You Must be a Trump Supporter: Political Identity Projections on the Social Web

Shubh Mittal[1 [0009-0005-1048-5990]], Tisha Chawla[1 [0009-0003-9722-5221]], and Ashiqur R. KhudaBukhsh[2 [0000-0003-2394-7902]]

[1] Vellore Institute of Technology, Vellore, Tamil Nadu 632014, India
{shubh.mittal2020, tisha.chawla2020}@vitstudent.ac.in
[2] Rochester Institute of Technology, Rochester, NY 14623, US
axkvse@rit.edu

Abstract. ⚠**This paper contains offensive content.** This paper assesses the extent of political polarization in the United States by demonstrating the phenomenon of political identity projection, where individuals attribute political affiliations to others based on political discourse. This aspect of political behavior, often found in interactions between authors on various social media platforms, remains relatively unexplored. To address this gap, our research utilizes a comprehensive dataset of comments on YouTube news videos from three prominent US cable news networks (Fox News, CNN, and MSNBC) to interpret expressions of political polarization. First, we assess the accuracy of LLMs in identifying political identity projections, exploring the potential biases these models may incorporate. Second, we conduct a user engagement analysis that highlights interaction patterns and their implications for understanding political identity projections across different news outlets.

Keywords: Political Identity Projections · Large Language Models · Social Web.

1 Introduction

Polarization [10, 34, 42, 50] has emerged as a pressing concern in the United States, with evident adverse effects [59] on multiple fronts, including the educational system [3], job market [2], healthcare facilities [58], religious institutions [8], and the political arena [22]. Scholarship disparities [25], alongside challenges in healthcare access and divisions within religious communities, underscores the breadth of this issue. Moreover, political polarization significantly shapes public discourse [38, 39] and policy-making processes [23, 60]. Particularly within social media platforms, discussions often reflect this divide, leading to detrimental outcomes such as misinformation [32, 36] and social fragmentation [44]. Studies highlight the pivotal role of platforms like Facebook [24], YouTube [20, 46], and Twitter [49] in shaping online engagement and influencing public opinion and policy debates [17].

Additionally, analyses of major US cable news networks' [13, 27, 41] content and audience engagement on platforms like YouTube reveal a distinct partisan [9, 34, 35, 55] and ideological split [63]. This polarization extends beyond online spaces, further

MSNBC	CNN	FOX
Calling someone you don't know a freeloader? That seems a little presumptuous, don't you think? Not to be so myself, but I'm thinking you must be a Trump supporter. Is it his policy or his cult of personality that you are drawn to?	And how did you come to that brilliant conclusion? Several paid agitators (many likely from out of state) organized to interrupt a policy speech at set intervals with their anti-white communist rhetoric, and suddenly the whole state hates him? You and anyone who believes you must be anencephalic.	This message is for Isa, I think you hate President Trump because he beat Hillary plus you must be a little snow flake, you should move to Disneyland to live a fake life and you can dance and sing all they long.
Ohh, so Fox News is owned by Democratic elites? The channel that claims to be the number 1 watched channel? Are you kidding me? You must be a troll because your logic is totally twisted around. Who are the elites who have purchased most of the radio waves in the US? It's name is Sinclair Corp and it is all right wing programming. I think the media has it's problems for sure. But to label it all left is crazy. I also think it's crazy to get your news from one news source (Fox)! Spread your wings, Cary!	It's a volcanic land mass. "Fact". Facts don't have anything to do with emotions or opinion. Funny how your care so much for my opinion on why I won't visit or live in a particular area. You must be Republican. Twisting things into a narrative so that you may assume some kind of none existant moral high ground. I don't do stupid shit, you also don't know a damn thing about me.	Really?? You must be a Hilary supporter and a CNN listener also. Trump talks about what he does to all the crowds at his rallies, at least he hasn't murdered anybody or lied to his supporters, unlike Hilary. I'm dillusional?? I think you are if you support Hilary and her crime sprees which makes you unpatriotic. Please don't use rhetoric which you don't understand just to make a comment.
you must be blind if you think hillary is involved in this russia nonsense. but thats what you want. keep saying her name as trump goes to jail	They are not killing all life...just black ones....That mean black lives matters too! That is all that means...and its people like you the problem....oh you must be a cop!	haha you must be thinking of downtown Chicago. Our economy is much more stable than Europe, the DOW is reaching record highs, and the Trump administration is about to create national infrastructure for a minimum of 4 years. Lack of commodities? We have the biggest amount of consumerism on the planet.

