Exploring Political Attitudes under Selective Media Consumption with Large Language Models

Josh Nguyen,[*] Upasana Dutta, Samar Haider, Bryan Li, Neil Fasching, and Duncan J. Watts

Keywords: Media Bias, News Consumption, Public Opinion, Language Models, Simulation

Extended Abstract

Whether through "echo chambers" (1) or "filter bubbles" (2), scholars have argued that *selective exposure* of news media—wherein people intentionally consume news that matches their preferences while avoiding potential counter-viewpoints (3, 4)—can lead to a litany of undesirable effects: an overall less-informed public (5), polarized individual-level attitudes (6), influence on political behavior and vote choice (7), and creation of divergent "shared realities" among exposed users (8). However, others argue that selective exposure is either overstated (9, 10) or does not have polarizing effects (11). Large-scale field experiments are required to adjudicate these competing claims, but they face prohibitive costs and many implementation challenges. To date, the only major field experiment in this area has focused on the effect of consuming cross-cutting media rather than selective exposure specifically (12).

Advances in artificial intelligence (AI), particularly large language models (LLMs), provide researchers with new tools to tackle the logistical obstacles of field experiments by substituting LLMs for human subjects. By synthesizing their training data, LLMs can reliably internalize political bias as well as the aggregate opinion of people who may consume the content in that data (13, 14). Moreover, recent work has reproduced both micro- and macro-behaviors of humans using LLMs such as moral judgments (15), economic decisions (16), voting preferences (17), and network homophily (18). Building upon this literature, we offer a first step toward establishing empirical evidence for the influence of selective exposure to partisan content from the news media on human attitudes using LLMs as representative surrogates.

We use data from the Media Bias Detector platform (19), which tracks ten mainstream U.S. news outlets since January 2024 by scraping their top-20 articles every four hours. We select all content classified under politics and published in October 2024, which totals $N = 10,127$ articles. Unsurprisingly, the most popular topic covered in this period was the 2024 U.S. presidential election ($N = 4,679$). We focus on two prominent and polarizing topics: immigration ($N = 440$) and crime ($N = 1,064$); see Figure 1. We further select only articles published by four outlets: *HuffPost*, *CNN*, *Fox News*, and *Breitbart*. To test whether selective exposure to a specific news outlet would influence one's attitudes about immigration and crime toward that outlet's ideological leaning, we introduce an LLM-based experimental setting as follows.

For each topic, we use identical copies of GPT-4o (20) as independent "subjects" in our experiment. Each subject is assigned to either a control or one of five treatment conditions ($N = 100$ subjects per condition). In the treatment conditions, subjects are asked to "read" 20 randomly sampled articles, presented in chronological order. In four of such conditions, subjects are presented articles by *only* one publisher, such as *CNN*. The other treatment condition, called "All Publishers," exposes subjects to 20 articles, five of which are from *each* of the four outlets. In the control, we present no articles. Afterward, all subjects are prompted to indicate their

[*] Corresponding author. Email: joshtn@seas.upenn.edu.

agreement—on a scale of 1 to 7, where 4 is neutral—with six statements related to the focal topic; see Figure 2. Two of these statements are conservative-leaning, two are liberal-leaning, and two are neutral.

The average rating of these statements within each condition is depicted in Figure 3. Our first observation is that, via the control group, GPT-4o-generated subjects exhibit an inherent liberal-leaning stance on both topics. These subjects tend to agree, often highly, with positions that are pro-immigration and social justice-oriented when it comes to crime. Secondly, when exposed to articles from a single source, subjects' average ratings tend to align closely with that source's ideological leaning. For instance, those who read only *Fox News* and *Breitbart* predominantly agree with stricter immigration and criminal policies, with one exception that they tend to agree that investing in social services is an effective way to prevent crime. Those who read *CNN* and *HuffPost* often demonstrate the opposite stance on these statements. Thirdly, we observe a qualitative moderating effect in the "All Publishers" condition, in which subjects' positions tend to be closer to the neutral line than those of the control group.

