Elections in the age of algorithms: evidence of political bias in search engines and large language models

Íris Damião[a,b], Paulo Almeida[a] , João Franco[a], Nuno Santos[c], Pedro C. Magalhães[d], and Joana Gonçalves-Sá[a,e]

[a] *Social Physics and Complexity Lab (SPAC), Laboratório de Instrumentação e Física Experimental de Partículas (LIP), Lisbon, Portugal*
[b] *Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal*
[c] *INESC-ID, Lisbon, Portugal*
[d] *Instituto de Ciências Sociais da Universidade de Lisboa, Lisbon, Portugal*
[e] *NOVA LINCS - NOVA Laboratory for Computer Science and Informatics, NOVA School of Science and Technology, NOVA University Lisbon, Lisbon, Portugal*

1. Introduction

Search engines (SEs) serve as main information gateways for billions of users worldwide (Semrush, 2026), and large language models (LLMs) are rapidly emerging as alternatives (Robarts, 2025). Although widely perceived as trustworthy (Pan et al., 2007), both algorithms rely on opaque, proprietary models that have been shown to produce problematic outputs. Search engines can exhibit racial and gender biases (Kay et al., 2015; Noble, 2018), while LLMs can generate inaccurate information (Huang et al., 2023) and reproduce biases embedded in their training data (Silberg & Manyika, 2019). As more people rely on online search for politics (Pew Research Center, 2025), this study examines how these algorithms shape access to election-related information, particularly during elections, when biased search rankings can shift the voting intentions of up to 20% of undecided voters (Epstein & Robertson, 2015; Epstein et al., 2025).

Hence, this study audits whether SEs and LLMs display political bias when responding to neutral, election-related queries in the days preceding two major 2024 elections: the European Parliament elections (June 9) and the US Presidential elections (November 5). Unlike prior work relying on explicitly political queries, sentiment analysis, or classification of media outlet leaning (Robertson et al., 2018; Gezici et al., 2021), we measure political visibility—the frequency with which political entities and issues are mentioned—using neutral, naturalistic queries that better reflect undecided voters' actual search behavior (Mustafaraj et al., 2020). We focus on visibility because it influences viability (Hopmann et al., 2010; Adams et al., 2024), and because even negative coverage has been shown to normalize fringe actors and ideas by signaling their relevance (Mancosu et al., 2021).

Methodology. We developed a privacy-preserving, bot-based auditing system in which automated web crawlers simulated human browsing behavior while operating from residential proxies in multiple geographic locations. Bots simultaneously queried four SEs—Google, Bing, DuckDuckGo, and Yahoo — and two LLMs — Microsoft Copilot (with internet access) and ChatGPT-4o (without internet access) — using neutral queries that did not mention politicians, political leanings, or ideological positions, in both English and local languages (e.g., "European Parliament elections 2024", "Who should I vote for?"). For the EU election, bots were deployed in five countries: Austria, Germany, Ireland, Poland, and Portugal. For the US, bots operated from 15 counties across eight states with varying partisan leanings. In total, we collected approximately 4,360 SE first-page results (six queries in the EU, in English and local languages, and 12 English-language queries in the US) and 205 LLM responses (4 queries per election), see Figure 1.

Political entities were extracted from SE URLs and headlines, and from full LLM responses, using automated LLM-based extraction with manual validation. For the EU, entities were mapped to five ideological groups — Radical Left, Mainstream Left, Greens, Mainstream Right, and Radical Right —

SEASON 2026, September 15–17, 2026, Hamburg, Germany
DOI: https://doi.org/10.48441/4427.3718

based on the Chapel Hill Expert Survey (Rovny et al., 2024). For the US, entities were classified as Democratic or Republican, and political issues were categorized by partisan salience using pre-electoral Pew Research (Pew Research Center, 2024) and YouGov (YouGov, 2025) surveys. Mention frequencies were benchmarked against media salience, pre-election polls, and prior electoral results.

Results. In the EU, political mentions appeared in 7–16% of SE results (grey bars in Figure 2), yet their distribution was highly skewed. Across all four SEs and ChatGPT, Radical Right entities were dramatically overrepresented relative to all external benchmarks (Figure 3). For Google, the Radical Right received 16–32 percentage points more mentions than expected from media coverage, polling, or prior electoral results—consistent across countries including Ireland and Portugal, where radical right parties had minimal electoral presence. Radical right entities appeared disproportionately in News and Reference results (e.g., Wikipedia), that is, sources perceived as authoritative. The Radical Left was significantly underrepresented across all platforms.