Table 1: Illustrative examples highlighting different identity projections in the YouTube comment sections of MSNBC (left column), CNN (middle column), and Fox News (right column).

exacerbating societal divisions [31] and challenging efforts to foster constructive dialogue and governance [28]. Despite the apparent ubiquity of these discussions, there remains a lack of comprehensive analysis focusing on political identity projection where user $\mathscr{A}$ assumes a political identity for user $\mathscr{B}$. Consider the comment *if you find this movie offensive, you must be a sensitive snowflake* .... In this comment, the comment author is projecting a political identity (*sensitive snowflake*) to her target. Table 1 lists a few illustrative examples.

How often do we notice such political identity projections happen in the social web political discourse? What is the broad nature of such identity projections? This research investigates political identity projection via a massive dataset with documented political dissonance [34, 35] of more than 80 million YouTube comments on news videos hosted by the official YouTube channels of CNN, MSNBC, and FOX. Our analysis seeks to understand the pervasive nature of political polarization and how it shapes user interactions and discussions in different settings. Specifically, our contributions are as follows:

- First, we identify a set of political identity projections frequently used on the social web. We examine how well large language models (LLMs) can identify these political identity projections.

– Second, we analyze user engagement patterns in the presence of political identity projections.

Our analyses reveal that

1. while humans can discern political leanings in identity projections consistently, not all large language models are equally astute;
2. beyond mere name-calling, social web posts containing political identity projections also contain politically polarizing texts;
3. political identity projections elicit more user endorsement (in the form of likes) than non-political identity projections; and
4. political identity projections often trigger an escalating behavior in the discourse.

2 Related Work

Political polarization in the US has reached a level where hyper-partisanship significantly impacts both governance and societal unity [19, 29, 40]. Research indicates that these ideological divides intensify during election cycles and are amplified by the dynamics of social media and traditional news consumption [24, 60]. This escalation in polarization is not just a matter of differing opinions but influences legislative gridlock [4, 6] and reduces the effectiveness of governance [43, 54]. Moreover attempts to expose individuals to opposing political views on social media can unintentionally intensify their original partisan biases [3], known as the backfire effect, complicating efforts to mitigate polarization through digital platforms. Content moderation efforts to keep the online environment safe can also become highly subjective to political leanings [52, 61].

Stereotypes in US politics [7], particularly those surrounding Democrats and Republicans, play a crucial role in shaping public perceptions [14] and voter behaviour [12]. Media portrayals often exacerbate these stereotypes, leading to polarized public opinion which, in turn, influences electoral outcomes [48] and political engagement [5, 45]. These stereotypes can drastically simplify complex political narratives, leading to a binary and often confrontational political landscape. Our research builds on these findings by investigating political identity projection, where individuals attribute political affiliations based on observed behaviors and discussions online [37, 56]. We investigate how digital interactions reflect and reinforce perceived political identities, exploring the impact of media narratives on these perceptions.

3 Dataset

The dataset comprises a collection of user comments on videos hosted by the official YouTube channels of three major US cable news networks: CNN, Fox News, and MSNBC. Recognized for their significant viewership [26] and considered among the most viewed cable news outlets in the United States, these networks collectively cater to a diverse audience spectrum [15]. Our corpus consists of a vast volume of data, including 46,990,892 comments and 35,177,044 replies, across 191,205 videos which span from 2015 to 2020 detailed in Table 2 and Table 1. This dataset is a reliable snapshot of US political discourse and has been previously used to study political polarization [34], election misinformation [35], and health misinformation [62].

In this research, we deal with comments and replies having the projection '*you must be*'. We acknowledge that identity projections can happen in several other ways (e.g., *it seems I am talking to*). That said, we believe this phrase is one of the most common ways to project political identities and the sheer scale of our data ensures a meaningful analysis. Prior literature also indicates similar text template-based searches to limit the focus to a small portion of relevant data (see, e.g., [21, 33]).

News channel	**Videos**	**Comments**	**Replies**
FOX News	64,696	20,156,002	12,650,882
CNN News	95,230	19,290,236	15,183,800
MSNBC News	31,279	10,241,259	7,342,362

Table 2: Distribution of our dataset.

4 Political identity projection Identification

We first narrow our search down to frequently used identity projections (e.g., *you must be a Trump supporter* or *you must be a Soros hack*). We next manually annotate these examples (one male and one female annotator) with three labels: *liberal*, *conservative*, and *unrelated*. This process mostly yielded consensus labels. The inter-annotator agreement was measured using Cohen's Kappa, resulting in a score of 0.98, indicating almost perfect agreement. Very few remaining disagreements were resolved by an expert social scientist [3].

