicos, também contém uma organização hierarquizada para analisar em diferentes níveis de granularidade.

No entanto, o IPTC não foi proposto no contexto de modelagem de tópicos. Assim, um desafio de pesquisa deste projeto é adaptar um método de modelagem de tópicos para incorporar o conhecimento da taxonomia do IPTC.

Espera-se que, ao responder essa questão de pesquisa, seja disponibilizado um novo método para modelagem de tópicos em notícias, sem a necessidade de anotação por parte de especialistas, mas respeitando a taxonomia do modelo do IPTC.

- **Como quantificar viés entre diferentes meios de comunicação?**

A quantificação de viés é um desafio em aberto e pouco explorado por meio de métodos não supervisionados, principalmente na língua portuguesa. Nesse projeto, espera-se extrair características que cada meio de comunicação utiliza para comunicar uma notícia em um determinado tópico. A extração dessas características será baseada em modelos de linguagem. A ideia geral é quantificar a relação entre palavras utilizadas dentro de um mesmo tópico por diferentes meios de comunicação, bem como explorar características estatísticas dessas palavras. Assim, é uma estratégia de quantificação baseada em conteúdo e que não depende de especialistas para anotação de dados.

Desenvolvimento

A Figura 1 ilustra o método de análise envolvendo todas as etapas, incluindo a ingestão dos dados, criação de um modelo de classificação de textos,

predição dos textos ingeridos e medição do viés, os quais serão descritos com mais detalhes nos capítulos seguintes.

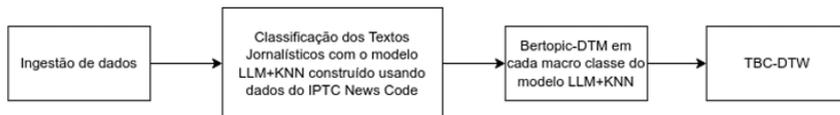


Figura 2: Esteira de desenvolvimento do trabalho.

Jornais

Para iniciar o processo de ingestão de notícias, foi realizado um levantamento dos três principais jornais do país: O Globo, Estadão e Folha de SP. Para realizar o teste do método, foi feita a busca no intervalo de tempo de 01/02/2018 até 31/12/2022. Cada veículo possui suas próprias formas de exibir notícias em HTML e pode bloquear requisições HTTP diretas.

Para contornar esses obstáculos, o projeto utiliza a biblioteca Selenium, que permite realizar chamadas por meio de um navegador, possibilitando varrer os jornais e coletar as notícias de forma automatizada.

Os detalhes das notícias extraídas podem ser armazenados em um formato adequado ao projeto, como um banco de dados, arquivos estruturados ou qualquer outro meio de armazenamento apropriado. A fim de padronizar a forma de busca das notícias nos jornais, o schema pode ser interpretado da seguinte forma:

- **ID URL NOTÍCIA:** URL do site da notícia;
- **DATA:** Data da notícia;
- **TÍTULO:** Título da notícia;
- **SUBTÍTULO:** Subtítulo da notícia;
- **CORPO:** Corpo da notícia;
- **VEÍCULO:** Nome do veículo da notícia

Classificação dos Textos

O primeiro passo em nossa abordagem é treinar o Modelo Fraco (WM) com base no algoritmo dos k-vizinhos mais próximos (k-NN).

Dado um conjunto de dados pré-processado, composto por amostras de texto x_1, x_2, \dots, x_n e suas respectivas etiquetas de classe y_1, y_2, \dots, y_n , nosso objetivo é usar uma função de incorporação $f: x \rightarrow R$ que mapeia cada texto x_i para um vetor d-dimensional que captura sua informação semântica.

Para capturar os dados para criar os modelos de classificação foi utilizado o IPTC Media Topic NewsCodes. IPTC Media Topic NewsCodes, também conhecido como IPTC News Codes, é um sistema de classificação e categorização usado para rotular e organizar notícias e conteúdos jornalísticos.

Desenvolvido pelo International Press Telecommunications Council (IPTC), esse sistema permite que as notícias sejam categorizadas de forma padronizada e consistente em diferentes plataformas e organizações de mídia. Os NewsCodes do IPTC são uma lista hierárquica de categorias e subcategorias que abrangem uma ampla gama de tópicos, temas e assuntos.

Cada código representa uma categoria específica, como política, esportes, entretenimento, saúde, economia, ciência, tecnologia, meio ambiente, entre outros. Essa estrutura hierárquica permite que as notícias sejam classificadas em níveis gerais e específicos, fornecendo uma granularidade adequada para a categorização.

O IPTC disponibiliza a versão desta lista em vários idiomas, como ilustrado na Figura 2.

