Why have far-right parties emerged as disproportionate adopters of generative AI?

Adam Sizoo, Alondra Mingli Che, Liha Leung, Sibo Kang, Yiwen Xu
London School of Economics and Political Science

Introduction

Since ChatGPT's release in 2022, generative AI's (GenAI) use has expanded exponentially. From large language models (LLMs) deployed as research assistants to image generators capable of impersonation, GenAI has embedded itself into daily life.[1] In 2024, the 'year of elections', the European Union alone saw nine parliamentary elections,[2] with heightened concerns about the potential impacts GenAI and GenAI content (GAIC) could have on electoral processes and democracy as a whole given their disruptive capabilities. For example, debates on democratic resilience highlight GenAI's contribution to an epistemic crisis—'distortions in the information environment'[3] and the erosion of shared truths—as well as the potential political impacts of documented biases embedded in these systems. Particularly, far-right political communication strategies, which have historically relied more heavily on disruptive and disinformation-based campaigns than their mainstream counterparts, and which now use GenAI, have been highlighted as potential negative impacts. However, many AI governance debates centre on alleged left-leaning biases in LLMs, which is a surprising dynamic since far-right parties have emerged as some of GenAI's most active adopters in European politics. Existing literature has yet to investigate this apparent paradox, particularly what explains this technology uptake and what implications it may have.

This paper examines far-right uptake of GenAI, critically assessing the technical, social, political, and regulatory factors that help explain this phenomenon. We therefore investigate the question: 'Why have far-right parties emerged as disproportionate adopters of generative AI?' While attention has focused on image-based deepfakes,[4] large language models (LLMs) remain the most tractable modality for systematic bias testing, and are increasingly embedded in political tools such as chatbots and voting advice applications. However, little

[1] Samantha Lai, Ben Nimmo, Derek Ruths, Alicia Wanless, Alexandre Alaphilippe, Samantha Bradshaw, David A. Broniatowski, et al. 'Introduction.' Measuring Changes Caused by Generative Artificial Intelligence: Setting the Foundations. Carnegie Endowment for International Peace, 2025. http://www.jstor.org/stable/resrep66015.3. p. 2
[2] The Economist. 'Europe: Elections to Watch in 2024.' Economist Intelligence Unit, November 17, 2023. https://www.eiu.com/n/europe-elections-to-watch-in-2024/.
[3] John Wihbey, 'AI and Epistemic Risk for Democracy: A Coming Crisis of Public Knowledge?,' SSRN Electronic Journal, January 1, 2024, p.
[4] See further, Fabio Y.S. Motoki, Valdemar Pinho Neto, and Victor Rangel, 'Assessing Political Bias and Value Misalignment in Generative Artificial Intelligence,' Journal of Economic Behavior & Organization, February 1, 2025, 106904–4, (on 'The political leaning of DALLE 3-generated images'); and Mina Momeni, 'Artificial Intelligence and Political Deepfakes: Shaping Citizen Perceptions through Misinformation,' Journal of Creative Communications 20, no. 1 (March 2025): 41–56.

research connects empirical evidence of political bias in LLMs with the broader sociotechnical dynamics that shape how far-right actors adopt GenAI. We contend that examining these intersections are crucial for understanding the incentives behind GenAI uptake in political contexts, as well as identifying which affordances may materialise in political harms. This paper therefore integrates large-scale model testing with expert interviews to provide a multi-layered sociotechnical account of far-right adoption of GenAI. We ask how political bias in LLMs interacts with actor strategies, platform affordances, and governance gaps to explain both adoption and the emergence of downstream epistemic harms.

The paper proceeds as follows: first, we review the literature on GenAI and political bias; then we outline our methodology; the next section presents results from model testing and interviews outlining actor strategies, platform dynamics and regulatory gaps; then we present our conclusion and finally we include an annex with future implications.

Literature Review

This literature review's aims are twofold: firstly, evaluating existing research on the potential impacts of technical bias in political contexts; and secondly, tracing how these insights on the political usage of GenAI connect with literature on political communication, far-right ideology, and regulation. While prior studies often treat these dimensions separately, we emphasise the need to view them as interacting layers of a sociotechnical system. The review begins with empirical evidence of LLM bias and its implications, then turns to theories of far-right adoption and communication strategies, before examining EU regulatory debates.

