Analyzing Political Stances on Twitter/X in the lead-up to the 2024 U.S. Election

Hazem Ibrahim, Farhan Kamrul Khan, Yasir Zaki, Talal Rahwan
New York University Abu Dhabi
Abu Dhabi, UAE
{yasir.zaki, talal.rahwan}@nyu.edu

Abstract

Social media platforms play a pivotal role in shaping public opinion and amplifying political discourse, particularly during elections. However, the same dynamics that foster democratic engagement can also exacerbate polarization. To better understand these challenges, here, we investigate the ideological positioning of tweets related to the 2024 U.S. Presidential Election. To this end, we analyze 1,235 tweets from key political figures and 63,322 replies, and classify ideological stances into Pro-Democrat, Anti-Republican, Pro-Republican, Anti-Democrat, and Neutral categories. Using a classification pipeline involving three large language models (LLMs)—GPT-4o, Gemini-Pro, and Claude-Opus—and validated by human annotators, we explore how ideological alignment varies between candidates and constituents. We find that Republican candidates author significantly more tweets in criticism of the Democratic party and its candidates than vice versa, but this relationship does not hold for replies to candidate tweets. Furthermore, we highlight shifts in public discourse observed during key political events. By shedding light on the ideological dynamics of online political interactions, these results provide insights for policymakers and platforms seeking to address polarization and foster healthier political dialogue.

Keywords: stance detection, political polarization, large language models, social media analysis, U.S. elections

1. Introduction

Social media platforms have become pivotal mediums for political discourse, often shaping public opinion through their recommendation algorithms and serving as hubs for information exchange. However, the same features that make social media valuable for democratic participation also render it susceptible to the spread of misinformation, polarization, and manipulation. As these dynamics increasingly influence elections and policy debates, understanding how political messaging and interactions unfold on social media is essential for fostering informed civic engagement and safeguarding democratic processes. Within this broader context, detecting political ideology stance on social media holds particular importance for policymakers, researchers, and social media platforms. Stance detection can provide valuable insights into the ideological divisions, shifting narratives, and public sentiment surrounding key issues. For policymakers, such insights can inform strategies to address polarization, identify misinformation, and gauge public support for initiatives. Moreover, by enabling a deeper understanding of online discourse, accurate stance detection can guide interventions to promote healthier digital ecosystems while ensuring diverse viewpoints are represented in public debates.

The political content published on social media platforms has been extensively analyzed over the last decade, ranging from examinations of the homophily of recommendations (Boutyline and Willer, 2017), radicalization pathways (Ibrahim et al., 2023), and the prevalence of echo chambers (Cinelli et al., 2021). Most recently, in the context of the 2024 U.S. election, Ye et al. (2024) examined political recommendations made on Twitter through a socket puppet driven experiment, finding a default right-leaning bias in content exposure, with reduced exposure to opposing viewpoints. To classify ideological positioning and stance specifically, work prior to the introduction of large language models (LLMs) has relied on access to user characteristics, such as their retweet behaviour (Stefanov et al., 2020) or network interactions (Aldayel and Magdy, 2019), to classify stance. Yet, the introduction of LLMs has offered researchers an increasingly reliable method to scale text annotation tasks previously done by human annotators or through supervised or unsupervised machine learning models. As shown by Ziems et al. (2024), through zero-shot prompting, LLMs can reliably identify text characteristics such as stance, ideology, misinformation, and humor, among others. Of these, stance detection was the characteristic most reliably classified by GPT-4. Wu et al. demonstrated GPT-3.5's ability to produce accurate classifications of the ideology of U.S. political figures via analyzing their public statements (Wu et al., 2023). Lastly, as shown by Gilardi et al. (2023), LLMs can accurately measure the topic of tweets, and the framing used by the authors of tweets to organize the message in a narrative manner. Taken together,

Proceedings of The 3rd Workshop on Natural Language Processing for Political Sciences (PoliticalNLP 2026) @ LREC 2026, pages 58–63
16 May 2026, Palma, Mallorca (Spain).

prior work suggests that accurate categorizations of political texts can be obtained through LLM-driven, human-validated annotation pipelines (Törnberg, 2024).

