

Análise Sobre Viés de Gênero no Youtube: um Estudo sobre as Eleições Brasileiras de 2018 e 2022

Andressa Kappaun¹, Jonice Oliveira¹

¹Programa de Pós-Graduação em Informática (PPGI)
Universidade Federal do Rio de Janeiro (UFRJ)
Rio de Janeiro – RJ – Brazil

kappaun@ppgi.ufrj.br, jonice@ic.ufrj.br

Abstract. *This paper proposes a method to investigate the presence of gender bias through engagement and sentiment analysis of comments on YouTube videos containing the names of men or women in its video title or description. The study encompasses the Brazilian elections of 2018 and 2022, focusing on the positions of president, governor, and senator. The quantitative analysis, statistically validated by the chi-square test, reveals higher engagement in videos featuring men, accompanied by a prevalence of negative sentiments in videos featuring women. The method proved effective in identifying gender bias and can be replicated for different scenarios, which can be considered a social contribution due to its humanistic values.*

Resumo. *Este trabalho propôs um método para investigar a presença de viés de gênero através da análise do engajamento e sentimento dos comentários em vídeos do YouTube que mencionam homens ou mulheres no título ou na descrição. Os cenários de estudo abrangem as eleições brasileiras de 2018 e 2022, focando nos cargos de presidente, governador e senador. A análise quantitativa, validada estatisticamente pelo teste de chi-quadrado, revelou um maior engajamento em vídeos com homens, acompanhado por uma predominância de sentimentos negativos nos vídeos com mulheres. O método demonstrou eficácia na detecção de viés de gênero, sendo replicável em diferentes contextos, representando, assim, uma contribuição social em termos de valores humanos.*

1. Introdução

O consumo de conteúdo de vídeo no Brasil foi muito grande em 2020, de acordo com pesquisa realizada pelo Kantar IBOPE [Media 2021], chegando a 99% do total de usuários de internet do país. Este consumo se deu em vários formatos (plataformas de *streaming*, redes sociais, videochamadas, etc.) e telas (televisões, tablets, celulares, etc.). Esta pesquisa também destaca que 80% dos brasileiros consumiram vídeo online gratuito, enquanto que a média mundial é de 65%. Em redes sociais, o consumo de vídeo dos brasileiros segue mais alto que a média mundial, respectivamente, 72% e 57%. Mais da metade dos brasileiros admitiram (58%) que o consumo foi maior após o isolamento social provocado pela pandemia de Covid-19.

Segundo [Social 2023], o Brasil teve um acréscimo de sete milhões de usuários de internet em 2023, em comparação com 2022 (mais 4%). Os autores destacam que o principal motivo de uso da internet para os usuários é a busca por informações. Destes,

quase 80% consomem conteúdos de vídeo no YouTube, sendo que a maioria são mulheres (52%).

Problemas antigos destas redes foram potencializados durante a pandemia, como por exemplo, a propagação de discursos de ódio e crimes de intolerância, principalmente xenofobia, misoginia, racismo, homofobia e transfobia [Stechemesser et al. 2020]. A ineficiência de mecanismos de punição para os que propagam preconceitos ou desinformação facilitou a disseminação de discursos de ódio. As desinformações ou *fake news* são disseminadas nas redes sociais com o objetivo de atingir uma pessoa, ou determinado grupo, logo, “se relacionam impulsionando um ao outro de forma quase simbiótica” [Sirius 2019].

Se há diferenças entre os gêneros no mundo virtual, há também no mundo físico. Em relação à desigualdade de gênero, no Brasil, de acordo com dados divulgados pelo Instituto Brasileiro de Geografia e Estatística (IBGE), indicam que, apesar do percentual de mulheres que completaram a graduação, na faixa etária de 25 a 44 anos de idade, ser de 21,5% e o de homens ser de 15,6%, no mercado de trabalho as mulheres ganham menos. Os rendimentos do trabalho das mulheres correspondem, em média, cerca de 75% do que os homens recebem [IBGE 2018]. Quanto à violência, de acordo com levantamento feito pelo G1¹, o feminicídio aumentou 5% em 2022, quando comparado ao ano anterior. O cenário não é diferente na esfera política: pesquisa realizada pelo GLOBO² (agosto de 2022) mostra que 87,5% das candidatas a cargos majoritários das eleições do mesmo ano sofreram algum tipo de ataque relacionado a gênero.

Desta maneira, urge a necessidade de aprofundar estudos sobre a representatividade de gênero no YouTube, adaptados para a realidade do Brasil. Em cenários recentes de polarização política e ideológica é fundamental investigar se os algoritmos de recomendação da plataforma reforçam os preconceitos e discriminações presentes na sociedade, em especial o machismo, a misoginia e o sexismo. Espera-se que os resultados possam auxiliar no desenvolvimento de políticas públicas no combate à problemática estudada.

Dentre os grandes desafios definidos pela Sociedade Brasileira de Computação (SBC), este trabalho perpassa e contribui com alguns deles. Nos “Grandes Desafios da Computação no Brasil”, acredita-se que há contribuição na “Gestão da Informação em grandes volumes de dados multimídia”, devido ao processamento e classificação de muitos dados relacionados à vídeos.

Quanto aos “Grandes Desafios em Sistemas de Informação no Brasil 2016-2026”, este trabalho perpassa o “Desafio 3: Complexidade dos Sistemas de Informação”, no subtópico de “Interoperabilidade Total”, que, em resumo, consiste na capacidade de aplicações e dados heterogêneos em compartilhar procedimentos para se comunicar, mesmo estando em diferentes equipamentos e plataformas. Esta falta de interoperabilidade restringe as análises a somente uma plataforma de vídeos (YouTube).

¹Reportagem disponível em: <https://g1.globo.com/monitor-da-violencia/noticia/2023/03/08/brasil-bate-recorde-de-feminicidios-em-2022-com-uma-mulher-morta-a-cada-6-horas.ghtml>

²Reportagem disponível em: <https://oglobo.globo.com/politica/noticia/2022/08/de-tentativas-de-mordaca-a-ameacas-de-morte-violencia-politica-contra-mulheres-chega-a-31-registros-em-um-ano-no-mpf.ghtml>

A principal **questão de pesquisa (QP)** deste trabalho é: Nomes de gêneros diferentes no título ou descrição de vídeos resultam em diferentes interações e sentimentos? Desta maneira, o **objetivo principal (OP)** desta pesquisa é “investigar se vídeos com nome de entidades de gêneros diferentes no título têm o mesmo sentimento e engajamento na rede social YouTube, tendo como cenário as eleições brasileiras, em 2018 e 2022”.