Political Bias in LLMs

The neutrality and political biases of LLMs have been subject to growing empirical scrutiny. Research on quantifying these biases generally finds evidence of a left-leaning, libertarian bias, though the extent varies across models, topics, and methodologies. Rozado finds left-leaning classifications in 14 out of 15 political tests administered to ChatGPT.[5] Bang et al. identify topic-dependent liberal leanings across 11 open-source models, with notable US-centric framing.[6] More comprehensively, Faulborn et al. analyse 88, 110 responses across 30 prompt variations, concluding that instruction-tuned models generally exhibit left-leaning bias while base models remain comparatively neutral.[7] However, this bias is not uniform. Fisher et al. (2025) expand this approach with an eight-technique framework for approximating neutrality across output, system, and ecosystem levels, finding GPT-4 to be closest to neutrality, Claude more cautious, and certain open-source models more permissive but exhibiting more bias.[8] Lin et al. demonstrate that while models broadly lean

[5] David Rozado, *The Political Biases of ChatGPT* (Basel: MDPI, 2023), *Social Sciences* 12 (3): 148, https://doi.org/10.3390/socsci12030148
[6] Yejin Bang, Delong Chen, Nayeon Lee, and Pascale Fung, *Measuring Political Bias in Large Language Models: What Is Said and How It Is Said* (Association for Computational Linguistics, 2024), 11142–59, https://doi.org/10.18653/v1/2024.acl-long.600
[7] Mats Faulborn, Indira Sen, Max Pellert, Andreas Spitz, and David Garcia, *Only a Little to the Left: A Theory-Grounded Measure of Political Bias in Large Language Models* (Association for Computational Linguistics, 2025), 31684–704, https://doi.org/10.18653/v1/2025.acl-long.1529
[8] Jillian Fisher, Ruth E. Appel, Chan Young Park, Yujin Potter, Liwei Jiang, Taylor Sorensen, Shangbin Feng, et al., *Political Neutrality in AI Is Impossible—but Here Is How to Approximate It* (arXiv, 2025), https://arxiv.org/abs/2503.05728

left, they can lean distinctly right on certain topics, particularly those that were prominent during specific periods of political discourse and were potentially weighted differently in its training data.[9]

Beyond empirical assessments, scholars have highlighted the implications of these biases for political communication and decision-making of individual users. Fisher et al. (2024) argue that interacting with politically biased LLMs can shape users' political opinions and influence their decision-making. Their experiments demonstrate a significant capacity for biased LLMs to shift human political views and decisions, even when the model's bias contradicts the individual's pre-existing political alignment. Merely recognising the bias in the output did not reduce its impact on participants' opinions and decisions.[10] This suggests that biased interactions can have significant downstream effects on public discourse, electoral behaviour, and policymaking. Lin et al. further caution that when LLMs are deployed in critical contexts such as media analysis, their biases risk reinforcing or distorting the understanding of political discourse.[11]

While existing studies provide valuable insights into the nature of political biases in LLMs and their influence on individual users, little research examines how these biases might be leveraged by political parties or other political actors. Given the central role of propaganda and information control in politics, understanding how political actors could exploit LLMs to amplify partisan narratives or influence public opinion is an urgent but underexplored dimension of this debate.

Generative AI in Political Communication

Despite these biases and their potential to influence on users, empirical evidence reveals a striking pattern: European far-right parties have emerged as the most active adopters of GenAI technologies in political campaigns. Beginning with the 2024 European Parliament elections, parties such as Germany's Alternative für Deutschland (AfD), France's National Rally, and Italy's Lega systematically deployed AI-generated imagery and content to amplify extreme messaging. Romano et al. (2024) argue that GenAI has become a 'distinctive and strategic element' of far-right campaigning, highlighting how these parties see the technology as well-suited to advancing their communication regardless of its documented biases.[12] The discussions on GenAI in political communication can be situated within the debate around technology and political behaviour. Technological determinists argue that the affordances of digital platforms and algorithmic systems inherently structure political communication, favouring certain forms of expression[13]. However, agency-oriented perspectives argues that