Here, we seek to understand the ideological positioning of political candidates and their constituents in the lead up to the 2024 U.S. election on X.

To this end, we analyze a dataset of 1,235 tweets made by major U.S. political figures from May 1st, 2024 to November 1st, 2024 on the social media platform Twitter/X. Furthermore, we analyze the 63,322 replies made to the aforementioned political figure tweets to answer the following research questions. **RQ1:** How do political candidates from either side of the political aisle position their tweets from an ideological perspective? **RQ2:** How are the replies made to political candidate tweets ideologically positioned? **RQ3:** Is engagement with political tweets correlated with a certain ideological positioning? **RQ4:** Did the ideological positioning of political tweets change in response to major political events?

2. Methodology

Data: The dataset used in this study, collected by Balasubramanian et al. (2024), encompasses large-scale social media discourse on Twitter related to the 2024 U.S. Presidential Election. Tweets were collected from May 1st to November 1st, 2024, and includes approximately 27 million publicly available political tweets.

Stance classification pipeline: To classify tweets from prominent political figures, we began by isolating tweets in the dataset made by Joe Biden, Kamala Harris, and Tim Walz of the Democratic party, and Donald Trump and JD Vance of the Republican party. These figures were selected as projected nominees for their respective parties at the time the dataset was collected, with Biden being replaced by Harris following his decision not to participate in the election in July 2024. This resulted in a total of 1,235 tweets from the aforementioned candidates. We then gathered all replies made to these tweets, amounting to 63,322 replies.

To classify tweets, we utilized three LLMs, namely GPT-4o, Gemini-Pro, and Claude-Opus. Here, we make a small alteration to standard stance detection classification tasks. While prior work has largely aimed to classify the stance of a given statement or tweet into one of three classes (support, against, or none), we further separate the first two of these classes to delineate the support or criticism of a given political party. Specifically, we classify a given tweet into one of the following five categories: "Anti-Democrat" (AD), "Anti-Republican" (AR), "Pro-Democrat" (PD), "Pro-Republican" (PR), and "Neutral" (N). This is done to identify differences in the framing of ideological alignment within tweets in support of the two political parties. Indeed, a tweet aligned with the Republican party, for instance, may be directly in support of the Republican party or its candidate (PR), or potentially indirectly in support of the Republican party through criticism of the Democratic party or its candidate (AD). Lastly, we classify a random sample of 1000 tweets daily in the two weeks surrounding major political events, namely the first presidential debate on June 27th, the supreme court ruling on July 1st, and Trump's attempted assassination on July 13th. This amounted to a set of 32,832 tweets.

	Candidate Tweets	Reply Tweets	Event Tweets
Accuracy (%)	91.0	92.2	90.2
Macro F1-Score	91.1	85.5	88.3
Rater-LLM agreement	0.89^{***}	0.88^{***}	0.86^{***}
Inter-LLM agreement	0.74^{***}	0.80^{***}	0.61^{***}
Inter-rater agreement	0.87^{***}	0.92^{***}	0.79^{***}

Table 1: LLM accuracy, Macro F1-score, rater-AI agreement (Fleiss's κ), inter-LLM agreement and inter-rater agreement (Krippendorf's α) for text classification tasks

In our analysis, we take the majority vote of all three LLMs. Given the use of three classifiers, a majority vote was always obtainable when at least two models agreed. In the remaining instances where no two-model consensus was reached (i.e., all three LLMs assigned different labels), the tweet was discarded from our analysis, although this only accounted for 1.1% of all tweets. Please refer to Balasubramanian et al. (2024) for the full political tweets dataset. To identify the ideological alignment of reply tweets, the models were given the following prompt:

> *Given the following tweet made by* **[CANDIDATE]** *who is a* **[CANDIDATE PARTY]**, *and its reply, classify the reply into one of the categories.*
>
> *Candidate Tweet:* **[CANDIDATE TWEET]**
>
> *Reply Tweet:* **[REPLY TWEET]**
>
> *Anti-Democrat, Anti-Republican, Pro-Democrat, Pro-Republican, Neutral*

For classifying candidate tweets, a similar approach was used, with the following prompt provided:

> *Given the following tweet made by* **[CANDIDATE]** *who is a* **[CANDIDATE PARTY]**, *classify the tweet into one of the categories.*
> *Candidate Tweet:* **[CANDIDATE TWEET]**
>
> *Anti-Democrat, Anti-Republican, Pro-Democrat, Pro-Republican, Neutral*

We also classify a random sample of 1000 tweets daily in the two weeks surrounding major political events, namely the first presidential debate on June 27th, the supreme court ruling on July 1st, and Trump's attempted assassination on July 13th. This amounted to a set of 32,832 tweets. To do so, we used the following prompt:

> *Given the following tweet, classify the tweet into one of the categories.*
> *Tweet:* **[TWEET]**
>
> *Anti-Democrat, Anti-Republican, Pro-Democrat, Pro-Republican, Neutral, Not political*

Here, we add the "Not political" class to filter out tweets that may have been incorrectly classified as political during the scraping process of the prior work from which we obtain the dataset, although this amounted to only 8.06% of the 32,832 tweets.

The outputs of the aforementioned text-based classification prompts were validated by three independent coders (67% Female, mean age = 21, IRB Protocol is redacted to preserve anonymity). In the case of reply tweet classification, we verify the accuracy of the model by manually annotating a sample of 250 replies made to Biden tweets and 250 replies made to Trump tweets. In the case of candidate tweet classification, we similarly annotate all 128 Republican candidate tweets as well as 128 randomly selected Democrat candidate tweets. Given that Donald Trump only returned to Twitter on August 12th, 2024 following a year-long break from the platform, the number of tweets made by Republican candidates within the dataset (128) were significantly fewer than those authored by Democrat candidates (1,107). Lastly, we manually annotated a random sample of 500 tweets out of those used to quantify the impact of major political events. All annotators were undergraduate political science students who were given the liberty to research any terms or events they were unfamiliar with. Annotators were tasked with the same prompt given to the LLM-ensemble, i.e., to classify a given tweet into one of the relevant categories. Disagreements between annotators were resolved through majority voting, consistent with the approach used for the LLM ensemble. Annotators were paid $100 USD for their work in accordance with relevant guidelines on participant payment rates. Links to both the annotated tweet data and reproduction code will be provided after acceptance to preserve anonymity. A summary of the accuracy of the pipeline, as well as the inter-LLM and inter-rater agreement, can be found in Table 1. As shown in the table, inter-rater agreement among human annotators, measured using Krippendorff's α, ranged from 0.79 to 0.92 across the three annotation tasks, indicating substantial to near-perfect agreement. Rater-LLM agreement, measured using Fleiss's κ, ranged from 0.86 to 0.89, suggesting strong concordance between the human gold standard and the LLM ensemble. While LLMs have been shown to be politically moderate or slightly left-leaning on average (Zhou and Zhang, 2024; Motoki et al., 2024; Gover, 2023), we find that the accuracy of our classification pipeline as validated through human annotators to be sufficiently high.