LLM responses were more balanced overall, but both Copilot and ChatGPT overrepresented the Radical Right relative to polls and prior results, while giving disproportionate attention to Green parties—a pattern inconsistent with electoral data, potentially reflecting fine-tuning and content moderation. This diverges from prior work characterizing LLMs as left-leaning (Potter et al., 2024; Motoki et al., 2024), likely because we use neutral prompts and measure visibility rather than eliciting political opinions.

In the US, party-level differences were smaller, consistent with the two-party system. Nevertheless, Republican entities were systematically overrepresented and Democratic entities underrepresented across all SEs relative to polls and prior results. More striking was the issue framing: Google dedicated 83% of classified issue mentions to Republican-prioritized topics, an overrepresentation of approximately 47 percentage points relative to voter-assigned importance. Bing, DuckDuckGo, and Yahoo were substantially more balanced. Both ChatGPT and Copilot also mentioned Republican-prioritized topics more frequently, though less extremely than Google.

Conclusions. Our results provide systematic evidence that both SEs and LLMs display political bias in response to neutral queries. This observed visibility bias cannot be fully explained by external factors: overrepresentation of right-leaning entities exceeds what media salience, polling, or prior results would predict, suggesting that algorithmic decisions actively amplify existing patterns (Norocel & Lewandowski, 2023). Because visibility and user engagement reinforce one another (Ekström et al., 2024), disentangling the origins of these patterns remains difficult, highlighting the need for independent auditing.

Limitations include a query set that does not capture the full diversity of real-world search behavior, a relatively small LLM sample, and rapidly evolving systems. Future longitudinal studies are needed, including across additional political systems and languages, particularly given the wide variation in LLM refusal rates by language observed here (Urman & Makhortykh, 2025).

However, regardless of the underlying mechanisms, algorithms used daily by millions of citizens seem to systematically give greater visibility to certain parts of the political spectrum than others. Research in political communication has consistently shown that visibility itself is a key determinant of political awareness and perceived viability (Hopmann et al., 2010; Adams et al., 2024), underscoring the need for greater transparency and independent oversight of these prominent platforms, as reflected in the EU's Digital Services Act, which mandates the assessment of systemic risks to democratic processes.

SEASON 2026, September 15–17, 2026, Hamburg, Germany
DOI: https://doi.org/10.48441/4427.3718