RQ 1: *Can large language models detect political identity projections?*

Our human annotation exercise reveals humans are highly consistent in categorizing political identity projections. But how good are large language models (LLMs) in this task? We sample the most frequent 1,000 instances of political identity projections evenly distributed across the three categories (*liberal*, *conservative*, and *unrelated*). We assess the accuracy of LLMs in identifying political identity projections (shown in Table 3). We consider the following LLMs: `GPT-4` [1], `Mistral` [30], and `Gemini` [57]. Table 4 details the LLM performance. The results demonstrate a significant difference in the performance of LLMs for political identity projection identification. We observe that `GPT-4`, with an accuracy of 95%, is much better suited for political identity projection identification. In contrast, `Mistral` and `Gemini` considerably underperform with an accuracy of 43% and 21%, respectively. `Gemini`'s poor performance perhaps suggests vulnerability to social biases [47]. The performance disparity highlights that while humans can easily interpret political identity projections, not all LLMs are equally adept at this.

RQ 2: *Beyond the projected identity, do comments containing conservative or liberal identity projections exhibit polarization?*

[3] Publicly available at Github link: https://bit.ly/46oQZHu.

Political Identity Projections	**Annotations**
you must be a trump supporter	conservative
you must be a Foxnews idiot	conservative
you must be a republican conman	conservative
you must be a hillary bot	liberal
you must be a democrat fucktard	liberal
you must be a soros hack	liberal
you must be a fucking genius	unrelated
you must be a god	unrelated
you must be a nice person	unrelated

Table 3: Example annotations associated with different political identity projections, showcasing high-confidence predictions from GPT-4 trained model to identify the political labels.

LLM	**Accuracy**
GPT-4	95%
Mistral	43%
Gemini	21%

Table 4: Accuracy of LLMs for identification of political identity projections.

We were curious to investigate if the political identity projections are mere name-calling, or if they also contain additional polarizing linguistic signals. We investigate this through a recent framework that leverages classification accuracy as a proxy for polarization [13]. Consider $\mathcal{D}_1$ and $\mathcal{D}_2$ are two datasets. If $\mathcal{D}_1$ and $\mathcal{D}_2$ are distributionally highly similar, a model that takes an instance of any of these two corpora as input and tries to predict which of the two corpora the instance is coming from, will have a very hard time predicting. On a balanced test set with an equal number of samples from $\mathcal{D}_1$ and $\mathcal{D}_2$, the accuracy of the model will be close to chance. In contrast, if $\mathcal{D}_1$ and $\mathcal{D}_2$ are distributionally highly dissimilar, the classification accuracy will be considerably better than chance.

We consider each comment to have two components: political identity projection and targeted content. For example, in our example *if you find this movie offensive, you must be a sensitive snowflake*, the projected identity is *sensitive snowflake* and the content stripped of the identity projection is *if you find this movie offensive*. If we create two corpora, one with the comments with liberal identity projections ($\mathcal{D}_{liberal}$) and the other with conservative identity projections ($\mathcal{D}_{conservative}$), we can run a similar experiment to estimate how linguistically dissimilar these two corpora are. However, if we do not remove the identity projections, the classification will be trivial due to shortcut learning [18]. If we strip both $\mathcal{D}_{liberal}$ and $\mathcal{D}_{conservative}$ off the projected identity and run a similar experiment, an accuracy considerably higher than chance possibly indicates the presence of polarizing content that extends beyond political identity projections.

To this end, we develop a content classifier using BERT [11]. We remove the identity projections from $\mathscr{D}_{liberal}$ and $\mathscr{D}_{conservative}$ and retain all comments containing more than 20 words to ensure that comments have sufficiently long context. This method evaluated how well the classifier could identify political projections after removing our target phrase as shown in Table 5. We achieved an accuracy of 70%, considerably more than chance. This result indicates that comments with political identity projections also contain polarizing texts beyond the juvenile name-calling. Few illustrative examples are listed in Table 5.