Does only reading *HuffPost*'s coverage in this period "cause" a subject to have more liberal views about immigration and crime? To estimate this "treatment effect," for each statement we fit a regression model predicting the response to that statement using the experimental conditions. As shown in Figures 4 and 5 (left), this exposure significantly causes subjects to change their attitudes. For instance, subjects in all conditions are much less likely, compared to the control, to agree that investing in social services is effective for reducing crime, thereby moderating the baseline attitude. While subjects assigned to the *CNN* and *HuffPost* conditions display no difference from the control on their position on stricter sentencing laws, *Fox News* and *Breitbart* readers are much more likely to agree with this statement. Additionally, the moderating effect of the "All Publishers" condition is evident in the positive regression coefficients for conservative-leaning statements and negative values for liberal-leaning claims.

Finally, we examine the robustness of these findings when using another LLM, LLaMA-3.1-8B (21), as an alternative to GPT-4o. The treatment effects are presented in Figures 4 and 5 (right). First, we find that most directional effects are preserved, especially for subjects which are exposed to *Breitbart* and *Fox News* articles. For instance, GPT- and LLaMA-generated subjects in these conditions are significantly more likely to agree that lenient criminal justice reforms have led to an increase in crime, compared to the control group. Interestingly, we find that the *HuffPost* and *CNN* groups display almost *no* difference from the control, suggesting that reading articles from these outlets does *not* alter the baseline attitudes of LLaMA subjects. Finally, we observe that most effect sizes are smaller in the LLaMA setting than in the GPT setting, even though their directions are largely preserved or remain insignificant.

Our work provides a clear, easily implementable and flexible framework to examine the effect of selective exposure on political attitudes. In future work, we aim to explore the following directions. First, we are considering a broader selection of LLMs as human surrogates. The two models studied here have demonstrated a liberal bias via the baseline attitudes of the control group, which is consistent with several recent findings (13, 22). Experimenting with other LLMs may reveal further insights into the effects of selective exposure, especially since their baseline responses may be biased in a different direction. Second, we are extending our analysis to a larger time horizon, covering more major events across more news outlets and involving more subjects. Most importantly, while the treatment effect reported here is pronounced for LLM-simulated subjects, it remains unclear whether this also holds for real humans. If so, can we generalize from these findings to make predictions about people's attitudes toward other statements and topics? Regarding robustness, we have found qualitatively that GPT-4o and LLaMA-3.1-8B give largely similar results, especially in the direction of the treatment effects.

However, effect sizes are highly different between the two models; future work will shed light on the accuracy of these models using the experimental human ground truth, and explore whether aggregating these models in a wisdom-of-crowd fashion (23) could be a superior approach.

References

1. Cass R Sunstein. *Republic.com 2.0*. Princeton University Press, Princeton, NJ, 2009.

2. Eli Pariser. *The Filter Bubble: What the Internet Is Hiding from You*. Penguin Press, New York, NY, 2011.

3. Larry M. Bartels. Beyond the running tally: Partisan bias in political perceptions. *Political Behavior*, 24(2):117–150, 2002.

4. Natalie Jomini Stroud. *Niche news: The politics of news choice*. Oxford University Press, 2011.

5. Natalie Jomini Stroud. Polarization and partisan selective exposure. *Journal of Communication*, 60(3):556–576, 2010.

6. Ro'ee Levy. Social media, news consumption, and polarization: Evidence from a field experiment. *American Economic Review*, 111(3):831–870, 2021.

7. Stefano DellaVigna and Ethan Kaplan. The fox news effect: Media bias and voting. *The Quarterly Journal of Economics*, 122(3):1187–1234, 2007.

8. Homa Hosseinmardi, Samuel Wolken, David M Rothschild, and Duncan J Watts. The diminishing state of shared reality on us television news. *arXiv preprint arXiv:2310.18863*, 2023.

9. Andrew M Guess. (almost) everything in moderation: New evidence on americans' online media diets. *American Journal of Political Science*, 65(4):1007–1022, 2021.

10. Matthew Gentzkow and Jesse M Shapiro. Ideological segregation online and offline. *The Quarterly Journal of Economics*, 126(4):1799–1839, 2011.

11. Brendan Nyhan et al. Like-minded sources on facebook are prevalent but not polarizing. *Nature*, 620(7972):137–144, 2023.

12. David E Broockman and Joshua L Kalla. Consuming cross-cutting media causes learning and moderates attitudes: A field experiment with fox news viewers. *The Journal of Politics*, 87(1), 2025.