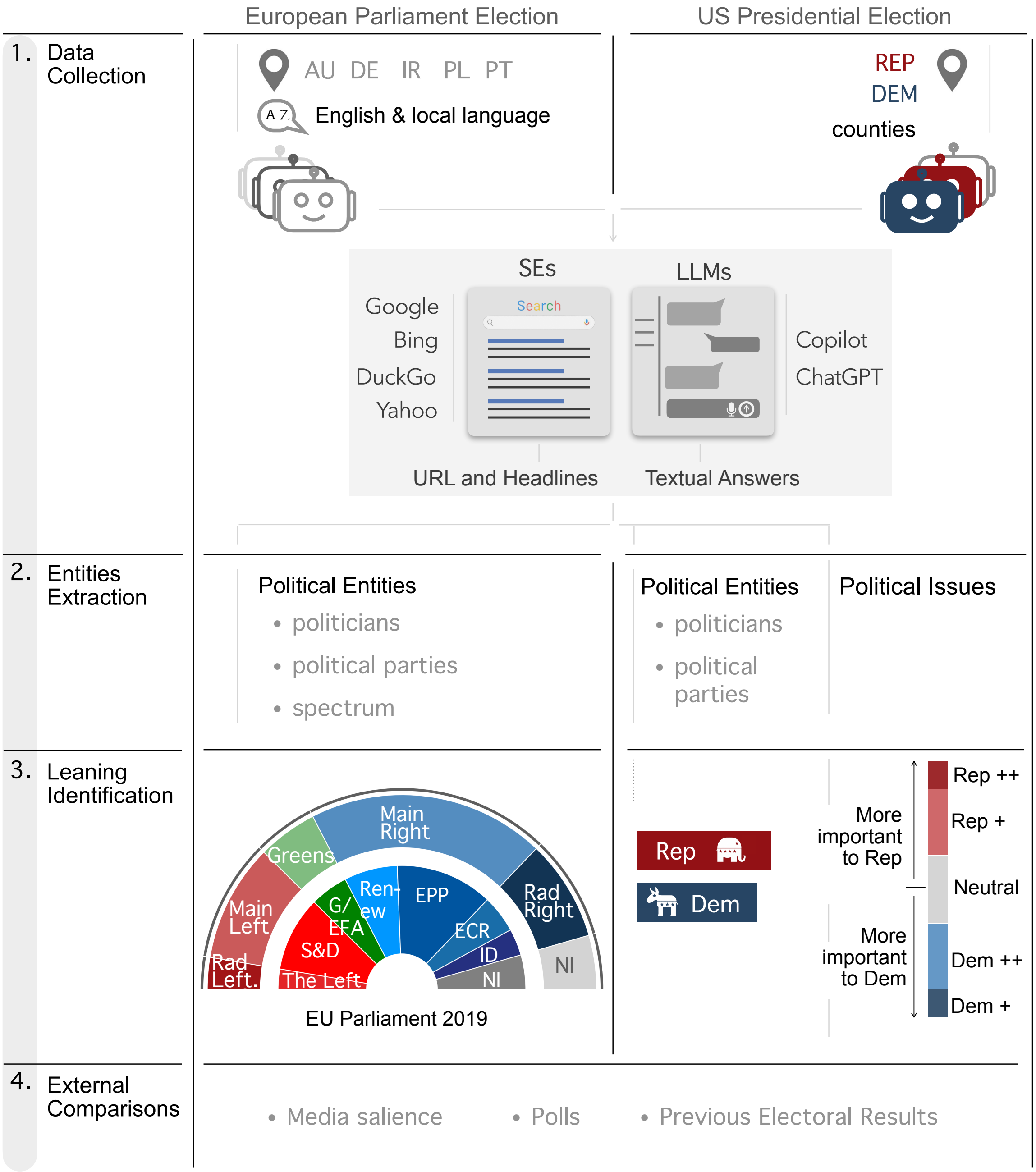

Figure 1: **Methodological scheme. (1)** Bots collected data from search engines (Google, Bing, DuckDuckGo, Yahoo) and LLMs (Copilot, ChatGPT), from different locations within the European Union (Austria - AU, Germany - DE, Ireland - IR, Poland - PL, and Portugal - PT) or from different counties in the United States of America. For LLMs, location was simulated by prompting the model in the most spoken language of each country (EU only). **(2)** Political entities were extracted from URLs and headlines for SEs, and from textual answers for LLMs. **(3)** Entities were mapped to ideological leanings (EU) or partisan issue associations (US). **(4)** Observed leanings were compared with media salience, polls, and prior electoral results.

SEASON 2026, September 15–17, 2026, Hamburg, Germany
DOI: https://doi.org/10.48441/4427.3718

SEASON 2026, September 15–17, 2026, Hamburg, Germany
DOI: https://doi.org/10.48441/4427.3718