Liberal	Conservative
~~You must be a demorat!~~? If this was a N.Korea you would not be here today! Are you a Clinton supporter? Cinton is dangerous for everybody think before judging your president.	This is the procedure of the rules passed by the Republicans during the Bengazi inquiry , moronic fool.you remember that nothing burger ran by Gowdy and Pompeo that went nowhere.so get your facts straight before you make yourself look like a fool.ah, it's too late.~~you must be Republican~~.
where the hell do you get your disinformation ? ~~You must be Obama loser~~. Trump is way more of a leader than Obama ever was. Our economy is the strongest it has been in thirty years. Under Obama unemployment was in double digits and you couldn't find a job and welfare was the highest it's ever been.	How is the weather in Siberia? If not, then ~~you must be a Trumptard~~ wallowing in deflection and whataboutism as always. Attempting to relay facts and reason with your kind is like trying to administer medicine to the dead, it's just futile. However, you keep attempting to evoke "ghetto" in a derogatory as if your people don't reside their too-meth infested trailer parks are the worst! Lastly, your orange Nazi-in-chief are the worst thing ever for this country, so you can't throw policy stones when you live in a cheap glass house. Jeez!
Who brought up "the left" or "nazis"?How are either related to my post about Israeli exceptionalism and influence in US politics and foreign policy?What an interesting knee jerk reaction to criticism of Israel- you must be a leftist or a nazi! Lmao is this the power of the Israeli online defense force? You criticize Israel? ~~You're a leftist nazi!~~The war in Iraq cost the US 2 trillion dollars and is projected to cost a total 6 trillion, and for what? Nothing. ISIS was formed, millions died in the conflicts after the war, and only Israel and Saudi Arabia benefited. The US paid the bill while our "greatest ally" didn't send a single soldier to help.	did i ever state prisons are full of innocent people lol ~~you must be republican~~ huh ive noticed a trend where repubs connect there own dots and put words in peoples mouths. What i said is that harsh punishments for minor crimes arent helping anyone i also stated selling a bit of weed not getting caught with a little weed intent to distribute carries a harsher punishment i would expect an expert on everyone and everything to understand that lol

Table 5: Example replies to feed into the content classifier with the target label '*you must be*' removed and categorized into two classes: liberal (left column) and conservative (right column).

5 User Engagment

5.1 Counter-projecting

We identified 6,944 comments across the three news channels where political identity projections occurred more than once within the replies section (Table 6 provides a detailed breakdown by channel). This highlights recurring patterns of targeted projections within the dataset. Further investigation into the reply thread revealed that, on average, the second instance of a target or insulting reply occurs after 17 subsequent replies to the initial comment. This finding suggests a probable escalation of negative interactions, indicating that the initial instance of target projections may predispose the conversation towards subsequent negative replies. The concept of counter-projections emerges in the analysis, referring to the reactionary patterns observed within the replies. These patterns indicate that subsequent replies may either oppose or affirm the political identity projections introduced in earlier conversations, contributing to a complex political interaction within online discussions.

News Channel	Total political identity projections
CNN	3,337
FOX	2,473
MSNBC	1,134

Table 6: Total political identity projections associated with each news channel.

5.2 Annual Disaggregation

The analysis in Fig. 1 shows an increase in negative comments during U.S. presidential election years, particularly in 2016 and 2020.

- *Election vs. Non-Election Years :* Across the three news networks, substantial variations in targeted comments have been observed during the election cycles. While FOX experienced a significant peak in political identity projecting comments during the 2020 election year, CNN displayed its highest level of such comments in 2019, contrary to an expected rise in an election year like 2020. This suggests that CNN's audience engagement and the nature of discourse may have been influenced by events leading up to the election year. MSNBC, on the other hand, showed a continuous increase in political identity projections, culminating in 2020, indicating heightened viewer engagement during the election period.
- *Comparative Channel Analysis :* Comparing the channels, FOX (30,889 identity projections) and CNN (38,455 identity projections) had higher instances of politically projecting comments than MSNBC (15,684 identity projections). This could be reflective of the different editorial content and audience engagement strategies employed by these networks. The higher numbers in FOX and CNN show that these channels may foster more polarized environments, which can increase during critical political periods.

5.3 Participation Disparities in Contributors

Over 44 users heavily contribute to the comment section, surpassing 50 comments, indicating concentrated activity among top contributors. Additionally, there are 531 middle-tier users with comment counts ranging from 10 to 50, contributing significantly, though less than the top tier but more than the less engaged users. Moreover,

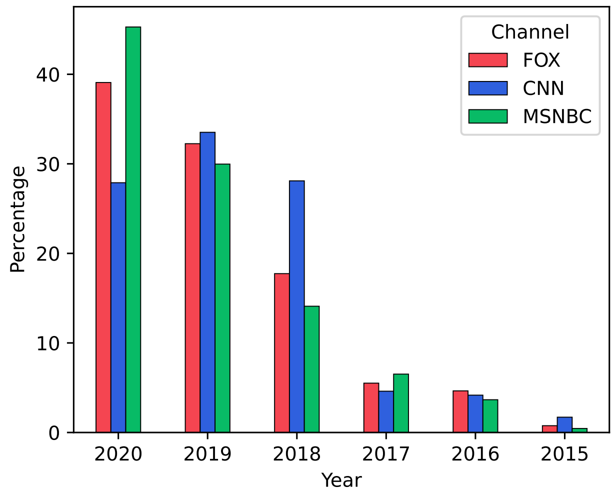

Fig. 1: Graphical representation of the yearly political identity projection count for the news channels.