13. Shangbin Feng, Chan Young Park, Yuhan Liu, and Yulia Tsvetkov. From pretraining data to language models to downstream tasks: Tracking the trails of political biases leading to unfair nlp models. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 11737–11762, 2023.

14. Eric Chu, Jacob Andreas, Stephen Ansolabehere, and Deb Roy. Language models trained on media diets can predict public opinion. *arXiv preprint arXiv:2303.16779*, 2023.

15. Danica Dillion, Niket Tandon, Yuling Gu, and Kurt Gray. Can AI language models replace human participants? *Trends in Cognitive Sciences*, 27(7):597–600, 2023.

16. Apostolos Filippas, John J. Horton, and Benjamin S. Manning. Large Language Models as Simulated Economic Agents: What Can We Learn from Homo Silicus? In *Proceedings of the 25th ACM Conference on Economics and Computation*, pages 614–615, New Haven CT USA, 2024. ACM.

17. Lisa P. Argyle et al. Out of One, Many: Using Language Models to Simulate Human Samples. *Political Analysis*, 31(3):337–351, 2023.

18. James K. He, Felix P. S. Wallis, Andrés Gvirtz, and Steve Rathje. Artificial intelligence chatbots mimic human collective behaviour. *British Journal of Psychology*, page bjop.12764, 2024.

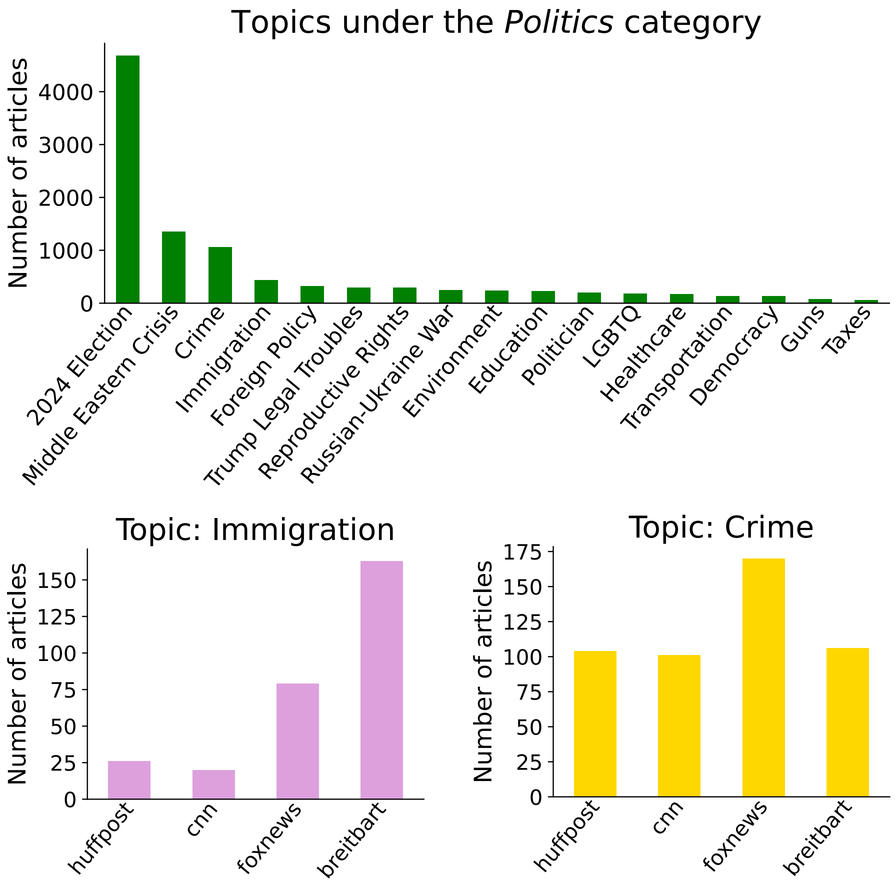
19. Jenny S Wang et al. Media bias detector: Designing and implementing a tool for real-time selection and framing bias analysis in news coverage. *arXiv preprint arXiv:2502.06009*, 2025.

20. Aaron Hurst et al. Gpt-4o system card. *arXiv preprint arXiv:2410.21276*, 2024.

21. Abhimanyu Dubey et al. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*, 2024.

22. Shibani Santurkar et al. Whose opinions do language models reflect? In *International Conference on Machine Learning*, pages 29971–30004. PMLR, 2023.