IPTC Media Topic NewsCodes as of 2023-03-31 (language: pt-PT)

Expand All Collapse All View language versions: Arabic | Chinese (Simplified) | Danish | English | Español | Français | German | Norwegian | 한국어 | Portuguese (Brazil) | Suomi | Svenska

Show related terms

MediaTopic ID	Nome (pt-PT)	Definição (pt-PT)
medtop_0000000	ambiente	Todos os aspetos relativos à proteção, estragos e preservação do ecossistema do planeta Terra e do seu ambiente.
medtop_0000048	alterações climáticas	Mudança significativa nas condições médias relativas ao clima (como temperatura, precipitação ou vento), que se mantem por um período longo.
medtop_0000042	conservação	Preservação de áreas selvagens, flora e fauna, incluindo extinção de espécies.
medtop_0000044	naturaleza	O mundo natural na sua totalidade.
medtop_0000045	poluição ambiental	Contaminação de ar, água, terra, etc., por substâncias nocivas.
medtop_0000040	recursos naturais	Quêntos ambientes naturais à disposição de recursos naturais.
medtop_0000000	artes, cultura, entretenimento e mídia	Todos as formas de arte, entretenimento, herança cultural e mídia.
medtop_0000002	arte e entretenimento	Todos as formas de arte e entretenimento.
medtop_0000008	cultura	Idiomas, costumes, artes, tradições de um grupo em particular.
medtop_0000049	mídia de comunicação	Método de comunicação que se destinam a um grande público, como a televisão, o rádio, as revistas, os jornais, etc.
medtop_0000000	ciência e tecnologia	Todos os aspetos relacionados com a interpretação humana da natureza e do mundo físico e com o desenvolvimento e aplicação desse conhecimento.
medtop_0000000	conflito, guerra e paz	Atos de violência ou de violência, cujos motivos podem ser sociais ou políticos, atividades militares, conflitos geopolíticos, estragos de resolução de conflitos.
medtop_0000000	crime, lei e justiça	Estabelecimento e/ou declaração das regras de comportamento em sociedade, aplicação dessas regras, quebra das regras e penas para os infratores, organizações e práticas envolvidas nessas atividades.
medtop_0000000	desastres, acidentes e emergências	Desastres naturais ou provocados pelo homem, de que resulta perda de vidas ou ferimentos de milhares de pessoas ou danos a objetos ou propriedades.
medtop_0000000	desporto	Exercício competitivo envolvendo esforço físico. Organizações e corpos envolvidos nessas atividades.
medtop_0000000	economia, negócios e finanças	Assuntos relativos à planificação, produção e troca de riqueza.
medtop_0000000	educação	Tudo o que contribua para o alargamento do conhecimento dos indivíduos, desde o nascimento até à morte.
medtop_0000000	estilo de vida e lazer	Atividades realizadas por grupos para relaxamento ou divertimento, tais de emergo, incluindo online e jogos.
medtop_0000000	interesse humano	Temas sobre indivíduos, grupos, animais, plantas ou outros objetos, com foco em aspetos emocionais.
medtop_0000000	meteorologia	Estudo, relatório e previsão de fenómenos meteorológicos.
medtop_0000000	política	Exercício do poder numa região; sistema e interações do país/países e relações entre governos e estados.
medtop_0000000	religião e crença	Todos os aspetos da existência humana que envolvem teologia, filosofia, ética e espiritualidade.

Figura 3: NewsCodes IPTC.

Na etapa seguinte, aproveita-se a saída do Modelo Fraco (WM) para gerar um prompt específico para cada texto, que servirá como entrada para o Modelo de Linguagem de Grande Escala (LLM) na classificação final.

O prompt é cuidadosamente construído para incorporar informações relevantes das k classes identificadas pelo WM. Isso pode incluir etiquetas de classe, descrições textuais ou outros recursos informativos contextualmente relevantes que auxiliem o LLM na geração de previsões precisas.

A geração do prompt é realizada por meio de uma abordagem de aprendizado de prompt few-shot.

Bertopic-DTM

Ao trabalhar com os termos atribuídos a cada texto jornalístico, realiza-se uma seleção no macro tópico do IPTC, como o exemplo 11000000, que se refere à categoria de política.

Em seguida, aplica-se o treinamento de um modelo de tópicos, como o BERTopic. O BERTopic é uma técnica que nos permite realizar a modelagem de tópicos de forma dinâmica, calculando a representação de tópicos em cada intervalo de tempo, sem a necessidade de executar o modelo completo várias vezes.

Para alcançar isso, inicia-se por ajustar o BERTopic como se não houvesse um aspecto temporal nos dados. Isso resulta na criação de um modelo de tópicos geral que captura a estrutura global dos principais tópicos, que podem variar ao longo de diferentes intervalos de tempo. Para cada tópico e intervalo de tempo, calcula-se a representação c-TF IDF (Term Frequency-Inverse Document Frequency personalizado). Essa abordagem nos permite obter uma representação específica dos tópicos em cada intervalo de tempo sem a necessidade de criar agrupamentos a partir das incrustações, uma vez que eles já foram gerados previamente.