17,670 users, have made fewer than 10 comments, illustrating a long tail of less involved users. This distribution follows a Pareto pattern, where a few users are highly active while the majority are minimally engaged.

5.4 User Reaction Distribution

We observe distinct patterns in the mean like counts of political identity projection comments across the three news channels (shown in Table 7), reflecting the political inclinations of their respective audiences. FOX, known for its Republican-leaning viewership [53], exhibits higher mean like counts for comments projecting liberals (2.0384) compared to those on CNN (1.0124) and MSNBC (0.8195), which are channels with predominantly Democratic-leaning audiences [16]. Conversely, comments about conservatives receive more likes on CNN (1.7175) and MSNBC (2.9395), considerably higher than that on FOX (1.3656).

This trend highlights that viewers are more likely to engage with and endorse comments that align with identity projections of the opposing political party, reinforcing the echo chamber effect [51] where ideological biases are perpetuated. The analysis also shows that comments categorized as ‘Unrelated,’ which contain identity projections not specific to any political group, receive relatively uniform mean like counts across all channels (CNN: 1.4989, FOX: 1.5738, MSNBC: 1.7981). This indicates that non-partisan identity projections are roughly equally likely to be engaged with, regardless of the audience’s political leanings.

Class	μ(CNN)	μ(FOX)	μ(MSNBC)
Liberal	1.0124	2.0384	0.8195
Conservative	1.7175	1.3656	2.9385
Unrelated	1.4989	1.5738	1.7981

Table 7: Mean like count of political identity projections across our data.

6 Conclusions and Discussion

This paper analyzes political identity projections in comments/replies on YouTube videos from prominent news channels such as CNN, FOX, and MSNBC. We utilize well-known LLMs such as `GPT-4`, `Mistral`, and `Gemini` to label the political identity projections in user comments with `GPT-4` demonstrating a superior performance as compared to other LLMs. Our results show a significant increase in political identity projections during U.S. election years, illustrating how political events intensify user engagement and alter online interactions. Our analysis also highlights that political leanings may influence user engagement with political identity projections suggesting political polarization.

Our research raises the following points:

- **Effectiveness of LLMs in Target Projection Identification:** We compared the performance of `GPT-4`, `Mistral`, and `Gemini` in labelling political projections and concluded that `GPT-4` provides the results that best align with human labels. On the other hand, `Mistral` and `Gemini` prediction show lesser alignment with human labels. This result indicates that before using large language models for tasks requiring understanding of political subtleties, we need to carefully evaluate the LLM's capability to understand political nuances.
- **Bias in Political Discourse:** Our analysis indicates that channels like CNN and FOX display a higher incidence of targeted comments against specific political groups, which may be influenced by the channel's editorial biases or the perceived political leanings of their audiences. This finding calls for further examination of how news media can influence or reflect political biases in public discourse.
- **Influence of Election Cycles on User Engagement:** The observed spike in user activity and polarized projections during election years suggests that these periods not only increase user engagement but also heighten political identity projections. Future research could explore how moderation strategies and community management can lessen polarization during election cycles, helping to mitigate divisiveness on social media platforms.

Our study has several limitations. We primarily focused on the phrase '*you must be*' to identify political identity projections within YouTube comments, potentially missing other similar expressions like '*you should be*' or '*you are an*'. Also, the vast scale of the dataset poses challenges in manually reviewing and identifying every potential expression of identity projections. Our reliance on a template-based method, while common in linguistic analysis, may not fully capture the distinct ways in which individuals project political identities. Further, while our dataset is a reasonable snapshot of US political discourse, it does not include other platforms like Reddit or Twitter, which could provide additional contexts for understanding online political discourse.

7 Ethical Statement

This research analyzes publicly available data from YouTube procured using publicly available YouTube API. We initially annotate this data by humans to identify instances of political identity projections, with subsequent labelling supported by LLMs. We consider both male and female annotators and compare multiple LLMs. We report results in aggregate form and reveal no individual user data. The objective of this research is not to defame any political group, news channel, or community. Instead, our goal is to shed light on the extent of political polarization evident in public discourse on major news channels' YouTube platforms, aiming to contribute to a broader understanding of political polarization through the lens of political identity projections.