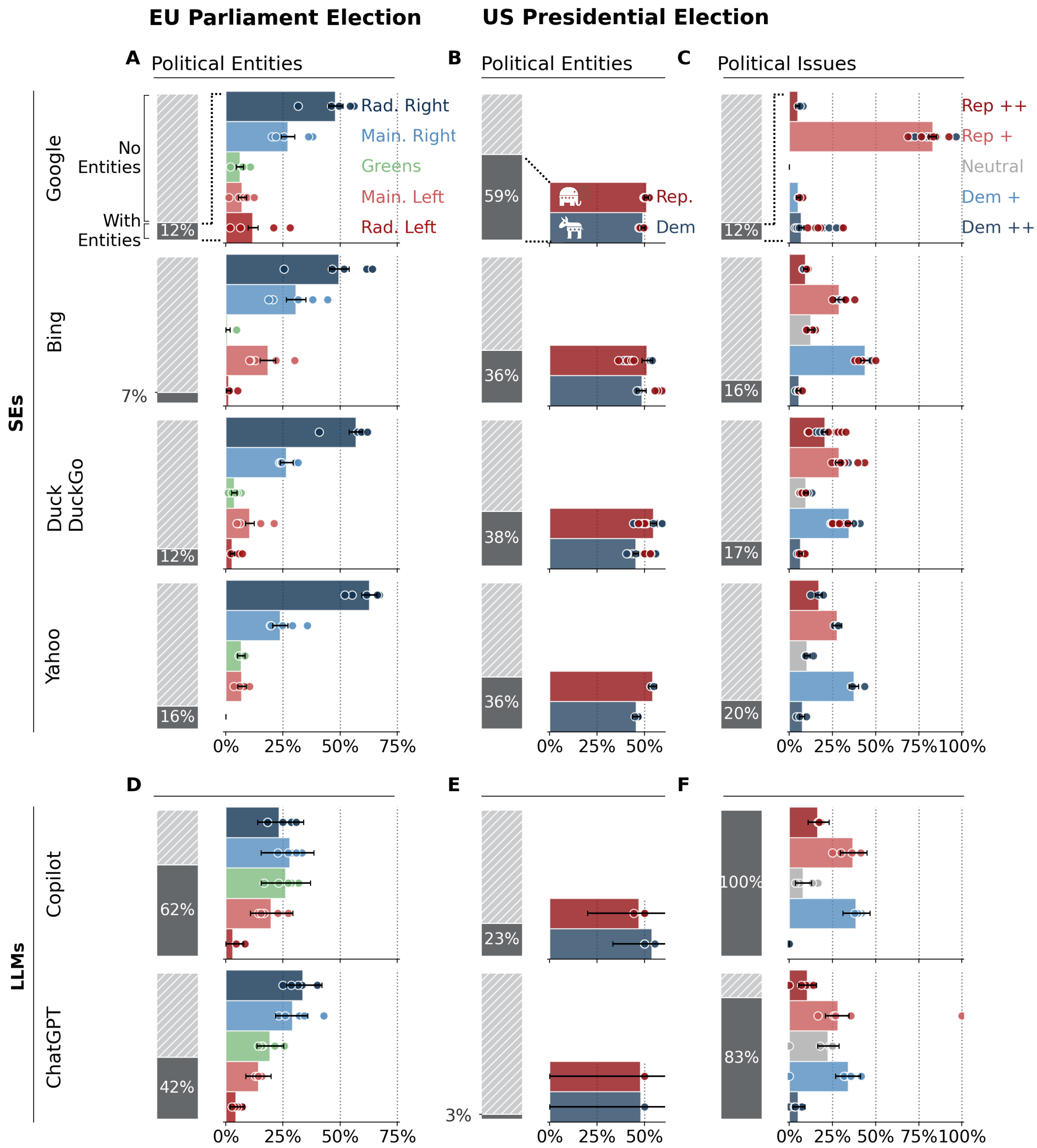

Figure 2: Search engine and LLM results classified by political leaning. Dark gray bars show the proportion of SE results or LLM answers containing relevant political mentions (political entities or issues). Colored bars show the average distribution of those mentions across political leanings or issue categories.

SEASON 2026, September 15–17, 2026, Hamburg, Germany
DOI: https://doi.org/10.48441/4427.3718