23. Philipp Schoenegger, Indre Tuminauskaite, Peter S. Park, Rafael Valdece Sousa Bastos, and Philip E. Tetlock. Wisdom of the silicon crowd: LLM ensemble prediction capabilities rival human crowd accuracy. *Science Advances*, 10(45):eadp1528, 2024.



Topic	Publisher	Title
Immigration	*HuffPost*	Trump Sparks Outrage With 'Bad Genes' Comment About Immigrants
	CNN	Trump says he would revoke Temporary Protected Status for Haitian migrants in Springfield if elected
	Fox News	'It's a mess': Vulnerable House Dem shreds Biden on border crisis in 'closing message' of campaign
	Breitbart	Democrat Strategist: Progressive 'Moment Is Well and Truly Over' Thanks to Open-Border Policies
Crime	*HuffPost*	Teen Girl's Dismembered Remains Found In Freezer Left Behind By Home's Previous Owner
	CNN	Active shooter at Atlanta hotel is taken into custody after multiple shots fired
	Fox News	Chinese citizen charged after allegedly voting illegally in key battleground state; vote will be counted
	Breitbart	Exclusive--'I Was Witnessing History': Butler, PA, Vendor Recounts 'Scary' Trump Assassination Attempt

Figure 1: Summary of the dataset. The bar plot on top shows the number of articles about each topic in the politics category, published in October 2024. The middle two bar plots show the number of articles about immigration or crime published by each of the four outlets. The table at the bottom shows some example titles in each topic.

```
You are about to read 20 online news articles.  After that, you will
be asked to indicate your opinion about 6 statements.  Here are the
articles:

#################
--- ARTICLE 1 ---
Date:
Title:
Body:
.
.
.
--- ARTICLE 20 ---
Date:
Title:
Body:
#################

TASK:
Below are 6 statements.  Indicate your opinion about these statements on
a 7-point Likert scale, where
1 = Strongly Disagree
2 = Disagree
3 = Somewhat Disagree
4 = Neither Agree nor Disagree
5 = Somewhat Agree
6 = Agree
7 = Strongly Agree.

Statements:
a.
b.
c.
d.
e.
f.

In total, there are 6 statements, and you should provide your agreement
with each one of them line by line.  For each statement, first provide
your Likert-scale response as ‘RESPONSE: [response]’, then in the next
line provide a short one-sentence explanation for your response as
‘EXPLANATION: [explanation]’.  Do not return anything other than the
response and explanation for each statement.
```

Figure 2: Prompt used for LLMs. Texts in black are common to both control and treatment conditions, whereas texts in blue are only for the treatment. In the treatment conditions, 20 articles are randomly sampled by a target outlet, such as *CNN*. Then, they are presented to the subject in chronological order. We do not reveal the outlet to the subject (e.g., that the articles it reads are from *CNN*), but the content within an article might help identify this information.