References

1. Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F.L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., et al.: Gpt-4 technical report. arXiv preprint arXiv:2303.08774 (2023)
2. Autor, D., et al.: The polarization of job opportunities in the us labor market: Implications for employment and earnings. Center for American Progress and The Hamilton Project **6**, 11–19 (2010)
3. Bail, C.A., Argyle, L.P., Brown, T.W., Bumpus, J.P., Chen, H., Hunzaker, M.F., Lee, J., Mann, M., Merhout, F., Volfovsky, A.: Exposure to opposing views on social media can increase political polarization. Proceedings of the National Academy of Sciences **115**(37), 9216–9221 (2018)
4. Barber, M., McCarty, N., Mansbridge, J., Martin, C.J.: Causes and consequences of polarization. Political negotiation: A handbook **37**, 39–43 (2015)
5. Barcelos, R.H.: To engage or not engage? the features of video content on youtube affecting digital consumer engagement. Journal of Consumer Behaviour **20**(1), 15–26 (2021)
6. Binder, S.A.: Stalemate: Causes and consequences of legislative gridlock. Rowman & Littlefield (2004)
7. Bordalo, P., Tabellini, M., Yang, D.Y.: Stereotypes and politics (2020)
8. Castle, J.J., Stepp, K.K.: Partisanship, religion, and issue polarization in the united states: A reassessment. Political Behavior pp. 1–25 (2021)
9. Chen, K., Babaeianjelodar, M., Shi, Y., Janmohamed, K., Sarkar, R., Weber, I., Davidson, T., De Choudhury, M., Huang, J., Yadav, S., et al.: Partisan us news media representations of syrian refugees. In: Proceedings of the International AAAI Conference on Web and Social Media. vol. 17, pp. 103–113 (2023)
10. Demszky, D., Garg, N., Voigt, R., Zou, J., Gentzkow, M., Shapiro, J., Jurafsky, D.: Analyzing polarization in social media: Method and application to tweets on 21 mass shootings (2019)

11. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018)
12. Ditonto, T.: Direct and indirect effects of prejudice: Sexism, information, and voting behavior in political campaigns. Politics, Groups, and Identities **7**(3), 590–609 (2019)
13. Dutta, S., Li, B., Nagin, D.S., KhudaBukhsh, A.R.: A murder and protests, the capitol riot, and the chauvin trial: Estimating disparate news media stance. In: Raedt, L.D. (ed.) Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI 2022. pp. 5059–5065. ijcai.org (2022). https://doi.org/10.24963/IJCAI.2022/702, https://doi.org/10.24963/ijcai.2022/702
14. Eagly, A.H., Nater, C., Miller, D.I., Kaufmann, M., Sczesny, S.: Gender stereotypes have changed: A cross-temporal meta-analysis of us public opinion polls from 1946 to 2018. American psychologist **75**(3), 301 (2020)
15. Emelu, M.N.: The us cable televisions' framing of mass shooting: a grounded discovery of competing narratives. Frontiers in communication **8**, e1174946–e1174946 (2023)
16. Ensor, K.: The partisan delivery of news: A content analysis of cnn and fox. Johnson & Wales University, ScholarsArchive@ JWU.[online] **16**(11), 2018 (2018)
17. Garimella, V.R.K., Weber, I.: A long-term analysis of polarization on twitter. In: Proceedings of the International AAAI Conference on Web and social media. vol. 11, pp. 528–531 (2017)
18. Geirhos, R., Jacobsen, J.H., Michaelis, C., Zemel, R., Brendel, W., Bethge, M., Wichmann, F.A.: Shortcut learning in deep neural networks. Nature Machine Intelligence **2**(11), 665–673 (2020)
19. Grossmann, M., Hopkins, D.A.: Ideological republicans and group interest democrats: The asymmetry of american party politics. Perspectives on Politics **13**(1), 119–139 (2015)
20. Halpern, D., Gibbs, J.: Social media as a catalyst for online deliberation? exploring the affordances of facebook and youtube for political expression. Computers in human behavior **29**(3), 1159–1168 (2013)
21. Halterman, A., Keith, K.A., Sarwar, S.M., O'Connor, B.: Corpus-level evaluation for event QA: the indiapoliceevents corpus covering the 2002 gujarat violence. In: Findings of the Association for Computational Linguistics: ACL/IJCNLP 2021, Online Event, August 1-6, 2021. Findings of ACL, vol. ACL/IJCNLP 2021, pp. 4240–4253. Association for Computational Linguistics (2021)
22. Hare, C., Poole, K.T.: The polarization of contemporary American politics. Polity **46**(3), 411–429 (2014)
23. Hefeker, C., Neugart, M.: Policy rules and political polarization (2024)
24. Heiss, R., Schmuck, D., Matthes, J.: What drives interaction in political actors' facebook posts? profile and content predictors of user engagement and political actors' reactions. Information, communication & society **22**(10), 1497–1513 (2019)
25. Hetherington, M.: Review article: Putting polarization in perspective. British Journal of Political Science **39**, 413 – 448 (04 2009). https://doi.org/10.1017/S0007123408000501
26. Hong, J., Crichton, W., Zhang, H., Fu, D.Y., Ritchie, J., Barenholtz, J., Hannel, B., Yao, X., Murray, M., Moriba, G., et al.: Analysis of faces in a decade of us cable tv news. In: KDD'21: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining (2021)
27. Hyun, K.D., Moon, S.J.: Agenda setting in the partisan tv news context: Attribute agenda setting and polarized evaluation of presidential candidates among viewers of nbc, cnn, and fox news. Journalism & Mass Communication Quarterly **93**(3), 509–529 (2016)
28. Iandoli, L., Primario, S., Zollo, G.: The impact of group polarization on the quality of online debate in social media: A systematic literature review. Technological Forecasting and Social Change **170**, 120924 (2021)
29. Iyengar, S., Sood, G., Lelkes, Y.: Affect, not ideology: A social identity perspective on polarization. Public opinion quarterly **76**(3), 405–431 (2012)