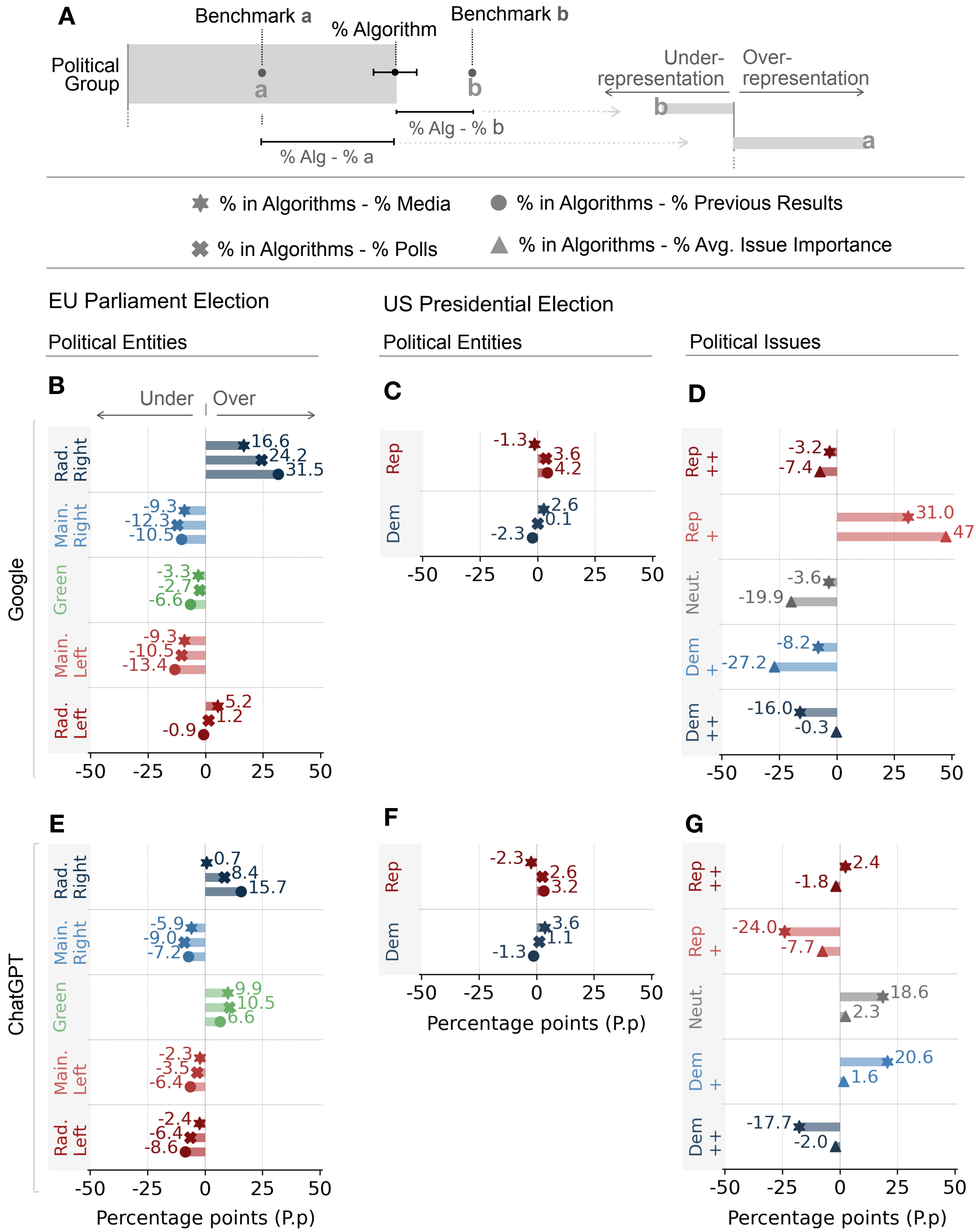

Figure 3: Comparison with external factors. (A) Explanatory scheme of the differences displayed in **(B-G)** Star = % in Algorithms -% Media; Circle = % in Algorithms - % Previous Results; Cross = % Algorithms - % Polls; Triangle = % Algorithms - % Average issue importance. Positive values indicate overrepresentation. **(B-D)** Google attention to EU and US political leanings vs. external benchmarks. **(E-G)** ChatGPT attention to EU and US political leanings and issues vs. the same benchmarks.

SEASON 2026, September 15–17, 2026, Hamburg, Germany
DOI: https://doi.org/10.48441/4427.3718

2. Acknowledgements

We thank members of the Social Physics and Complexity (SPAC) group at LIP. We thank BrightData for generously providing proxy services. This work has received funding from the European Research Council (ERC) under the European Union's (EU) Horizon 2020 and Horizon Europe research and innovation programs (Grant agreements 853566 and 101100653), granted to J.G.-S., and from the Fundação para a Ciência e Tecnologia (FCT) PhD fellowship (2022.12547.BD), granted to I.D.

3. References

Adams, J., Merrill, S., III, & Zur, R. (2024). How much does issue salience matter? A model with applications to the UK elections. European Journal of Political Research, 63(2), 798–809.

Ekström, A. G., Madison, G., Olsson, E. J., & Tsapos, M. (2024). The search query filter bubble: Effect of user ideology on political leaning of search results through query selection. Information, Communication & Society, 27(5), 878–894. https://doi.org/10.1080/1369118X.2023.2230242

Epstein, R., Newland, A., & Tang, L. Y. (2025). The "digital personalization effect" (DPE): A quantification of the possible extent to which personalizing content can increase the impact of online manipulations. Computers in Human Behavior, 166, Article 108578. https://doi.org/10.1016/j.chb.2025.108578

Epstein, R., & Robertson, R. E. (2015). The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections. Proceedings of the National Academy of Sciences, 112(33), E4512–E4521.