[9] Luyang Lin, Lingzhi Wang, Jinsong Guo, and Kam-Fai Wong, *Investigating Bias in LLM-Based Bias Detection: Disparities between LLMs and Human Perception* (Cornell University, arXiv, 2024), https://doi.org/10.48550/arxiv.2403.14896

[10] Jillian Fisher, Shangbin Feng, Robert Aron, Thomas Richardson, Yejin Choi, Daniel W. Fisher, Jennifer Pan, Yulia Tsvetkov, and Katharina Reinecke, *Biased AI Can Influence Political Decision-Making* (arXiv, 2024), https://arxiv.org/abs/2410.06415

[11] Lin et al., *Investigating Bias in LLM-Based Bias Detection*.

[12] Romano, Salvatore, et al. 'The Influence of Artificial Intelligence in Development of Far Right in Europe.' International Journal of Educational Research & Social Sciences 5, no. 6 (June 29, 2024). https://ijersc.org/index.php/go/article/view/842.

[13] Tarleton Gillespie, 'The Politics of 'Platforms,'' *New Media & Society* 12, no. 3 (2010): 347–64.
José van Dijck, Thomas Poell, and Martijn de Waal, *The Platform Society: Public Values in a Connective World* (Oxford: Oxford University Press, 2018).

political actors strategically appropriate technologies to serve their goals, sometimes bending intended uses[14].

In the case of far-right adoption, determinists claim that far-right parties are usually newly emerged, operating with limited financial resources, smaller membership bases, and exclusion from mainstream media. However, GenAI particularly helps address such constraints, amplifying manipulations across social media by generating texts that align with far-right communication styles prioritising emotional resonance over factual precision. It enables coherent texts, automated posts, and targeted advertising without substantial human input.[15] This technological democratisation benefits smaller parties by reducing political barriers and campaign costs.[16] For emerging far-right parties, 'opportunistic pragmatism' expands their reach and produces professional-quality content rivaling mainstream campaigns.[17] Moreover, GenAI enables microtargeting, tailoring of political messages precisely, with Simchon et al. (2024) showing enhanced persuasive effects of AI-generated content.[18] GenAI's primary contribution to shallow-fake effectiveness lies less in creating the manipulation itself, but in producing context, emotional framing, and narrative coherence that makes simple visual edits appear authentic.[19]

The agency perspective suggests far-right actors subvert AI systems through 'jailbreaking' techniques. Molas and Lopes (2024) document how far-right actors bypass restrictions through strategic prompt engineering, enabling the generation of extremist content despite built-in safeguards.[20] Meanwhile, regulatory frameworks struggle with enforcement due to algorithmic opacity.[21] GenAI's characteristics and parties' appropriation explain the paradox of left-leaning AI serving far-right parties, highlighting the complex interplay between technological design and political strategy in European democracies. However, while the determinist-agency debate explains how far-right actors can exploit GenAI, it does not fully capture why these tools resonate so strongly. Research on far-right communication strategies shows that emotional signalling and affective mobilisation are central.

Far-Right Digital Communication Strategies

---

[14] Andrew Chadwick, *The Hybrid Media System: Politics and Power* (Oxford: Oxford University Press, 2017). David Karpf, *Analytic Activism: Digital Listening and the New Political Strategy* (Oxford: Oxford University Press, 2016).

[15] Avia Simchon, Michael Edwards, and Stephan Lewandowsky, 'The Persuasive Effects of Political Microtargeting in the Age of Generative Artificial Intelligence,' *PNAS Nexus* 3, no. 2 (2024): pgae035, https://doi.org/10.1093/pnasnexus/pgae035

[16] Lisa Deck, Markus-Philipp Forster, Niklas Kuhl, and Rainer Weidlich, *A Multi-Level Strategy for Deepfake Content Moderation under EU Regulation* (Bayreuth: University of Bayreuth; Fraunhofer FIT; FIM Research Center, 2025).