30. Jiang, A.Q., Sablayrolles, A., Mensch, A., Bamford, C., Chaplot, D.S., Casas, D.d.l., Bressand, F., Lengyel, G., Lample, G., Saulnier, L., et al.: Mistral 7b. arXiv preprint arXiv:2310.06825 (2023)
31. Jost, J.T., Baldassarri, D.S., Druckman, J.N.: Cognitive–motivational mechanisms of political polarization in social-communicative contexts. Nature Reviews Psychology **1**(10), 560–576 (2022)
32. Juneja, P., Bhuiyan, M.M., Mitra, T.: Assessing enactment of content regulation policies: A post hoc crowd-sourced audit of election misinformation on youtube. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. pp. 1–22 (2023)
33. KhudaBukhsh, A.R., Bennett, P.N., White, R.W.: Building effective query classifiers: A case study in self-harm intent detection. In: Proceedings of the 24th ACM International Conference on Information and Knowledge Management, CIKM 2015, Melbourne, VIC, Australia, October 19 - 23, 2015. pp. 1735–1738. ACM (2015)
34. KhudaBukhsh, A.R., Sarkar, R., Kamlet, M.S., Mitchell, T.: We don't speak the same language: Interpreting polarization through machine translation. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 35, pp. 14893–14901 (2021)
35. KhudaBukhsh, A.R., Sarkar, R., Kamlet, M.S., Mitchell, T.M.: Fringe news networks: Dynamics of us news viewership following the 2020 presidential election. In: 14th ACM Web Science Conference 2022. pp. 269–278 (2022)
36. Kouzy, R., Abi Jaoude, J., Kraitem, A., El Alam, M.B., Karam, B., Adib, E., Zarka, J., Traboulsi, C., Akl, E.W., Baddour, K.: Coronavirus goes viral: quantifying the covid-19 misinformation epidemic on twitter. Cureus **12**(3) (2020)
37. Kubin, E., Von Sikorski, C.: The role of (social) media in political polarization: A systematic review. Annals of the International Communication Association **45**(2), 188–206 (2021)
38. Kubin, E., Von Sikorski, C.: The role of (social) media in political polarization: a systematic review. Annals of the International Communication Association **45**(3), 188–206 (2021)
39. Lee, C., Shin, J., Hong, A.: Does social media use really make people politically polarized? direct and indirect effects of social media use on political polarization in south korea. Telematics and Informatics **35**(1), 245–254 (2018)
40. Lee, F.E.: How party polarization affects governance. Annual review of political science **18**, 261–282 (2015)
41. Marozzo, F., Bessi, A.: Analyzing polarization of social media users and news sites during political campaigns. Social Network Analysis and Mining **8**, 1–13 (2018)
42. McCarty, N., Poole, K.T., Rosenthal, H.: Polarized America: The dance of ideology and unequal riches. mit Press (2016)
43. McCoy, J., Rahman, T., Somer, M.: Polarization and the global crisis of democracy: Common patterns, dynamics, and pernicious consequences for democratic polities. American Behavioral Scientist **62**(1), 16–42 (2018)
44. Minh Pham, T., Kondor, I., Hanel, R., Thurner, S.: The effect of social balance on social fragmentation. Journal of the Royal Society Interface **17**(172), 20200752 (2020)
45. Munaro, A.C., Hübner Barcelos, R.: Youtube as a source of patient information for coronavirus disease (covid-19): a content-quality and audience engagement analysis. Reviews in Medical Virology **30**(3), 2105 (2021)
46. Munaro, A.C., Hübner Barcelos, R., Francisco Maffezzolli, E.C., Santos Rodrigues, J.P., Cabrera Paraiso, E.: To engage or not engage? the features of video content on youtube affecting digital consumer engagement. Journal of consumer behaviour **20**(5), 1336–1352 (2021)
47. NPR: Google races to find a solution after ai generator gemini misses the mark (2024), https://www.npr.org/2024/03/18/1239107313/google-races-to-find-a-solution-after-ai-generator-gemini-misses-the-mark