3. Results

RQ1: We begin by analyzing how political candidates on either end of the political aisle frame their tweets from an ideological perspective. As can be seen in Figure 1A, of the 1,107 tweets made by Democrat political candidates, 48.1% were framed in a manner such that they are in support of the Democratic party. In contrast, 26.4% were positioned as criticisms of the Republican party, while a similar 25.5% were neutrally positioned. On the other hand, of the 128 tweets made by Republican candidates, 40.2% were positioned as in support of the Republican party, while 40.6% were positioned as criticisms of the Democratic party. Statistical tests indicate that there were no significant differences with regards to the proportion of tweets in support of a given party's ideology (chi-squared test; $\chi^2 = 1.98$, $p = 0.16$), while there was a significant difference in the proportion of tweets framed as a criticism of the opposite party. Specifically, Republican candidates authored significantly more tweets in criticism of the Democratic party than vice-versa ($\chi^2 = 11.55$, $p < 0.001$). Indeed, these results were consistent both when focusing on the period of time prior to Trump's return to Twitter on August 12th, as well as after Trump's return (Pre August 12th: AD tweets by Rep. candidates: 54%, AR tweets by Dem. candidates: 26%, χ^2 = 13.9, $p < 0.001$; Post August 12th: AD tweets by Rep. candidates: 29%, AR tweets by Dem. candidates: 19.3%, χ^2 = 4.77, $p = 0.028$).

RQ2: While candidate-authored tweets were largely positioned in support of the respective candidate's party (either pro-party or anti-opposite party), replies to such tweets did not exhibit the same pattern. As can be seen in Figure 1B, of the 36,254 replies to Democrat candidates, the majority were in opposition to the Democratic party (AD), with 69.2% of replies being classified as such per candidate tweet on average. However, only 31.7% of the 27,048 replies made to Republican candidate tweets were classified as AR on average, a significantly smaller proportion (two-sided independent t-test; $z = 15.19$, $p < 0.001$). In contrast, while 49.1% of replies made to Republican candidates were PR, only 17.2% of replies made to Democrat candidates were PD (two-sided independent t-test;