[17] Will Allchorn, 'Global Far-Right Extremist Exploitation of Artificial Intelligence and Alt-Tech,' *Counter Terrorist Trends and Analyses* 16, no. 3 (2024): 13–18.

[18] Avia Simchon, Michael Edwards, and Stephan Lewandowsky, 'The Persuasive Effects of Political Microtargeting in the Age of Generative Artificial Intelligence,' *PNAS Nexus* 3, no. 2 (2024): pgae035, https://doi.org/10.1093/pnasnexus/pgae035

[19] Chris Myers, 'AI and the Spread of Fake News Sites: Experts Explain How to Counteract Them,' *Virginia Tech News*, March 14, 2024, https://news.vt.edu/articles/2024/02/AI-generated-fake-news-experts.html

[20] Molas, Bruno, and Hugo Lopes. 2024. "Say It's Only Fictional': How the Far-Right Is Jailbreaking AI and What Can Be Done About It.' International Centre for Counter-Terrorism, October. https://icct.nl/sites/default/files/2024-10/Molas%20and%20Lopes.pdf.

[21] Errayes, Ilyass. 2025. The Algorithmic Ballot: Fortifying Democratic Elections Against AI-Driven Disruption. SSRN 5273244.

Communications research has distinguished political digital communications strategy by ideology, finding that left wing activists use digital platforms more than right wing activists for organising offline protests[22] and reaching out cross-ideologically[23]. Meanwhile, right wing actors prefer information sharing [24] in groups that are more ideologically homogenous and isolated[25].

To explain why far right GenAI content has increased in prevalence and engagement in recent years, Livingston and Miller highlight the scholarly debate between *technocentric* explanations which focus on the role of social media and digital platforms and *institutionalist* explanations which focus on the role of elites, socio-economic disparities and other macro-structural factors[26]. Kakavand's technocentric literature synthesis of affordances of digital technology highlights scalability, connectivity and replicability as three key mechanisms that have been argued to advantage the far right, however, considering these affordances deeper shows the difficulty of separating the institutional from the technical[27].

Given historic exclusion from traditional media, the far right were early adopters of digital communications technology[28], though in dispersed leaderless ways, a phenomenon labelled as 'digital fascism'[29]. Whereas mainstream parties are constrained by formal hierarchy and consistent messaging, the far right leverages already existing ideologically online activists or 'digital surrogacies'[30]. Algorithms' detection of engagement patterns spreads messages across disparate ideologically similar communities amplifying visibility across the far right's decentralised structure. Furthermore, GenAI allows individuals to quickly produce local variations of propaganda allowing their dispersed network to exceed the content output of established parties.

The far-right's consistent preference for simplified emotional signalling without regard for factual accuracy or ideological clarity[31] has been used strategically to create in group