2. Trabalhos Relacionados

Com o objetivo de identificar os trabalhos nacionais, tendo em vista que o cenário de análise e atuação desta pesquisa é o Brasil, foi escolhido como principal fonte a biblioteca da SOL³, da Sociedade Brasileira de Computação. Duas strings de busca praticamente iguais foram utilizadas, mudando apenas o idioma (português e inglês):

String de busca em português: (“viés de gênero”OR “sexismo”OR “preconceito”OR “machismo”OR “misoginia”) AND (“redes sociais”OR “youtube”OR “mídias sociais”OR “facebook”OR “twitter”OR instagram OR whatsapp OR tiktok OR streaming)”

String de busca em inglês: “((sexism OR prejudice OR “male chauvinism” OR misogyny) AND (youtube OR “social networks” OR “social media” OR facebook OR twitter OR instagram OR whatsapp OR tiktok))”

O período de busca foi de 01/01/2013 a 12/10/2023. Foram retornados seis artigos no total, sendo um referente aos termos em inglês e cinco em português. Explorando a relevância do tema no cenário brasileiro, a partir dos artigos descobertos foi feito o processo de *snowballing* com objetivo de aumentar a identificação de trabalhos relevantes na área.

Além disso, para ter uma maior amplitude de trabalhos no cenário internacional, também aplicamos a busca de termos em inglês na plataforma IEEE para compor este mapeamento, realizando um processo de filtragem utilizando como critérios de seleção:

- publicação entre 01/01/2013 e 12/10/2023;
- assunto de publicação como *gender issues*, selecionando trabalhos relacionados a questões de gênero;
- publicações em revistas.

Ao final desta etapa, foram encontrados seis artigos na SOL, oito com *snowballing* e quatro na IEEE. Onde apenas quatro trabalhos abordam algum tipo de questão de gênero utilizando dados de YouTube. Apenas um destes trata simultaneamente a problemática de viés de gênero no Brasil dentro da esfera política [Kappaun and Oliveira 2023]. Os outros três trabalhos abordam aspectos mais gerais do tema [O. Plath et al. 2022],[Wotanis and McMillan 2014] e [Wegener et al. 2020].

3. Framework de identificação de viés de gênero

A identificação de viés de gênero em vídeos do YouTube, dentro de qualquer cenário, necessita de alguns passos metodológicos, a saber: (A) descoberta e pré-processamento de dados do cenário de estudo; (B) levantamento e extração de dados de canais relevantes ao cenário de estudo e (C) cruzamento, processamento e análise dos dados (Figura 1).

³Disponível em: <https://sol.sbc.org.br/busca/index.php/integrada/index>

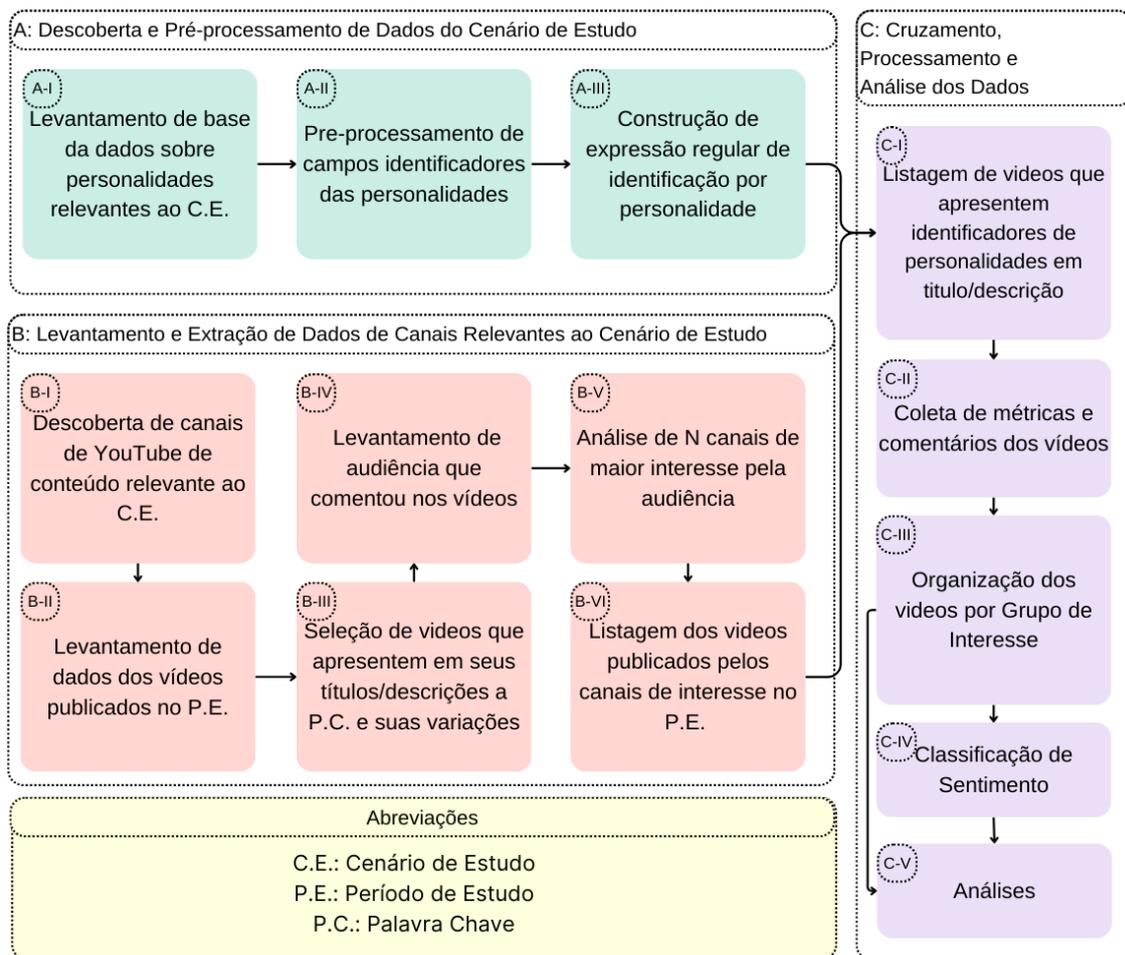


Figura 1. Fluxograma metodológico.