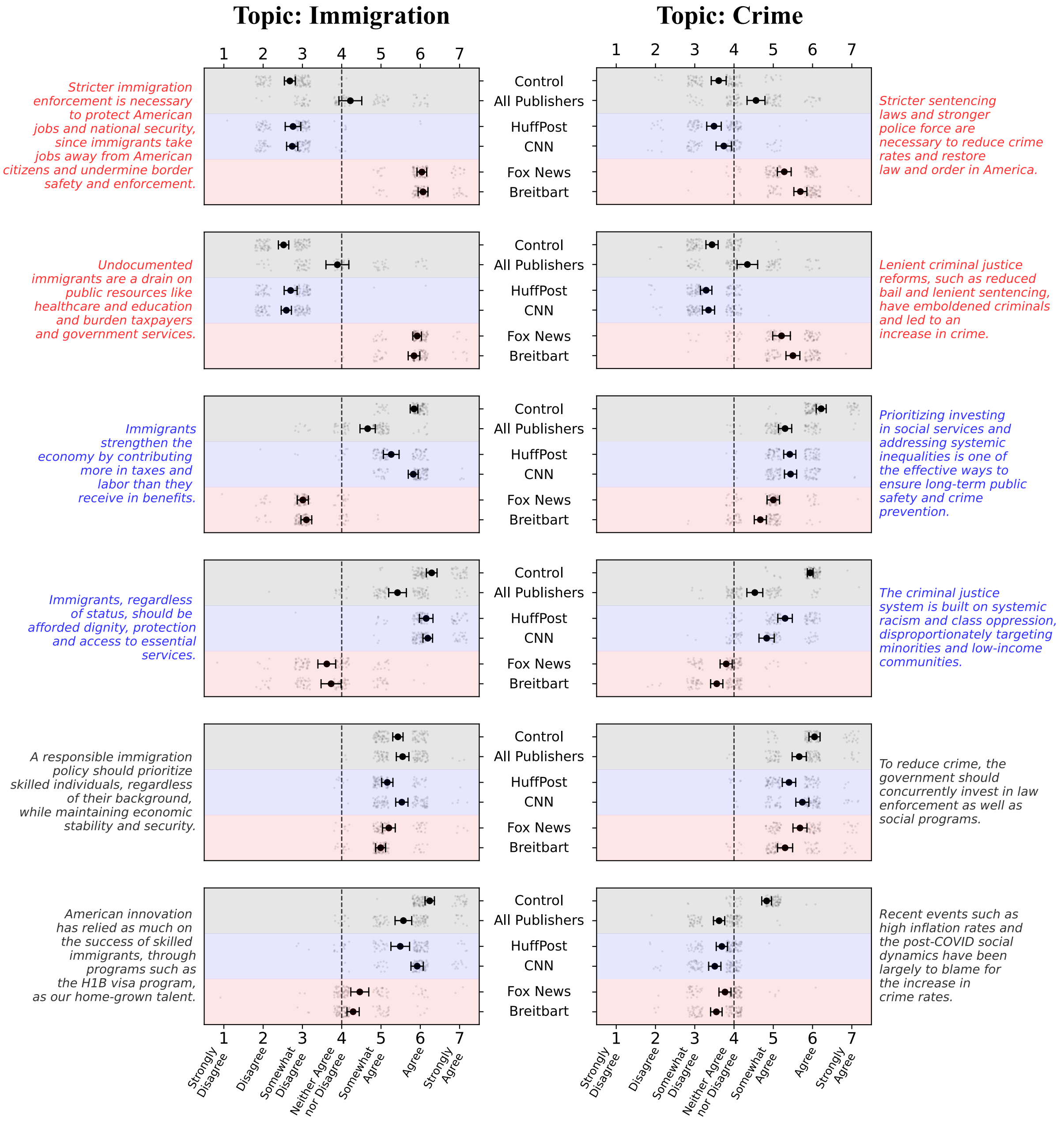


Figure 3: Attitudes toward immigration and crime. For each topic, we ask subjects to indicate, on a scale of 1 to 7, their agreement with six statements. Two of these statements have a conservative leaning (in red), two have a liberal leaning (in blue), and the other two have a neutral leaning (in black). In the control condition, we ask for GPT-4o's attitudes without giving it any article. In the "All Publishers" condition, we ask GPT-4o to "read" 20 randomly sampled articles from 4 outlets, *HuffPost*, *CNN*, *Fox News* and *Breitbart*, before asking for its attitudes. In the four other conditions, we ask GPT-4o to "read" 20 randomly sampled articles but *only* by the corresponding outlet, such as *CNN*. There are $N = 100$ subjects per condition. Displayed are all answers of these subjects toward each statement, as well as their estimated means and 99% confidence intervals.