TBC-DTW

Nesta seção, discutiremos os resultados do modelo BERTopic DTM, que produziu dois conjuntos de dados temporais representados por df1 e df2. O modelo BERTopic-DTM é uma abordagem avançada que combina tópicos gerados pelo BERTopic com modelagem de séries temporais (DTM- Dynamic 52 Topic Modeling).

Os conjuntos df1 e df2 contêm informações sobre a dinâmica de tópicos em um contexto temporal. df1 e df2 podem ser interpretados como resultados da aplicação do modelo BERTopic-DTM em duas fontes diferentes de dados, como dois jornais, blogs ou sites.

A análise dos resultados inclui a aplicação de métricas de similaridade, como BLEU, para avaliar o quão semelhantes são os tópicos em df1 e df2. Além disso, é realizada a comparação das séries temporais geradas a partir dos tópicos para entender a evolução dos tópicos ao longo do tempo.

A partir dos conjuntos de dados df1 e df2, são calculadas métricas de distância, como a métrica DTW (Dynamic Time Warping), para avaliar a dissimilaridade entre as séries temporais.

A métrica TBC-DTW, que combina a similaridade BLEU e a distância DTW, é aplicada para obter uma métrica global que quantifica a relação entre as duas fontes de dados.

A equação para a métrica TBC-DTW($journal1, journal2$) é definida como:

$$\frac{1}{|T_1|} \sum_{t_1 \in T_1} \frac{1}{T_2(t_1)} \sum_{t_2 \in T_2(t_1)} D(F(df_1, t_1), F(df_2, t_2)) \cdot B(t_1, t_2, n)$$

Onde T_1 representa o conjunto de tópicos em df_1 , $T_2(t_1)$ representa o conjunto de tópicos em df_2 selecionados com base nos $topn$ maiores valores de similaridade BLEU com t_1 , e D é a métrica de distância DTW. O algoritmo de alinhamento temporal dinâmico DTW é usado para comparar e alinhar sequências temporais que podem ter diferentes comprimentos, ritmos ou distorções. Ele encontra um mapeamento ótimo entre os pontos correspondentes das duas sequências, considerando a similaridade entre eles e minimizando a distância acumulada.

Dada a hipótese de que esteja sendo analisado o tópico de saúde em um jornal específico, como o Estadão, o algo ritmo BERTopic nos fornecerá uma lista de palavras-chave que são frequentes nesse tópico ao longo do tempo. Em seguida, compara-se essa lista de palavras-chave com os tópicos relacionados à saúde em outros jornais. Ao realizar essa comparação, pode-se calcular uma medida de similaridade ou alinhamento médio. Isso é feito analisando todas as matrizes de palavras-chave relacionadas ao tópico de saúde em diferentes períodos de tempo e calculando uma média ponderada das correspondências encontradas. O mesmo vale para qualquer tópico do IPTC.

As etapas anteriores forneceram os passos principais para a coleta de textos de várias mídias, pré-processamento e categorização em tópicos de interesse. Também foi apresentado o conceito de DTW (Dynamic Time Warping) para analisar séries temporais. Um diferencial do método proposto neste trabalho é adaptar o DTW para identificação de viés entre diferentes mídias sociais, aqui denominado TBC-DTW (Topic-based Bias Comparison with Dynamic Time Warping).

De forma geral, a proposta envolve comparar tópicos e sua evolução temporal a partir de bases de notícias de duas mídias, com o uso do IPTC e BERTopic. Na série temporal de um tópico, cada observação representa a frequência das palavras do tópico em um intervalo de tempo específico.

Uma etapa importante do DTW é o cálculo da matriz de custo. No TBC-DTW, a matriz de custo é calculada levando em consideração não apenas a distância entre os pontos das séries temporais, mas também a dissimilaridade das palavras de cada tópico em cada período de tempo. Obtida a matriz de custo, o processo de alinhamento tradicional do DTW é aplicado, em que é calculada a matriz de distâncias acumuladas e, em seguida, o caminho de menor custo nesta matriz.

O valor obtido pelo TBC-DTW é uma medida de distância de alinhamento entre as séries temporais das mídias para os tópicos selecionados. Uma distância menor indica um viés mais semelhante entre as mídias, enquanto uma distância maior indica diferenças de viés.

As vantagens da proposta TBC-DTW envolvem tanto uma (1) técnica de comparação objetiva de viés, quanto a (2) tolerância a variações de tempo. Nesse último caso, é possível entender como as mídias reagem a eventos específicos, sem a necessidade de um alinhamento perfeito entre as duas mídias para compará-las.