Na fase de descoberta e pré-processamento de dados do cenário de estudo (A), é fundamental a escolha de uma base de dados relevante ao cenário escolhido (A-I), que contenha identificadores nominais como, por exemplo, nome, apelido etc. Após esta etapa, tem-se o pré-processamento de quais campos são identificadores das personalidades dentro deste cenário (A-II). Neste momento, é possível construir a expressão regular de cada personalidade, onde a diversidade de seus nomes pré-processados no passo anterior é agrupada (A-III).

Para o levantamento e extração de dados, é importante uma análise de quais são os canais relevantes para o seu cenário de estudo, sempre tentando evitar grupos amostrais de canais viesados para não comprometer os resultados e análises (B). Recomenda-se partir de fontes neutras (por exemplo: pesquisas e relatórios de órgãos oficiais), para construir uma rotina eficiente de escolha dos canais mais representativos ou relevantes para o seu cenário, considerando o período de estudo (B-I). A partir disso, começa o processo de seleção de vídeos com palavras-chave referentes ao seu cenário para posterior levantamento da audiência destes (B- II a IV). Por fim, faz-se a seleção dos canais que obtiveram maior engajamento pela audiência encontrada no passo anterior (B-V), para a listagem de vídeos publicados pelos mesmos no período de estudo (B-VI).

A última fase contempla o cruzamento, processamento e análise dos dados (C), partindo do cruzamento da expressão regular de cada personalidade gerada em A-III com os vídeos listados em B-VI (C-I). Neste momento inicia-se a coleta de métricas (curtidas, número de comentários totais, entre outros, por exemplo) e amostras de comentários publicados nos vídeos (C-II), para em seguida, organizar os dados obtidos por grupo de interesse (gênero, faixa etária, etnia etc.) (C-III). Por fim, classifica-se o sentimento dos comentários coletados, para analisá-los, juntamente com as métricas (C-IV e V).

3.1. Framework Instanciado: Eleições Brasileiras de 2018 e 2022

Este trabalho aplica o framework proposto (Figura 1) no cenário brasileiro nos anos de 2018 e 2022. O estudo consiste na organização e avaliação dos dados coletados para uma análise quantitativa e estatística baseada nas métricas dos vídeos e na extração de uma amostra de comentários de cada vídeo, sendo no mínimo 5 e no máximo 500 comentários, cada um com até 500 caracteres. No processo foi empregado o teste de chi-quadrado para verificação da hipótese de existência de viés de gênero em vídeos do YouTube publicados pelos canais de interesse levantados e que apresente nomes de candidatos no título ou na descrição do vídeo. Tais vídeos são referentes às eleições de 2018 e 2022, abrangendo as disputas eleitorais dos cargos de: presidente, vice-presidente, governador, vice-governador e senador.

Os dados utilizados neste trabalho são provenientes de duas fontes, a saber: 1) Tribunal Superior Eleitoral⁴ (TSE) e 2) YouTube⁵. Os dados do TSE podem ser baixados no formato CSV, através do seu domínio público. Os dados do YouTube são coletados automaticamente via API, através de código em *Python*. Vale mencionar que todos são gratuitos e disponibilizados pelas suas entidades proprietárias.

Para a escolha de canais de YouTube relevantes para a pesquisa foram utilizados estudos do Ibope sobre audiência de TV aberta no Brasil no ano de 2022⁶. Para cada uma dessas emissoras foram levantados seus principais canais de conteúdo oficial dentro da plataforma YouTube. Os vídeos publicados nos anos de 2018 e 2022 por esses canais, com título que deram match com a expressão regular “*eleições|eleição|elei..o*”, foram considerados como conjunto inicial de vídeos a serem trabalhados, denominado de vídeos de *seed*.

A partir dos vídeos de *seed* e suas respectivas audiências (pessoas que comentaram nos vídeos), é constatado que as mesmas também engajaram com conteúdos de outros canais publicados na plataforma. Priorizando os anos de 2018 e 2022, foi levantado o ranking de canais de maior interesse pela audiência do *seed*. Destes, separamos os 50 primeiros para estudo neste trabalho. A partir dos canais de estudo mapeados é iniciada a extração dos dados de cada canal, vídeos e seus comentários.

Baseando-se nos dados dos candidatos às eleições de 2018 e 2022 disponibilizados no portal oficial do TSE⁷, foram criadas automaticamente expressões regulares para cada

⁴<https://sig.tse.jus.br/ords/dwapt/r/seai/sig-eleicao/home>

⁵<https://developers.google.com/youtube/v3/docs>

⁶<https://www.uol.com.br/splash/noticias/ooops/2022/02/04/veja-o-ranking-de-ibope-da-tv-aberta-redetv-ja-ronda-o-traco.htm>

⁷Disponível em: <https://sig.tse.jus.br/ords/dwapt/seai/r/sig-eleicao-resultados/resultado-da-elei%C3%A7%C3%A3o>

candidato de acordo com o cargo disputado. Ao cruzar expressões regulares com os títulos ou descrições dos vídeos, as correspondências positivas acionam a coleta de métricas dos vídeos, incluindo total de curtidas, visualizações e compartilhamentos. Adicionalmente, é obtida uma amostra dos comentários associados a essas correspondências.

Para melhor sistematização dos dados, optou-se por separá-los em três grupos, a saber: 1) grupo F: vídeos apenas com nome de mulheres candidatas no título; 2) grupo M: vídeos apenas com nome de homens candidatos no título; 3) grupo MF: vídeos que aparecem nomes de homens e de mulheres candidatas simultaneamente.

3.2. Análises

Este trabalho utiliza duas abordagens para o estudo do tema de viés de gênero, a saber: 1) engajamento e 2) análise de sentimento negativo dos vídeos.