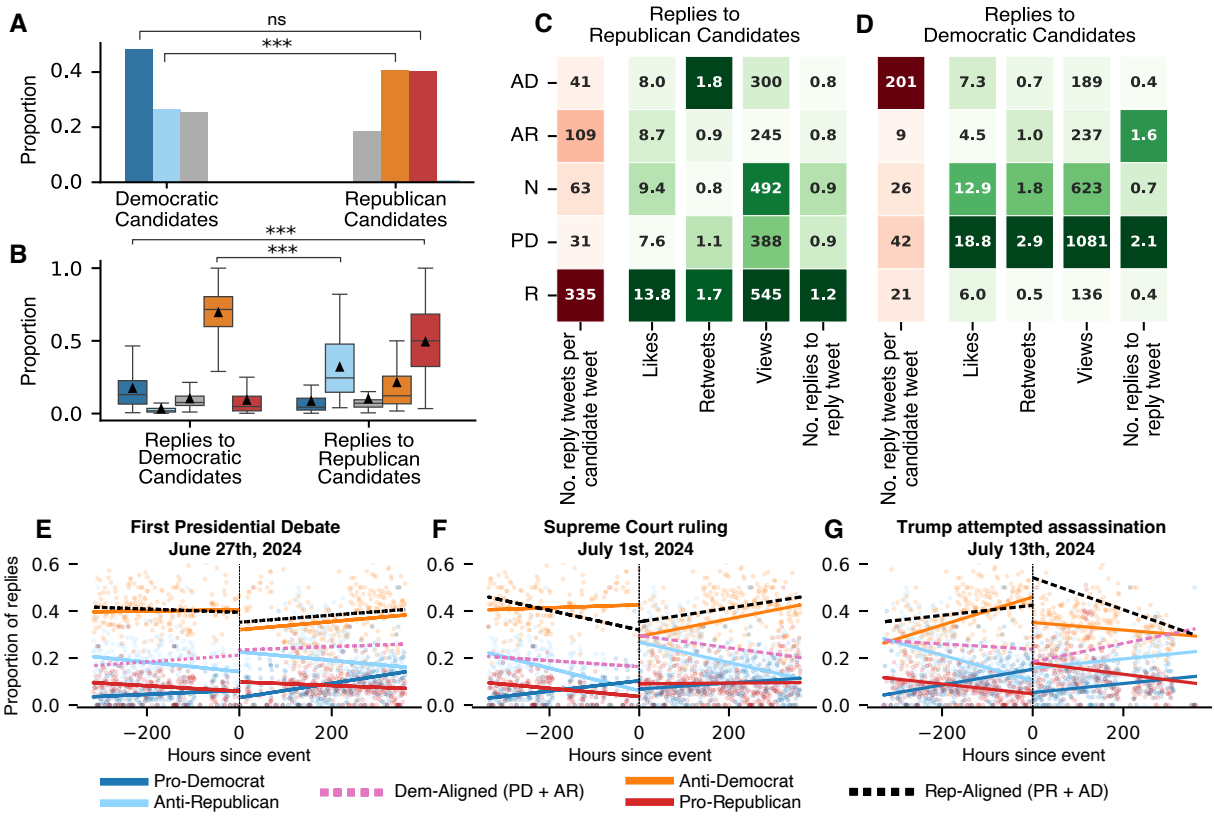

Figure 1: The proportion of candidate tweets (A) and candidate tweet replies (B) with a given ideological stance (*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$). The average number of replies of a certain stance made to (C) Democrat and (D) Republican candidates per candidate tweet, as well as the average engagement metrics (likes, retweets, views, replies) such replies receive. RDiT analysis of the ideological stance of tweets in the two weeks surrounding major political events. (E) First presidential debate between Biden and Trump; (F) Supreme Court ruling that a president has absolute immunity from criminal prosecution for core constitutional powers; (G) Trump's attempted assassination during a rally in Pennsylvania.

$z = 12.60$, $p < 0.001$), suggesting that Republican-aligned Twitter users were more active in replying to the tweets of both Democrat and Republican candidates. Indeed, this effect was amplified after Trump's return to Twitter. While there was no significant impact on the reply distribution to Democrat candidate tweets (PD: $p = 0.08$, AD: $p = 0.98$), replies to Republican candidates (i.e., both Trump and JD Vance, as opposed to only JD Vance), were significantly less likely to be of an Anti-Republican stance ($p < 0.01$), and significantly more likely to be Pro-Republican ($p < 0.01$).

RQ3: Given that replies to political candidates on both sides of the political aisle skewed Republican (either Anti-Democrat or Pro-Republican), next, we analyze the engagement these replies received. Figures 1C and 1D depict the average number of replies of a certain ideological stance made to Democrat candidates (Fig. 1C) and Republican candidates (Fig. 1D), as well as the average number of likes, retweets, views, and replies each reply tweet receives on average. Here, each column represents a separate heat map, or in other words, cell colors are computed relative to other values within the cell's column. Columns colored in red represent the average number of replies a candidate tweet receives on average, while those colored in green represent the various engagement metrics in question. As shown in Fig. 1C, for replies to Republican candidates, PR replies were both the most common type of reply (335 replies on average), and received the most engagement on average. However, for Democratic candidates, this is surprisingly not the case. As can be seen in Figure 1D, while the majority of replies received by Democratic candidates were classified as AD (201 replies on average), these replies received far less engagement than PD replies. This dichotomy in reply rate to reply engagement seen by Democratic candidates offers potential future inquiries into why this difference exists. Indeed, this could be driven by a number of factors, such as the "muting" habits of supporters of the Democratic party, bot comments, or a result of algorithm-driven reply visibility.

RQ4: To estimate the impact of major political events on the average ideological stance of tweets,

we use a regression discontinuity in time (RDiT) design which is commonly used to study the treatment effect in quasi-experiments. Specifically, we focus on three major political events in the lead up to the 2024 U.S. election, namely, the first presidential debate between Biden and Trump on June 27th, the Supreme Court's ruling on presidential immunity on July 1st, and Trump's attempted assassination on July 13th. Specifically, here, we isolate a sample of 1000 tweets per day for each day two weeks before and after a given event.