Os resultados esperados pelo método proposto são, de certa forma, ambiciosos. A ideia é apresentar um dos primeiros métodos objetivos para quantificar o quanto a análise da cobertura jornalística pode ser imparcial sobre as informações divulgadas, quando comparadas entre seus pares.

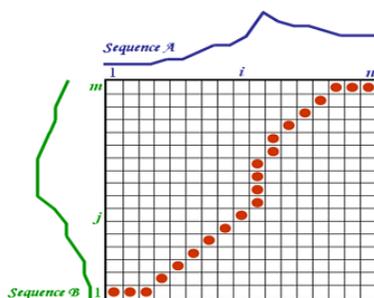


Figura 4: Matriz DTW-Retirada de Sakoe and Chiba (1978).

Essa estratégia também fornece ferramentas para avaliar se vozes e opiniões diversas estão sendo representadas, assim visa promover um discurso mais inclusivo e plural. Por fim, o método é um avanço para apoiar a tomada de decisões informadas, pois ao compreender o viés entre mídias, as pessoas podem tomar decisões informadas considerando como interpretar e ponderar as informações recebidas.

É importante ressaltar também que o uso do algoritmo TBC-DTW proporciona uma análise mais refinada e granular em comparação com abordagens tradicionais de análise de tópicos. Ao levar em consideração a evolução temporal dos tópicos e a correspondência entre eles, podemos obter visões mais detalhadas sobre como os jornais abordam e discutem questões ao longo do tempo.

Essas informações são fundamentais para compreender a dinâmica da cobertura jornalística

e a evolução das narrativas.

Portanto, a aplicação do algoritmo TBC-DTW na análise de alinhamento ideológico na cobertura jornalística é uma forma de examinar e compreender as complexas interações entre a mídia, a ideologia e a sociedade, proporcionando visões mais profundas sobre como as informações são apresentadas, interpretadas e influenciam a percepção pública, em consonância com as perspectivas de Althusser e Gramsci.

Referências

Hallin, D. (2004). *Comparing media systems: Three models of media and politics*, Cambridge University Press.

Sakoe, H. and Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition, *IEEE transactions on acoustics, speech, and signal processing* 26(1): 43–49.

Notas Finais

Para as notas finais desta revista, gostaríamos de agradecer imensamente a todos os que contribuem para o avanço científico e o debate social por um mundo melhor.

Graças às pessoas que se organizam, debatem e praticam a política, hoje podemos vislumbrar novos horizontes e, a partir deles, caminhar rumo a futuros alternativos. A construção desta revista está longe de ser uma realização individual; o QuebraDev jamais teria alcançado o que alcançou sem pessoas que acreditaram e apoiaram o projeto.

Cada episódio, palestra e conversa, cada pessoa que nos apoiou financeiramente ou não, tornou este projeto possível e, graças a isso, milhares de pessoas poderão ouvir o que a periferia constrói com técnica, ciência e política.

Gostaríamos de expressar nossos agradecimentos especiais à empresa Opará-Tecnologia de Guerrilha, que nos ofereceu o espaço necessário para realizarmos nossa conferência.

Agradecemos também imensamente ao Ateliê do Bixiga, que desde o primeiro contato se mostrou totalmente disponível para receber nosso evento.

Nossa eterna gratidão a Andreza Rocha e ao Afroya, um espaço afrofuturista de autonomia criativa voltado ao desenvolvimento de projetos colaborativos que promovam, incentivem e apoiem a inserção e o pertencimento de pessoas negras no ecossistema de tecnologia e inovação. Agradecemos igualmente a cada fornecedor que possibilitou a infraestrutura para a realização desta revista e desta conferência.

Nossa gratidão a cada membro do coletivo que dedicou seu esforço para a construção deste evento: Raquel Queiroz, Kaio Teixeira, Vitor Leonardo, Gustavo Castilião, Fernanda Souza, Sara

Oliveira, Marcos Eptácio, Pedro Ellio e Reginaldo Luna e também todos os voluntários que cederam tempo para trabalhar no evento. O esforço individual e o trabalho em equipe tornaram este projeto realidade mais uma vez; vocês são o efeito colateral que o sistema fez.

Encerramos esta edição com uma alegria imensa e o ânimo de construir novas conferências e seguir debatendo o futuro, sob a perspectiva de nós para nós.

Todo poder à periferia! Que a tecnologia seja a ferramenta que o sistema não use para nos eliminar, mas sim para nos salvar dele!



C:\

Nesta revista, buscamos expor novas ideias sobre como a tecnologia e a sociedade funcionam e se conectam em nossas vidas, e de que maneira podemos utilizá-las para avançar rumo a novos caminhos possíveis. A coletânea de palestras e artigos aqui reunidos explora esses futuros, e esperamos que você, leitor, possa ler e extrair suas próprias reflexões