Para a análise do engajamento dos vídeos em relação ao seu grupo de interesse, adotou-se como engajamento a soma de todas as interações válidas na plataforma, compreendendo a soma de comentários, respostas a comentários e curtidas nos vídeos. A abordagem analítica baseia-se na comparação entre a média de engajamento de todos os vídeos no cenário estudado e a média específica obtida para cada grupo de interesse.

Quanto à classificação de sentimento dos comentários, o classificador utilizado neste trabalho é o *pySentimiento*, que é disponibilizado na forma de *Toolkit* em Python para análise de sentimentos e processamento de linguagem natural focado em textos de redes sociais. Esta ferramenta foi escolhida por possuir bom desempenho na análise de sentimentos em português-brasileiro, com confiabilidade média (F1) de 75,3 com desvio padrão de 0,5 [Pérez et al. 2021].

Para cada vídeo de interesse é extraída uma amostra de comentários, compreendida em um intervalo de no mínimo 5 e no máximo 500 comentários, cada um com tamanho máximo de 500 caracteres. A classificação de sentimento foi aplicada para todos os comentários, resultando na atribuição de um dicionário para cada comentário. Esses dicionários contêm as probabilidades de classificação em cada domínio, abrangendo as categorias positivo, neutro e negativo. A análise é baseada na comparação entre a proporção de comentários negativos do cenário abordado com a proporção de comentários negativos obtida em cada grupo de interesse.

4. Resultados

Para estudo de viés de gênero no cenário escolhido foi avaliado os dados demográficos das candidaturas em paralelo a duas análises: uma baseada em engajamento e outra baseada na ocorrência de sentimentos negativos.

O cenário eleitoral é analisado a partir da distribuição demográfica de candidatos frente às questões demográficas de gênero e raça, comparando os volumes de candidatos por cargos nos seus anos eleitorais de disputa.

Para análise dos vídeos foram utilizados dois tipos de estudos diferentes referentes ao engajamento e ocorrência de sentimento negativo. Foram aplicadas nos seguintes cenários de estudo, a saber: 1) geral: todos os vídeos coletados no ano eleitoral; 2) cargo disputado: vídeos contendo o nome de um candidato à disputa no ano eleitoral. Em ambos os cenários foi feito um comparativo entre os anos eleitorais abordados (2018 e 2022).

É fundamental esclarecer que, foram considerados neste estudo apenas vídeos divulgados no segundo semestre de cada ano, seguindo o calendário eleitoral correspondente, abrangendo as campanhas de ambos os turnos e a repercussão pós-eleição até o final do ano.

Para validar os resultados encontrados foi aplicado o teste de independência chi-square ou Chi Quadrado de Pearson [Greenwood and Nikulin 1996]. Este teste ajuda a entender se a associação entre duas variáveis categóricas é estatisticamente significativa, ou seja, quando se deseja constatar se existe uma diferença real nos dados baseado em alguma afirmação ou conceito teórico. O teste é aplicado a partir da tabela de contingência, do cálculo do grau de liberdade dos dados, pré-determinação do alfa, chi quadrado tabelado e chi quadrado calculado. Sobre análises são levantadas as seguintes hipóteses:

- H_0 - Engajamento: não existe relação entre a presença de gênero no título/descrição de um vídeo e maior volume de engajamento.
- H_0 - Sentimento Negativo: não existe relação entre a presença de gênero no título/descrição de um vídeo e a ocorrência de comentários de sentimento negativo.

4.1. Análise das candidaturas e cenário eleitoral

Analisando a distribuição de gênero do eleitorado (Tabela 1), observa-se que o volume de eleitoras mulheres é maior nos dois anos, seguindo o perfil da população brasileira que é predominante por mulheres, de acordo com o censo do IBGE⁸.

Tabela 1. Resumo de dados de gênero do eleitorado Nacional e Internacional referente às eleições de 2018 e 2022 (Fonte: TSE).

Ano	Área de Voto	Gênero		
		Feminino	Masculino	Não Informado
2018	Brasil	77.047.366 (52,48%)	69.694.781 (47,47%)	63.401 (0,04%)
	Exterior	292.531 (58,42%)	208.196 (41,58%)	-
2022	Brasil	81.965.109 (52,62%)	73.755.042 (47,35%)	36.782 (0,02%)
	Exterior	408.055 (58,54%)	289.023 (41,46%)	-

Apesar de um maior volume de eleitoras é observada uma baixa representatividade do gênero feminino ao ver a distribuição de homens e mulheres nos cargos abordados. Ainda que o número de candidatas mulheres em 2022 tenha sido aproximadamente 8% maior do que em 2018 (sendo 2018 contando com 162 candidatas e 2022 com 175 candidatas), ainda existe uma discrepância significativa na distribuição de gênero no último ano, onde nenhum dos cargos chega ao menos a 40% de representatividade feminina, como explicitado na figura 2.

4.2. Análise geral comparativa 2018 x 2022

A partir dos dados coletados referentes a 2018 e 2022, destacados na tabela 2, observa-se um volume de vídeos desproporcional entre os grupos de interesse. Em ambos os anos, o grupo M têm grande maioria dos vídeos quando comparado aos grupos F e MF. Essa predominância pode ser atribuída ao fato de ambas as eleições possuem maior volume de