Starting with the debate (Fig. 1E), we estimate that this event did not have a significant impact on Democrat-aligned ($p = 0.41$) or Republican-aligned ($p = 0.30$) tweets as a whole. However, further delineating the type of alignment within each set shows that the debate caused a 20.1% drop in AD tweets ($p < 0.01$) coupled with a 66.0% increase in PR tweets ($p < 0.01$). Conversely, the debate caused a sharp 58.8% increase in AR tweets ($p < 0.001$) and a 49.4% decrease in PD ($p < 0.05$) tweets, suggesting differing ideological reactions to the debate from supporters of each party. With regards to the Supreme Court ruling (Fig. 1F), while we do not find a significant impact on Republican aligned tweets as a whole ($p = 0.37$), we do see a significant increase in PR tweets (148%, $p < 0.01$) and a significant drop in AD tweets (-31.6%, $p < 0.001$). On the other hand, we do find a significant change in Democrat-aligned tweets (81%, $p < 0.001$), primarily driven by a 336% increase in AR tweets ($p < 0.001$), coupled with a 34% drop in PD tweets ($p < 0.05$). Lastly, in response to Trump's attempted assassination (Fig. 1G), we find significant differences in both Democrat-aligned (-22%, $p < 0.05$) and Republican-aligned tweets (27%, $p < 0.001$). The drop in Democrat-aligned tweets was driven by 65% ($p < 0.001$) drop in PD and a 51% ($p < 0.001$) increase in AR tweets. In contrast, the rise in Republican-aligned tweets were primarily due to a 277% ($p < 0.001$) increase in PR tweets, with a 23% ($p < 0.001$) drop in AD tweets. We repeat this analysis while isolating tweets made in the 7 and 21 days surrounding each event, and find largely similar results.

Across all three events, we see two common patterns. Specifically, after each event, there was a significant increase in both AR and PR replies, coupled with significant decreases in both PD and AD replies, suggesting that discourse surrounding each event centered around support or criticism of Trump or the Republican party generally, from constituents of both parties. These results, as a whole, offer insights into the ideological positioning of tweets made both by political candidates specifically, and Twitter users broadly, in the lead-up to the 2024 U.S. election.

4. Discussion and Limitations

Due to the nature of the dataset analyzed, there are a number of limitations with regards to our analysis. Firstly, the dataset is not a comprehensive list of all political tweets on Twitter, and therefore, the results illustrated above only represent a sample of both candidate tweets and their replies. Nonetheless, given that our analysis focuses on a comparison in political alignment rates of candidate tweets and their replies, limitations based on tweet collection frequency should apply uniformly to tweets from both ends of the political spectrum.

Second, our classification pipeline relies on zero-shot prompting of three general-purpose LLMs. While the resulting accuracy and agreement with human annotators are high (Table 1), we did not compare the ensemble against supervised, fine-tuned encoder-based models such as BERTweet (**?**) or RoBERTa (**?**). Prior work has shown that such models can perform competitively on structured classification tasks while being considerably more computationally efficient. A systematic comparison between LLM-ensemble and fine-tuned approaches remains an important direction for future work.

Third, our analysis of the reply corpus does not incorporate bot detection or filtering for coordinated inauthentic behavior. Given the well-documented prevalence of automated accounts during electoral periods on Twitter/X (**?**), it is possible that some proportion of the reply-level stance distributions are influenced by non-human activity. The engagement discrepancies observed in RQ3—where Anti-Democrat replies to Democratic candidates were frequent but received relatively low engagement—may partly reflect such automated behavior. Future work should incorporate bot detection methods to assess the sensitivity of the observed patterns to the presence of automated accounts.

Fourth, our analysis is conducted exclusively on Twitter/X, which underwent significant algorithmic and policy changes throughout 2024 under its new ownership, including modifications to content moderation practices, verification systems, and recommendation algorithms. These platform-level shifts may have influenced content visibility, user engagement patterns, and the composition of replies in ways that are difficult to disentangle from genuine ideological dynamics. Extending this analysis to other platforms such as TikTok, Bluesky, or Truth Social would help contextualize the extent to which the observed patterns are specific to Twitter/X or reflective of broader online political discourse.