Gezici, G., Lipani, A., Saygin, Y., & Yilmaz, E. (2021). Evaluation metrics for measuring bias in search engine results. Information Retrieval Journal, 24(5), 400–432. https://doi.org/10.1007/s10791-020-09386-w

Hopmann, D. N., Vliegenthart, R., De Vreese, C., & Albæk, E. (2010). Effects of election news coverage: How visibility and tone influence party choice. Political Communication, 27(4), 389–405.

Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., Chen, Q., Peng, W., Feng, X., Qin, B., et al. (2023). A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. arXiv. https://arxiv.org/abs/2311.05232

Kay, M., Matuszek, C., & Munson, S. A. (2015). Unequal representation and gender stereotypes in image search results for occupations. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (pp. 3819–3828). Association for Computing Machinery.

Mancosu, M., Seddone, A., & Bobba, G. (2021). Negativity in the news and electoral behavior between first- and second-order elections. Italian Political Science Review/Rivista Italiana di Scienza Politica, 52(3), 391–403. https://doi.org/10.1017/ipo.2021.8

Motoki, F., Pinho Neto, V., & Rodrigues, V. (2024). More human than human: Measuring ChatGPT political bias. Public Choice, 198(1), 3–23. https://doi.org/10.1007/s11127-023-01097-2

SEASON 2026, September 15–17, 2026, Hamburg, Germany
DOI: https://doi.org/10.48441/4427.3718

Mustafaraj, E., Lurie, E., & Devine, C. (2020). The case for voter-centered audits of search engines during political elections. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT* '20) (pp. 559–569). Association for Computing Machinery. https://doi.org/10.1145/3351095.3372835

Noble, S. U. (2018). Algorithms of oppression: How search engines reinforce racism. New York University Press.

Norocel, O. C., & Lewandowski, D. (2023). Google, data voids, and the dynamics of the politics of exclusion. Big Data & Society, 10(1). https://doi.org/10.1177/20539517221149099

Pan, B., Hembrooke, H., Joachims, T., Lorigo, L., Gay, G., & Granka, L. (2007). In Google we trust: Users' decisions on rank, position, and relevance. Journal of Computer-Mediated Communication, 12(3), 801–823.

Pew Research Center. (2024, September 9). Issues and the 2024 election. https://www.pewresearch.org/politics/2024/09/09/issues-and-the-2024-election/

Pew Research Center. (2025, September). News platform fact sheet. https://www.pewresearch.org/journalism/fact-sheet/news-platform-fact-sheet/

Potter, Y., Lai, S., Kim, J., Evans, J., & Song, D. (2024). Hidden persuaders: LLMs' political leaning and their influence on voters. In Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP) (pp. 4244–4275). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.emnlp-main.244

Robarts, S. (2025, September). AI challenges the dominance of Google search. BBC News. https://www.bbc.com/news/articles/c1dx9qy1eeno

Robertson, R. E., Jiang, S., Joseph, K., Friedland, L., Lazer, D., & Wilson, C. (2018). Auditing partisan audience bias within Google search. Proceedings of the ACM on Human-Computer Interaction, 2(CSCW). https://doi.org/10.1145/3274417

Rovny, J., Bakker, R., Hooghe, L., Jolly, S., Marks, G., Polk, J., Steenbergen, M., & Vachudova, M. (2024). 25 years of political party positions in Europe: The Chapel Hill Expert Survey, 1999–2024 [Working paper]. Chapel Hill Expert Survey.

Semrush. (2026, February). Top websites in the world: January 2026 most visited and popular rankings. Semrush Traffic Analytics. https://www.semrush.com/website/top/

Silberg, J., & Manyika, J. (2019, June). Notes from the AI frontier: Tackling bias in AI (and in humans). McKinsey Global Institute.

Urman, A., & Makhortykh, M. (2025). The silence of the LLMs: Cross-lingual analysis of political bias and false information prevalence in ChatGPT, Google Bard, and Bing Chat. Telematics and Informatics. https://doi.org/10.1016/j.tele.2024.102211

YouGov. (2025). Most important issues facing the U.S. https://today.yougov.com/topics/politics/trackers/most-important-issues-facing-the-us

SEASON 2026, September 15–17, 2026, Hamburg, Germany
DOI: https://doi.org/10.48441/4427.3718

SEASON 2026, September 15–17, 2026, Hamburg, Germany
DOI: https://doi.org/10.48441/4427.3718