⁸Censo do IBGE, de 2022, disponível em: <https://censo2022.ibge.gov.br/>

Cargo x Proporção de Gênero dos Candidatos

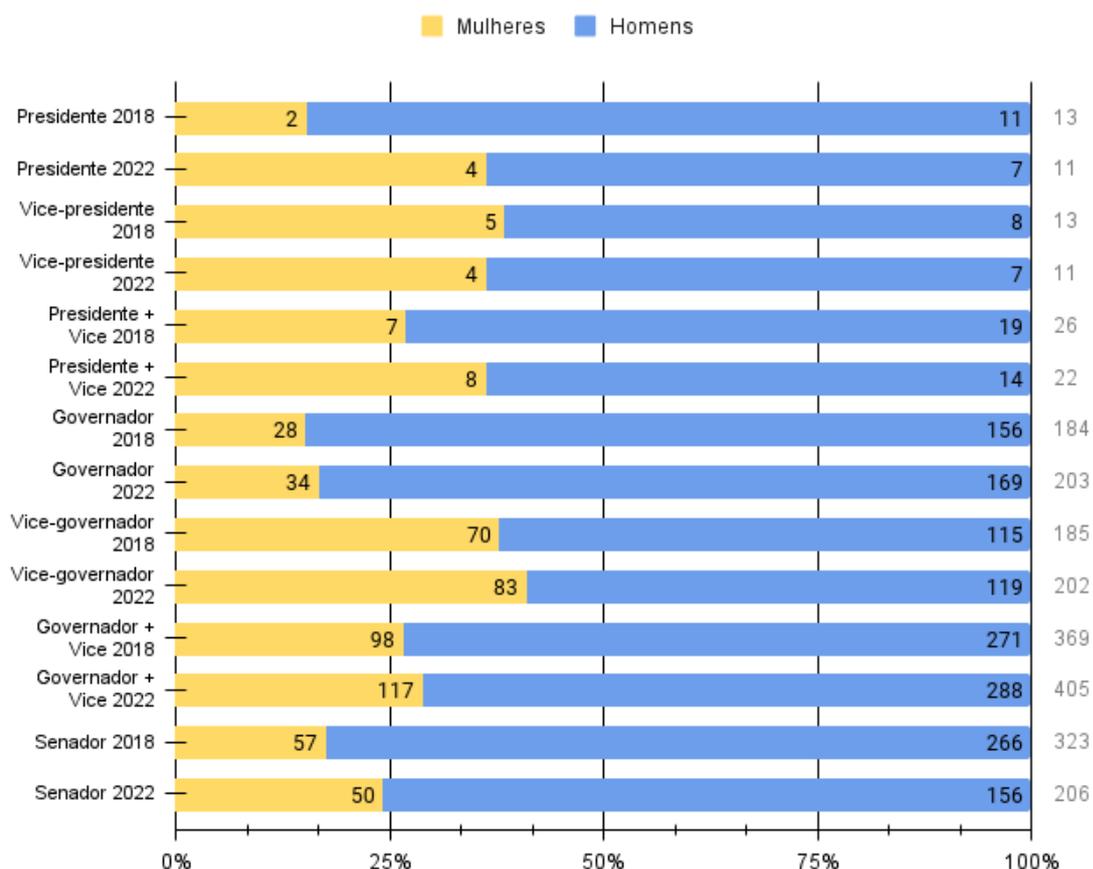


Figura 2. Comparativo entre a proporção de gênero entre os cargos de interesse disputados nas eleições de 2018 e 2022.

candidatos homens em cargos de maior destaque. Há também esta desproporcionalidade entre as datas, onde 2022 tem um volume de vídeos muito maior. Nas demais linhas da tabela 2, separadas por cargo, observa-se que a discrepância observada nos dados gerais da tabela se repete em todos os cargos entre os grupos de interesse.

Tabela 2. Resumo do total de vídeos publicados e vídeos utilizados nas análises divididos por grupo de interesse, ano eleitoral e por cargo disputado.

Tipo de análise	M		F		MF		Total	
	Engajamento	Sent. Neg.						
Cenário analisado								
2018/2 Geral	2.749	2.341	95	39	87	68	2931	2.448

2022/2 Geral	29.269	25.546	618	429	1.376	1.147	31.263	27.122
2018/2 Presidente	2.624	2.282	73	33	78	67	2.775	2.382
2022/2 Presidente	27.792	24.594	467	341	1.337	1.123	29.596	26.058
2018/2 Governador	75	39	1	1	4	2	80	42
2022/2 Governador	1.708	1.212	18	14	87	63	1.813	1.289
2018/2 Senador	164	127	21	5	25	16	210	148
2022/2 Senador	960	749	133	74	108	88	1.201	911

4.2.1. Análise de engajamento

Considerando os vídeos de interesse coletados no segundo semestre dos anos eleitorais, temos no total 34.194 vídeos e seus respectivos engajamentos. No segundo semestre de 2018, foram coletados 2.931 vídeos publicados no período, com uma média de engajamento geral no valor de 6.042. No mesmo período de 2022, foram coletados ao todo 31.263 vídeos com uma média de engajamento geral de 4.183. Em ambos os períodos de análise, o grupo de interesse F obteve engajamento médio por vídeo menor que os outros grupos (M e MF) e menor que a média geral (Figura 3). Isto pode indicar que a presença do nome de um homem em um vídeo aumenta o interesse das pessoas em interagir com o conteúdo.

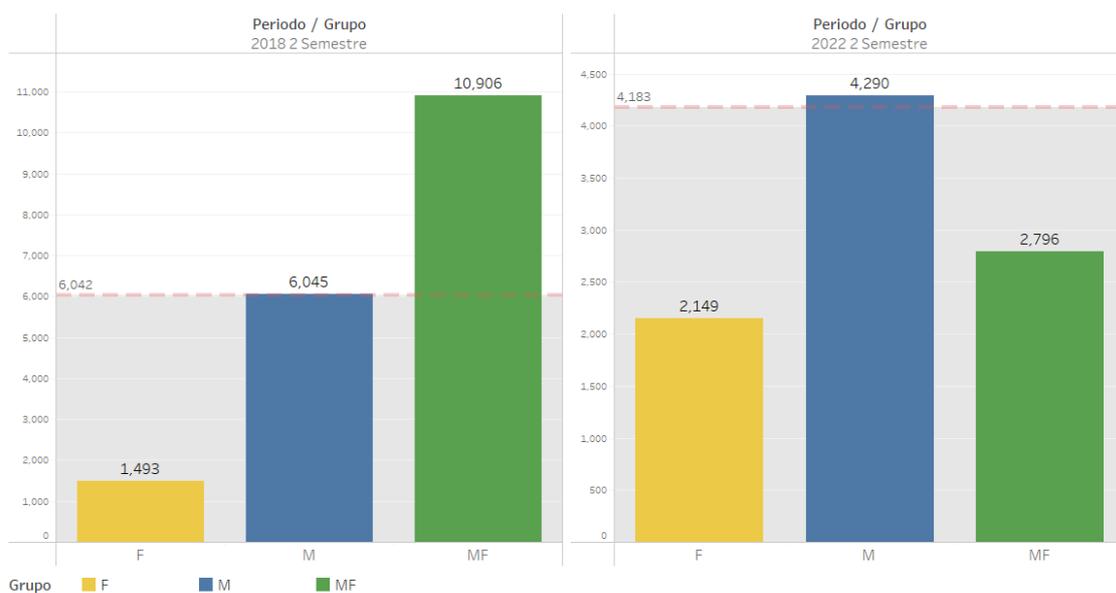


Figura 3. Engajamento médio para cada grupo de interesse por ano de análise.