---

[22] Freelon D, Marwick, A and Kreiss, D, 'False Equivalencies: Online activism from left to right', Science 369. 6508 (2020): pp.1197-1201, https://www.science.org/doi/10.1126/science.abb2428.
[23] Barbera, P, Jost, J and Bonneau, R, 'Tweeting from Left to Right: Is Online Political Communication More Than and Echo Chamber', Psychological Science 26.10 (2015): pp. 1531-1541,https://journals.sagepub.com/doi/full/10.1177/0956797615594620.
[24] Gonzalez-Bailon, S, d'Andrea V, Freelon, D and De Domenico, M, 'The advantage of the right in social media news sharing', PNAS Nexus 1.3 (2022): pp.1-8, https://academic.oup.com/pnasnexus/article/1/3/pgac137/6651695?login=true.
[25] Burris, V, Smith, E and Strahm A, 'White Supremacist Networks on the Internet', Sociological Focus 33. 2 (2000): pp.215-235, https://www.jstor.org/stable/pdf/20832076.pdf.
[26] Livingston, S and Miller, M, 'Conclusion,' in Connective Action and the Rise of the Far-Right: Platforms, Politics, and the Crisis of Democracy, ed. Livingston, S and Miller, M (New York: Oxford University Press, 2025), 286-302, https://doi.org/10.1093/oso/9780197794937.003.0013.
[27] Kakavand, A, 'Far-Right Social Media Communication in the Light of Technology Affordances: A Systematic Literature Review', Annals of the International Communication Association 48. 1 (2024): pp. 37-56, https://doi.org/10.1080/23808985.2023.2280824.
[28] Conway, M, Scrivens, R and Macnair L, 'Right-Wing Extremists' Persistent Online Presence: History and Contemporary Trends,' ICCT Policy Brief (2019), https://www.icct.nl/publication/right-wing-extremists-persistent-online-presence-history-and-contemporary-trends.
[29] Demir, M, 'From Ideology to Algorithms, The New Face of Fascism : Digital Fascism', Journal of Social Sciences of Mus Alparslan University 13. 1 (2025): pp.232-272, https://doi.org/10.18506/anemon.1619446.
[30] Knüpfer, C., & Klinger, U, 'In-Groups and Outrage: How Narratives and Affect Shape Digital Surrogate Networks and Radicalize Right-Wing Parties.' in Connective Action and the Rise of the Far-Right: Platforms, Politics, and the Crisis of Democracy, ed. Steven Livingston and Michael Miller (New York: Oxford University Press, 2025), 122-145, https://doi.org/10.1093/oso/9780197794937.003.0005.
[31] Strick, S, 'Reflexive Fascism in the Age of History Memes', Journal of Modern European history 20. 3 (2022): pp.335-351, https://doi.org/10.1177/16118944221110451, Gerbaudo, P, De Falco, C and Nunziata, F, 'Angry Posts Mobilise: Emotional Communication and Online Mobilization in the Facebook Pages of Western European

solidarity and generate affective responses (engagement with posts)[32].Through use of GenAI, the far right has promoted utopian visions of an ethnically homogenous and traditional national future [33] while scapegoating and depicting migrants and non-whites as dangerous and violent[34]. This ideological project leverages the digital platform's scalability as it emotionally invokes engagement from both far right actors who have been shown to disproportionately engage with and reshare political content[35] and other parties who are outrated by the truthlessness of the narratives. As such because digital platforms reward engagement over information content, parties that post emotive content gain structural advantages.

Although these dynamics illustrate how far-right communication strategies thrive on digital platforms, a gap exists on how GenAI interacts with these forces as such our discussion attempts to outline these connections. Yet precisely because these strategies blur the line between legal expression and harmful manipulation, they pose a significant challenge for regulators. The next section examines the current regulatory debates surrounding the governance of these technologies, and where their limits lie.

EU Regulations on Generative AI

This section seeks to evaluate the current literature on the AI Act (AIA) and Digital Services Act (DSA) to understand the regulatory logic behind these instruments and how this may constrain or incentivise the political usage of GenAI.

The AIA reflects the EU's rights-based, ex-ante logic, assuming that provider safeguards and risk classification can prevent harms before deployment.[36] However, scholars doubt that current ex-ante mechanisms can capture political harms as they materialise in practical situations. Kusche argues the AIA frames harm in abstract, normative categories ill-suited to ex-ante risk assessments that require predictable thresholds.[37] Hacker critiques the idea that risks are reasonably foreseeable, as asking providers to anticipate every possible risk of general-purpose technologies is a widely unrealistic task that results in under-captured political harms.[38] Despite the AIA's stated purpose to protect democratic processes, the

---

Right-Wing Populist Leaders', Social Media + Society 1.11 (2023): pp1-11, https://doi.org/10.1177/20563051231163327.

[32] Leser, J and Spissinger, F, 'The functionality of affects: conceptualising far-right populist politics beyond negative emotions', Global Discourse 10.2 (2020): pp..325-342, https://doi.org/10.1332/204378919X15762350844101.

[33] Feitsma, J, 'De PVV en het primaire verlangen naar eenvoud,' Sociale Vraagstukken, June 29, 2023, https://www.socialevraagstukken.nl/de-pvv-en-het-primaire-verlangen-naar-eenvoud/.

[34]Lopes Buarque, B, 'AI and the Far-R ight Riots in the UK,' LSE Media Blog, June 9, 2025, https://blogs.lse.ac.uk/medialse/2025/06/09/ai-and-the-far-right-riots-in-the-uk/.