Fifth, this study focuses exclusively on the two major U.S. political parties and their candidates. Third-party candidates—such as those from the Green or Libertarian parties—were not included

in the analysis. While these candidates received comparatively less attention on Twitter/X during the 2024 election cycle, their exclusion means that our five-category classification scheme may not fully capture the breadth of ideological positioning present in online discourse. Future work could incorporate a broader set of political actors to provide a more comprehensive picture of the ideological landscape.

Future work may explore both the detection and influence of bot accounts in the context of promoting particular political ideologies. Furthermore, future work could examine the evolution of ideological positioning outside of the two-party system, incorporating third-party candidates and stances. Moreover, while our results aim to capture ideological positioning on Twitter, studies could look to collect and analyze such sentiments on other social media platforms.

5. Ethics

All data analyzed in this study is publicly available. Informed consent was retrieved from the human annotators who validated model outputs. We confirm that all text in this paper was written by the authors. AI-based writing assistants (Grammarly and Writeful) were used solely for grammar and spelling checks and to improve the clarity of the author-written text. The content and intellectual contributions remain entirely those of the human authors.

6. Bibliographical References

Abeer Aldayel and Walid Magdy. 2019. Your stance is exposed! analysing possible factors for stance detection on social media. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):1–20.

Ashwin Balasubramanian, Vito Zou, Hitesh Narayana, Christina You, Luca Luceri, and Emilio Ferrara. 2024. A public dataset tracking social media discourse about the 2024 us presidential election on twitter/x. *arXiv preprint arXiv:2411.00376*.

Andrei Boutyline and Robb Willer. 2017. The social structure of political echo chambers: Variation in ideological homophily in online networks. *Political psychology*.

Matteo Cinelli, Gianmarco De Francisci Morales, Alessandro Galeazzi, Walter Quattrociocchi, and Michele Starnini. 2021. The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, 118(9):e2023301118.

Fabrizio Gilardi, Meysam Alizadeh, and Maël Kubli. 2023. Chatgpt outperforms crowd workers for text-annotation tasks. *PNAS*, 120(30):e2305016120.

Lucas Gover. 2023. Political bias in large language models. *The Commons: Puget Sound Journal of Politics*, 4(1):2.

Hazem Ibrahim, Nouar AlDahoul, Sangjin Lee, Talal Rahwan, and Yasir Zaki. 2023. Youtube's recommendation algorithm is left-leaning in the united states. *PNAS nexus*.

Fabio Motoki, Valdemar Pinho Neto, and Victor Rodrigues. 2024. More human than human: measuring ChatGPT political bias. *Public Choice*, 198(1):3–23.

Peter Stefanov, Kareem Darwish, Atanas Atanasov, and Preslav Nakov. 2020. Predicting the topical stance and political leaning of media using tweets. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*.

Petter Törnberg. 2024. Large language models outperform expert coders and supervised classifiers at annotating political social media messages. *Social Science Computer Review*, page 08944393241286471.

Patrick Y Wu, Jonathan Nagler, Joshua A Tucker, and Solomon Messing. 2023. Large language models can be used to scale the ideologies of politicians in a zero-shot learning setting, april 2023. *arXiv preprint arXiv:2303.12057*.

Jinyi Ye, Luca Luceri, and Emilio Ferrara. 2024. Auditing political exposure bias: Algorithmic amplification on twitter/x approaching the 2024 us presidential election. *arXiv preprint arXiv:2411.01852*.

Di Zhou and Yinxian Zhang. 2024. Political biases and inconsistencies in bilingual GPT models—the cases of the US and China. *Scientific Reports*, 14(1):25048.

Caleb Ziems, William Held, Omar Shaikh, Jiaao Chen, Zhehao Zhang, and Diyi Yang. 2024. Can large language models transform computational social science? *Computational Linguistics*, 50(1):237–291.