Aplicando o teste estatístico proposto, H_0 é rejeitada em ambos os cenários, aceitando H_1 que relaciona a ocorrência de maior engajamento em vídeos dos grupos M e MF quando comparado a vídeos do grupo F. Isto demonstra que há maior interesse da audiência em vídeos com a presença de nomes masculinos no nome ou na descrição. A seguir serão detalhadas as análises de sentimentos dos comentários nos vídeos.

4.2.2. Análise de sentimento negativo

Quanto à classificação de sentimento dos comentários, é analisada a proporção de ocorrência de cada tipo de sentimento (Figura 4) para cada ano de eleição estudado. É possível observar que há um predomínio claro dos sentimentos neutros e negativos, (Figura 4), sendo a proporção de ocorrência de comentários negativos similar entre ambos os anos. Apesar do desbalanceamento do número de comentários nos vídeos entre os anos de análise (Tabela 3). É interessante notar que na eleição de 2022, bem mais polarizada, o número de comentários positivos foi bem menor.

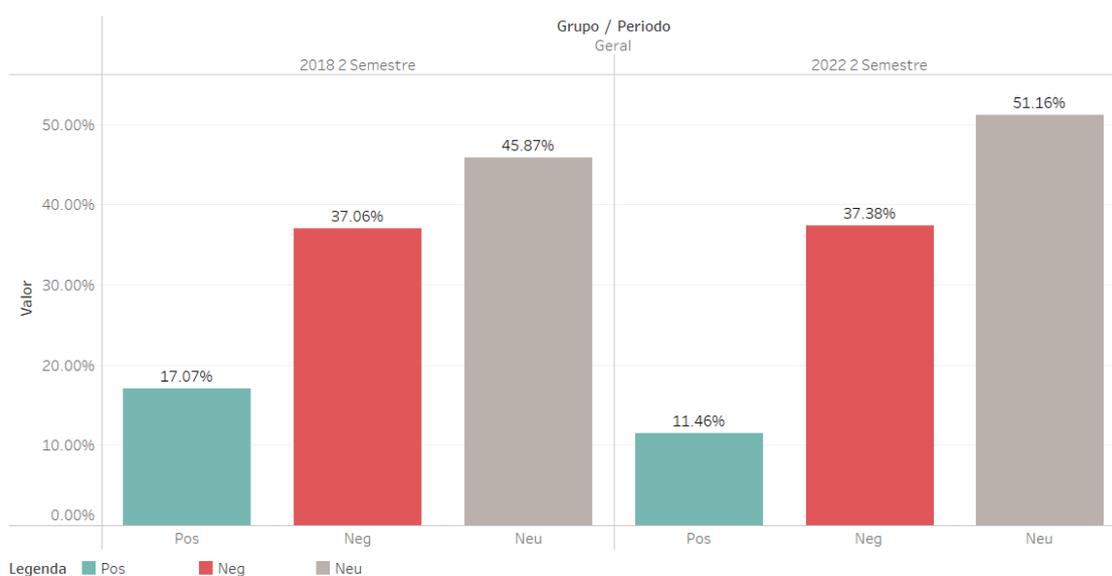


Figura 4. Proporção de sentimento dos comentários por eleição.

Tabela 3. Volume de comentários analisados por grupo de interesse.

	F	M	MF	Total
2018.2	3.393	415.997	16.167	435.557
2022.2	52.080	4.579.419	184.197	4.815.696

Separando estes dados por grupos de interesse e por ano eleitoral, focando apenas nos comentários negativos, observa-se que, em 2018, o grupo F esteve abaixo do grupo M e abaixo da proporção média (37,06%) e, em 2022, acima do grupo M e da proporção média (37,38%) de comentários negativos (figura 5).

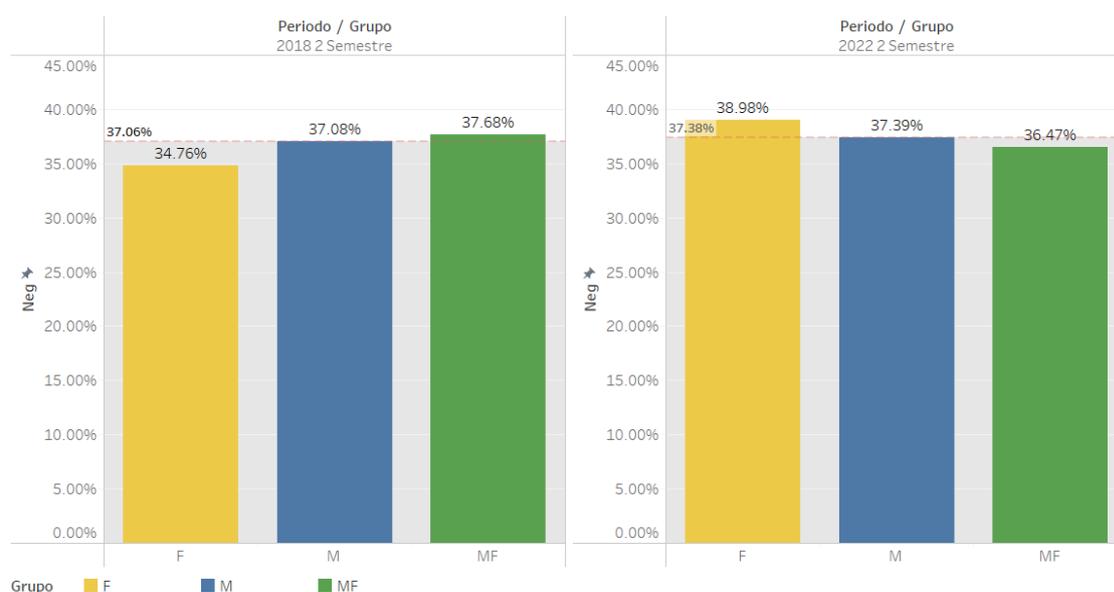


Figura 5. Proporção de sentimento dos comentários por grupo em 2018 e 2022.