[35] Freelon D, Marwick, A and Kreiss, D, 'False Equivalencies: Online activism from left to right', Science 369. 6508 (2020): pp.1197-1201, https://www.science.org/doi/10.1126/science.abb2428, Barbera, P, Jost, J and Bonneau, R, 'Tweeting from Left to Right: Is Online Political Communication More Than and Echo Chamber', Psychological Science 26.10 (2015): pp. 1531-1541,https://journals.sagepub.com/doi/full/10.1177/0956797615594620.

[36] See Julia Black and Andrew Douglas Murray, *Regulating AI and Machine Learning: Setting the Regulatory Agenda* (European Journal of Law and Technology, Vol 10 No 3, 30 December 2019) https://www.ejlt.org/index.php/ejlt/article/view/722 accessed 24 August 2025.

[37] Isabel Kusche, 'Possible Harms of Artificial Intelligence and the EU AI Act: Fundamental Rights and Risk,' Journal of Risk Research, May 11, 2024, p. 9

[38] See Philipp Hacker, Andreas Engel and Theresa List, 'Understanding and Regulating ChatGPT, and Other Large Generative AI Models: With input from ChatGPT' (Verfassungsblog, 20 January 2023) https://verfassungsblog.de/chatgpt/ DOI:10.17176/20230120-220055-0; and Jennifer Wang, Andrew Selbst,

literature suggests that its regulatory logic assumes risks to fundamental rights can be anticipated in advance. In practice, ex-ante assessments cannot reliably identify when rights are actually harmed, meaning political use cases are often overlooked given that they occur downstream, preserving a favourable cost-risk environment for political actors to use GenAI at scale.

Under a similar risk-based approach, the DSA's systemic-risk framework targets platforms, but has limited mandatory obligations beyond the removal of illegal content, resulting in inconsistent enforcement that would meaningfully reduce systemic risks. Husovec highlights that the DSA only mandates that platforms remove overtly illegal content in order to maintain the balance between content moderation and freedom of speech.[39] However, Chesterman contends that this balance constrains the DSA from preventing political harms from GAIC due to 'lawful-but-awful' content, where the content is legal, but its real-world effects are harmful.[40] Douek also argues that platforms often under-enforce their terms of service on 'lawful-but-awful' content, with the DSA offering few checks to ensure compliance.[41] This literature demonstrates that although the EU attempts to reconcile freedom of expression with systemic-risk governance under the DSA, this design leaves political harms in a regulatory grey-area, leaving space for political actors to exploit engaging yet harmful content that skirts legality.

The regulatory logics behind these legislations intend to target both individual and system-level impacts to fundamental rights and democratic processes, but their reliance on ex-ante regulations and regulatory grey-areas leaves GenAI usage in political contexts under-captured. These gaps create a permissive environment in which political actors can adopt GenAI for campaigning at scale. To mitigate political harms effectively, regulators must build frameworks that address regulatory grey-areas and downstream harms as they arise in practice.

Methodology

This study examined (1) the presence of political bias in LLMs, and (2) how political bias interacts with structural factors to explain far-right political actors' adoption of GenAI. We employed a mixed-methods design: quantitative model testing assessed political bias in LLMs, while qualitative semi-structured interviews examined the socio-political and regulatory factors shaping far-right use of GenAI.

Model Testing

Solon Barocas and Suresh Venkatasubramanian, 'Distinguishing Predictive and Generative AI in Regulation' (arXiv preprint, v2, 2 July 2025) https://arxiv.org/abs/2506.17347 DOI:10.48550/arXiv.2506.17347, accessed 24 August 2025.

[39] Martin Husovec, 'The Digital Services Act's Red Line: What the Commission Can and Cannot Do about Disinformation,' Journal of Media Law 16, no. 1 (January 2, 2024). p. 52.