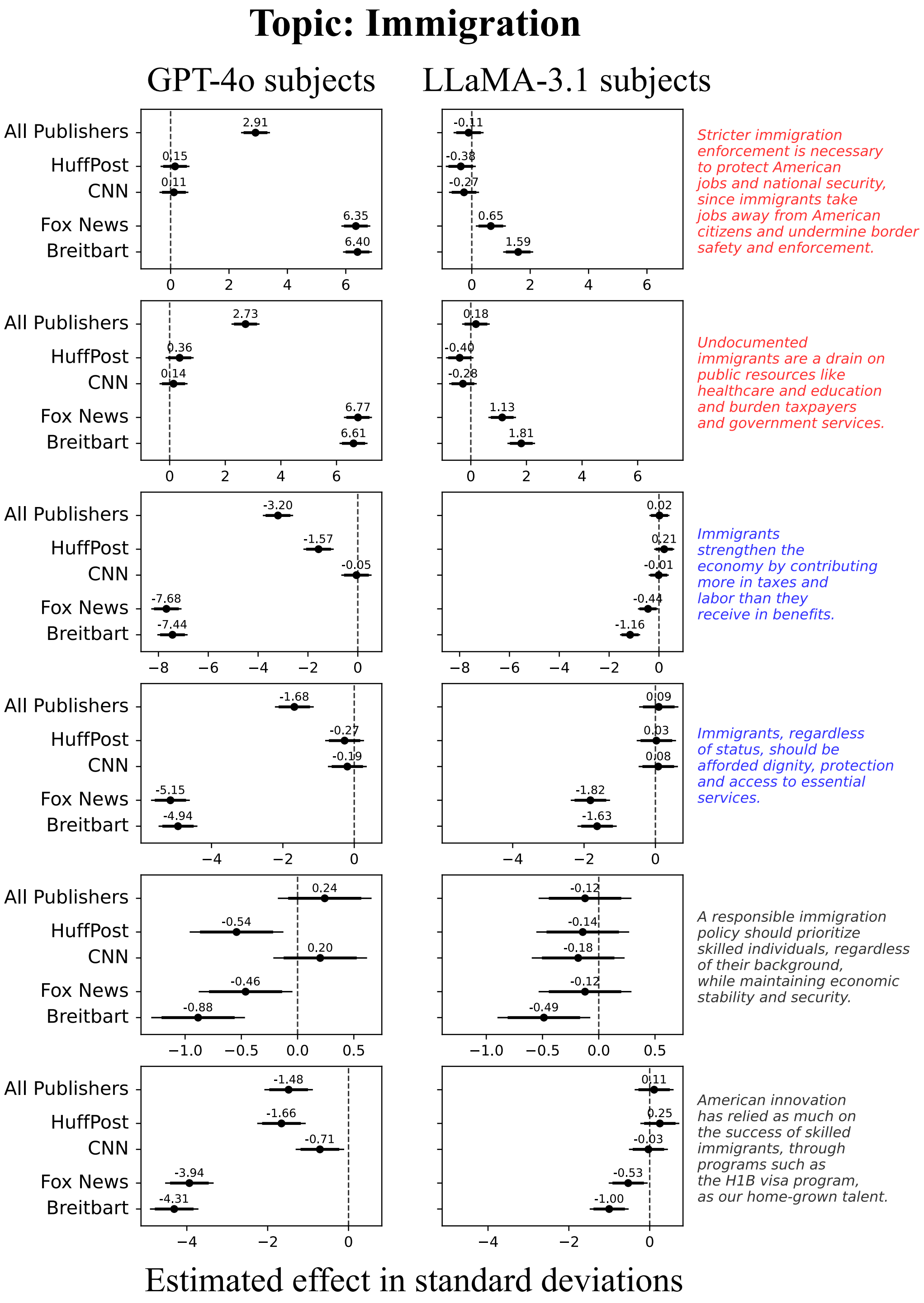
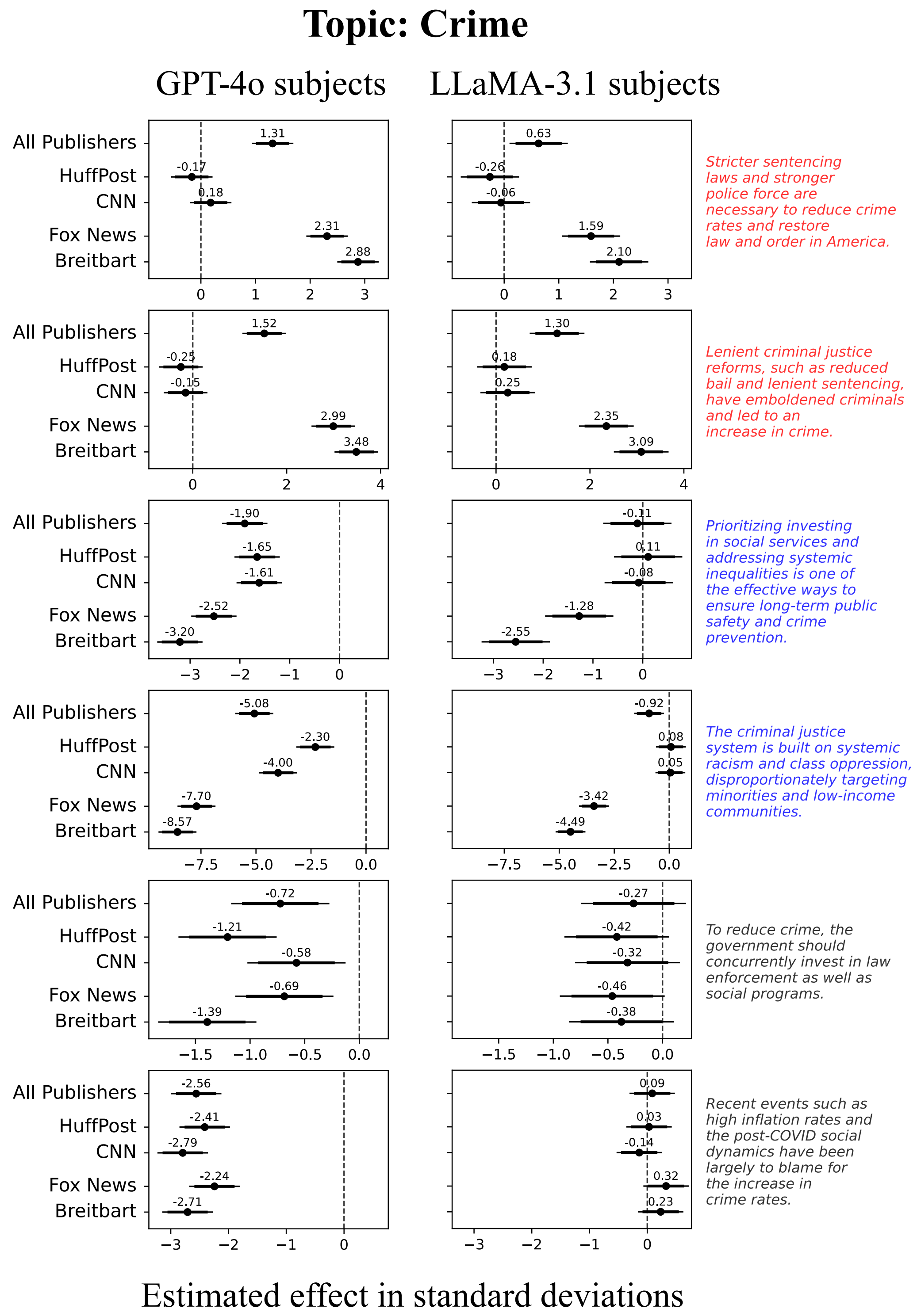