48. Papakyriakopoulos, O., Zuckerman, E.: The media during the rise of trump: Identity politics, immigration," mexican" demonization and hate-crime. In: Proceedings of the International AAAI Conference on Web and Social Media. vol. 15, pp. 467–478 (2021)
49. Park, C.S.: Does twitter motivate involvement in politics? tweeting, opinion leadership, and political engagement. Computers in human behavior **29**(4), 1641–1648 (2013)
50. Poole, K.T., Rosenthal, H.: The polarization of american politics. The journal of politics **46**(4), 1061–1079 (1984)
51. Ross Arguedas, A., Robertson, C., Fletcher, R., Nielsen, R.: Echo chambers, filter bubbles, and polarisation: A literature review (2022)
52. Sap, M., Swayamdipta, S., Vianna, L., Zhou, X., Choi, Y., Smith, N.A.: Annotators with attitudes: How annotator beliefs and identities bias toxic language detection. In: Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL 2022. pp. 5884–5906. Association for Computational Linguistics (2022)
53. Schroeder, E., Stone, D.F.: Fox news and political knowledge. Journal of Public Economics **126**, 52–63 (2015)
54. Somer, M., McCoy, J.L., Luke, R.E.: Pernicious polarization, autocratization and opposition strategies. Democratization **28**(5), 929–948 (2021)
55. Stanley, A.: How msnbc became fox's liberal evil twin. The New York Times (2012), https://www.nytimes.com/2012/08/31/us/politics/msnbc-as-foxs-liberal-evil-twin.html, online; accessed 01-September-2020
56. Szmuda, T., Syed, M.T., Singh, A., Ali, S.: Youtube as a source of patient information for coronavirus disease (covid-19): a content-quality and audience engagement analysis. Reviews in Medical Virology **30**(3), 1–9 (2020)
57. Team, G., Anil, R., Borgeaud, S., Wu, Y., Alayrac, J.B., Yu, J., Soricut, R., Schalkwyk, J., Dai, A.M., Hauth, A., et al.: Gemini: a family of highly capable multimodal models. arXiv preprint arXiv:2312.11805 (2023)
58. Tesler, M.: The spillover of racialization into health care: How president obama polarized public opinion by racial attitudes and race. American Journal of Political Science **56**(3), 690–704 (2012)
59. Tucker, J.A., Guess, A., Barberá, P., Vaccari, C., Siegel, A., Sanovich, S., Stukal, D., Nyhan, B.: Social media, political polarization, and political disinformation: A review of the scientific literature. Political polarization, and political disinformation: a review of the scientific literature (March 19, 2018) (2018)
60. Weber, T., Hydock, C., Ding, W., Gardner, M., Jacob, P., Mandel, N., Sprott, D.E., Van Steenburg, E.: Political polarization: Challenges, opportunities, and hope for consumer welfare, marketers, and public policy. Journal of Public Policy & Marketing **40**(2), 184–205 (2021)
61. Weerasooriya, T.C., Dutta, S., Ranasinghe, T., Zamperi, M., Homan, C.M., KhudaBukhsh, A.R.: Vicarious offense and noise audit of offensive speech classifiers: Unifying human and machine disagreement on what is offensive. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). pp. 11648–11668 (2023)
62. Yoo, C.H., KhudaBukhsh, A.R.: Auditing and robustifying COVID-19 misinformation datasets via anticontent sampling. In: Thirty-Seventh AAAI Conference on Artificial Intelligence, AAAI 2023. pp. 15260–15268. AAAI Press (2023)
63. Gil de Zúñiga, H., Correa, T., Valenzuela, S.: Selective exposure to cable news and immigration in the us: The relationship between fox news, cnn, and attitudes toward mexican immigrants. Journal of Broadcasting & Electronic Media **56**(4), 597–615 (2012)