A partir dos resultados obtidos, H_0 é rejeitada no período de 2022. Para o período de 2018 o resultado do teste é inconclusivo e não se pode rejeitar H_0 .

É interessante observar que, a partir dos dados referentes ao ano da eleição de 2022, foi possível associar estatisticamente a relação entre a presença apenas de nome feminino com menor engajamento e também maior ocorrência de sentimento negativo. Importante destacar que foi o período de maior número de candidatas mulheres dentre os anos de estudo, ou seja, período de menor disparidade no volume de candidatos de gêneros diferentes, principalmente no cargo de presidente, que possui maior destaque.

4.3. Análise por cargo disputado

A visibilidade e destaque de candidatos estão diretamente ligados à relevância do cargo disputado. Esta seção aborda a análise de dados por cargo, permitindo uma discussão mais detalhada das diferenças entre eles. Dados para presidentes/vice-presidentes e governadores/vice-governadores foram agrupados por formarem uma mesma chapa. Cada grupo de cargos será submetido a análises de engajamento e sentimento negativo.

4.3.1. Presidente e Vice-presidente

Quanto aos vídeos referentes à presidência, foram observados no total 2.775 vídeos em 2018 e 29.596 em 2022. Mulheres somaram ao todo 7 candidatas em 2018 e 8 em 2022, quanto aos homens, somam 19 em 2018 e 14 em 2022. Sendo 2 candidatas titulares à presidência em 2018 e 4 em 2022.

Nos dois anos analisados, observou-se que o engajamento médio do grupo F é inferior ao de todos os outros grupos de interesse. Quanto à ocorrência de sentimentos negativos, verificou-se que, em ambos os anos, cerca de 37% dos comentários nos cenários apresentaram esse tipo de sentimento. Em 2018, o grupo F teve 33.9% de seus

comentários classificados como negativos, enquanto os grupos M e MF ficaram ligeiramente acima da média. Em 2022, o grupo F se destacou com 38.8% do total, superando a média geral de 37.4%. Em ambos os períodos, o grupo M permaneceu próximo à média.

Validando a relação estatística dos dados é possível aceitar rejeitar H_0 de engajamento em ambos os anos, e também para ocorrência de sentimento negativo em 2022.

4.3.2. Governador e Vice-governador

Devido ao reduzido número de vídeos relacionados aos candidatos a governador/vice-governador em 2018, não foi possível identificar uma predominância entre os grupos de interesse. No entanto, em 2022, os testes estatísticos não rejeitam H_0 para as análises de engajamento e ocorrência de sentimento negativo.

4.3.3. Senador

Em ambos os anos foram observados que o engajamento médio do grupo F é menor dentre todos os outros grupos de interesse e menor que a média. Quanto à ocorrência de sentimento negativo, a partir de amostras de comentários dos vídeos relevantes para análise, foi observado que no cenário geral, 43% dos comentários em 2018 tiveram sentimento negativo, enquanto em 2022 foram 37%. Comparando com a proporção de comentários negativos de cada grupo de interesse, é visto que em 2018 o grupo F tem 36.3% dos seus comentários classificados como negativos, número abaixo da média geral. Em 2022, ocorre o contrário, o grupo F tem maior ocorrência de comentários negativos, com 41,76% do total, acima da média. Seguindo o teste estatístico aplicado é considerada válida a relação apenas para a maior ocorrência de sentimento negativo para vídeos sobre senador em 2022. Não rejeitando H_0 nos demais casos.

5. Contribuições

5.1. Sociais

Ao lançar luz sobre uma questão política pouco abordada no Brasil, este estudo revela o viés de gênero presente em vídeos no YouTube, quantificando um menor engajamento e um maior sentimento negativo associado a vídeos com nomes de mulheres no título ou na descrição. Tal constatação tem o potencial de fomentar debates e conscientização entre as pessoas, destacando uma questão frequentemente negligenciada. O método proposto neste trabalho se mostrou eficaz ao evidenciar o viés de gênero no contexto eleitoral e, por se tratar de um *framework* baseado em análise de texto em mídias sociais, pode ser replicado em outros cenários, constituindo, assim, uma contribuição significativa para a sociedade. Em um país extremamente desigual como o Brasil, o acesso às tecnologias e redes sociais, por exemplo, nem sempre vem acompanhado de aspectos humanos, envolvendo também “valores no desenvolvimento e utilização das tecnologias”, muitas vezes desconsiderados ou menos valorizados na concepção e design de sistemas.

5.2. Científicas

Esta pesquisa tange alguns dos Grandes Desafios da área de SI descritos em [Boscarioli et al. 2017]. Inicialmente, as redes sociais online podem ser consideradas

“Sistemas de Sistemas”, trazendo desafios para a sua construção e principalmente para a gestão do seu conteúdo. Conforme dito no capítulo “Sistemas de Informação e os desafios do mundo aberto”: “...Nos defrontamos com a complexidade que nos oferece a construção de sistemas cada vez menos caracterizados como um artefato fechado e intraorganizacional e cada vez mais como um organismo de conexão intra, inter e socio organizacional, de comportamento imprevisível e emergente”. A criação de conteúdos nas redes sociais online está ligada ao desafio de mundo aberto, tendo em vista que todo o processo de consumo de informação - que engloba a criação, disseminação, uso e alteração de um conteúdo nas mídias sociais - não possui mais fronteiras entre as plataformas, geográficas, institucionais e a rápida proliferação de um conteúdo tóxico (como por exemplo, discriminatório) pode trazer consequências danosas e imprevisíveis. Consequentemente, surgem desafios no desenvolvimento deste tipo de sistema. Desafios no quesito técnico, como identificar e controlar a recomendação deste tipo de conteúdo e no quesito ético. Outro desafio é uma abordagem Sociotécnica, envolvendo uma “Visão Sistêmica e Socialmente Consciente”, tendo em vista os impactos sociais do problema tratado e da solução.