Figure 4: Estimated effect of exposure to partisan news content on attitudes about **immigration**. For each statement, we standardize all responses using the mean and standard deviation of the control group's responses. Then, we fit an ordinary least-squares model of the form $y_i = \alpha + \boldsymbol{\beta}_i^T \boldsymbol{X}_i + \varepsilon_i$, where y_i is subject i's response to the statement (after standardization), $\boldsymbol{X}_i$ is the 5-vector of dummies representing the condition subject i is in (with "Control" as reference), and $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$ is an unbiased error. Displayed in this figure are estimates of the coefficients $\boldsymbol{\beta}_i$ as well as their 95% (thick bars) and 99% (thin bars) confidence intervals. A positive coefficient of δ implies that subjects in that condition are δ standard deviations more likely than the control group to agree with the corresponding statement.



Topic: Crime

GPT-4o subjects LLaMA-3.1 subjects

Stricter sentencing laws and stronger police force are necessary to reduce crime rates and restore law and order in America.

Lenient criminal justice reforms, such as reduced bail and lenient sentencing, have emboldened criminals and led to an increase in crime.

Prioritizing investing in social services and addressing systemic inequalities is one of the effective ways to ensure long-term public safety and crime prevention.

The criminal justice system is built on systemic racism and class oppression, disproportionately targeting minorities and low-income communities.

To reduce crime, the government should concurrently invest in law enforcement as well as social programs.

Recent events such as high inflation rates and the post-COVID social dynamics have been largely to blame for the increase in crime rates.

Estimated effect in standard deviations

Figure 5: Estimated effect of exposure to partisan news content on attitudes about **crime**. The calculation of the effect sizes is identical to that in Figure 4.