O trabalho contribui cientificamente no sentido de testar métodos consolidados pela literatura, porém integrados no cenário brasileiro e no idioma português-brasileiro, pois poucos trabalhos abordam questão de gênero em português aplicada a plataformas de vídeos. O diferencial deste trabalho é a apresentação de um método capaz de identificar viés de gênero (ou outras questões de interesse) em diversos cenários, o que torna-o replicável e, portanto, uma importante contribuição científica. Neste trabalho, a identificação de viés de gênero apontada no artigo publicado no BraSNAM em 2023 [Kappaun and Oliveira 2023] foi ratificada, porém com um escopo maior (de dados e de métricas, com a adição do engajamento) e identificação em mais cargos eleitorais.

5.3. Tecnológica

O método revela-se eficiente na detecção de viés de gênero e destaca-se por sua extrema acessibilidade, pois utiliza dados gratuitos e públicos, além de aproveitar ambientes de desenvolvimento sem custo adicional, os quais empregam processamento em nuvem. Essa abordagem não demanda investimentos expressivos para sua implementação.

Adicionalmente, neste estudo, foi elaborado um dataset contendo informações sobre os vídeos, juntamente com os comentários correspondentes. Nesse conjunto de dados, é possível encontrar para cada vídeo informações como gênero, raça e cargo disputado pelo candidato mencionado nos títulos ou descrições, assim como os nomes dos autores de cada comentário e seus identificadores individuais. O dataset, disponibilizado de maneira pública e gratuita por meio da plataforma Zenodo⁹, é percebido como uma valiosa contribuição que pode ser empregada para enriquecer pesquisas na área. Por exemplo, ele pode ser utilizado para explorar a aceitação de gênero em relação à audiência, bem como para outras análises relevantes por parte de membros da comunidade interessados em diversos temas.

6. Conclusão

O método se mostrou eficiente, pois conseguiu levantar e classificar uma grande quantidade de dados com uma acurácia satisfatória, além de responder a QP e alcançar o OP.

⁹<https://zenodo.org/records/10109864>

Vale ressaltar que pode ser replicado para outros cenários, não somente à política, e com um escopo maior de dados. Os resultados foram satisfatórios e condizentes com o método proposto. Foi importante observar o viés de gênero explicitado na Literatura internacional, só que na realidade brasileira. Vale ressaltar que as últimas eleições no Brasil foram muito marcadas pela polarização e pelo uso de vídeos na internet. O método estatístico proposto provou-se sólido em apontar o viés de gênero nas eleições de 2022.

Quanto às limitações do método, o fato de 2018 ter poucos vídeos e apenas duas candidatas mulheres à presidência, um dos cargos de maior destaque entre os demais, pode ter influenciado em seus resultados inferiores aos observados em 2022. Outra dificuldade encontrada foi a similaridade de nomes dos candidatos com outras figuras políticas influentes (familiares, por exemplo), o que demanda mais cuidado nas análises para evitar eventuais falsos positivos. Não se pode afirmar se o viés de gênero identificado é da plataforma ou só um reflexo da sociedade. Além disso, não foram analisados o conteúdo dos vídeos. Em relação às perspectivas desta pesquisa, pretende-se ampliar o escopo, considerando mais canais de notícias e agregando outras áreas políticas do executivo e legislativo (prefeitos e deputados, por exemplo). Por fim, para aprofundar esta pesquisa, vale abordar outras características do texto comentado, como as emoções (raiva, tristeza, deboche, desconfiança, alegria, etc) e identificação do discurso de ódio.

Referências

- Boscarioli, C., de Araujo, R. M., Maciel, R. S., Neto, V. V. G., Oquendo, F., Nakagawa, E. Y., Bernardini, F. C., Viterbo, J., Vianna, D., Martins, C. B., et al. (2017). I grandsibr. *Sociedade Brasileira de Computação*.
- Greenwood, P. E. and Nikulin, M. S. (1996). *A guide to chi-squared testing*, volume 280. John Wiley & Sons.
- IBGE (2018). *Coordenação de População e Indicadores Sociais. Estatísticas de gênero: indicadores sociais das mulheres no Brasil*. n. 38. IBGE, Rio de Janeiro. Acesso em: 17 mar. 2023.
- Kappaun, A. and Oliveira, J. (2023). Análise sobre viés de gênero no youtube: Um estudo sobre as eleições presidenciais de 2018 e 2022. In *Anais do XII Brazilian Workshop on Social Network Analysis and Mining*, pages 127–138, Porto Alegre, RS, Brasil. SBC.
- Media, K. I. (2021). Consumo de vídeo bate recorde no brasil. <https://www.kantaribopemedia.com/consumo-de-video-bate-recorde-no-brasil/>. Acesso em: 01/03/2022.
- O. Plath, H., O. Paiva, M. E., L. Pinto, D., and D. P. Costa, P. (2022). Detecção de discurso de Ódio contra mulheres em textos em português brasileiro: Construção da base mina-br e modelo de classificação. *Revista Eletrônica de Iniciação Científica em Computação*, 20(3).
- Pérez, J. M., Giudici, J. C., and Luque, F. (2021). pysentimiento: A python toolkit for sentiment analysis and socialnlp tasks. *arXiv preprint arXiv:2106.09462*.
- Sirius, A. (2019). Das fake news aos discursos de Ódio: Uma análise à luz da constituição cidadã nas mídias sociais.
- Social, W. A. (2023). Digital 2023.
- Stechemesser, A., Wenz, L., and Levermann, A. (2020). Corona crisis fuels racially profiled hate in social media networks. *EClinicalMedicine*, 23.

Wegener, C., Prommer, E., and Linke, C. (2020). Gender representations on youtube: The exclusion of female diversity. *M/C Journal*, 23(6).

Wotanis, L. and McMillan, L. (2014). Performing gender on youtube: How jenna marbles negotiates a hostile online environment. *Feminist Media Studies*, 14(6):912–928.