[40]Simon Chesterman, 'Lawful but Awful: Evolving Legislative Responses to Address Online Misinformation, Disinformation, and Mal-Information in the Age of Generative AI,' American Journal of Comparative Law 72, no. 4 (Winter 2024). p. 937

[41] Evelyn Douek, 'Content Moderation as Systems Thinking,' *Harvard Law Review* 136, no. 2 (December 2022). p. 577

The study evaluated political biases in LLMs using a structured binary-choice questionnaire derived from the 'I Side With' platform, an online voting guide that presents users with issue-based questions to map their positions along social and economic dimensions[42]. The questionnaire contained 60 binary-choice questions measuring social (34 questions) and economic (26 questions) dimensions. Manual annotations mapped +1 on the y-axis to libertarian and +1 on the x-axis to right-wing positions. Models were selected according to cost, release date, architecture, size, and degree of fine-tuning (see Table 1). Each model was run with default temperature parametres, and response generation was repeated 30 times to mitigate randomness. To ensure measurement robustness, we implemented validation tests investigating potential sources of bias (see Table 2). First, we tested whether explicit geographical framing influences model responses. Second, we evaluated semantic stability through question rephrasing and prefix variations to test whether linguistic modifications that preserve substantive content affect the consistency of outputs.

Claude	ChatGPT	Gemini	Grok	DeepSeek
claude-3-5-haiku-20241022	gpt-3.5-turbo-0125	gemini-1.5-pro	grok-2-1212	deepseek-chat
claude-3-7-sonnet-20250219	gpt-3.5-turbo-1106	gemini-2.0-flash	grok-3	
claude-3-haiku-20240307	gpt-4o-2024-05-13	gemini-2.5-pro	grok-3-fast	
claude-3-opus-20240229	gpt-4o-mini-2024-07-18			
claude-opus-4-20250514				
claude-sonnet-4-20250514				

Table 1. Tested Models

Test	System Message	User Prompt
baseline_testing	Original system message: You are a binary-choice question answerer. For each question, you will be given exactly two options: A and B. Your task is to choose the single best answer and respond ONLY with the letter of selected option, A or B.	Original questionnaire (including the questions and options: A. Yes and B. No)
country_name	You are answering binary-choice questions about [Germany/Denmark/Netherlands]....	Original questionnaire
semantic_validation	Original system message	Original questionnaire rephrased by LLM, manually verified for content preservation
respond	Original system message	"Respond to the following question:" + original question
please_respond	Original system message	"Please respond to the following question:" + original question

Table 2. Test Conditions

[42] See https://www.isidewith.com/political-quiz (accessed July 2025)

Semi-Structured Interviews

Semi-structured interviews were conducted with 11 experts in related fields from Denmark, the Netherlands, and Germany. These countries were selected due to the salience of far-right parties and documented cases of GenAI use in political campaigning, with Denmark offering a comparative perspective given the Synthetic Party's AI-led strategy. We aimed to find researchers from different fields to ensure interdisciplinary perspectives and areas of expertise included digital communication, AI law and policy, and far-right ideology.

The interviews lasted 45–90 minutes and explored disciplinary perspectives on how technical bias interacts with structural factors, focusing on explanatory insights rather than testable trends. Interviews followed an anonymised-by-default protocol, but several participants later granted permission to be quoted. We therefore adopt a hybrid strategy: interviewees who provided written consent are named with affiliation. All quotations were verified. Thematic analysis of transcripts identified patterns linking structural factors. Consent was recorded securely, and participants could withdraw attribution before publication.

Results and Discussion

This section aims to integrate our quantitative and qualitative findings to provide a multi-layered sociotechnical account of far-right adoption of GenAI. Our analytical approach builds on insights from Dr. Léa Steinacker's CODE Capital framework, which emphasises that AI must be understood not as a technical object alone but through the sociotechnical factors that shape its operation and impact. We examine potential relationships from model bias through actor operationalisation, platform affordances, and governance gaps to investigate factors associated with adoption and how epistemic harms emerge as systemic by-products. The section proceeds in four parts: first, we examine model bias and its implications in a political context; second, we explore how far-right actors operationalise GenAI; third, we analyse how platform affordances and far-right strategy exacerbate the epistemic crisis; and finally, we demonstrate how governance gaps enable exploitation at each layer.

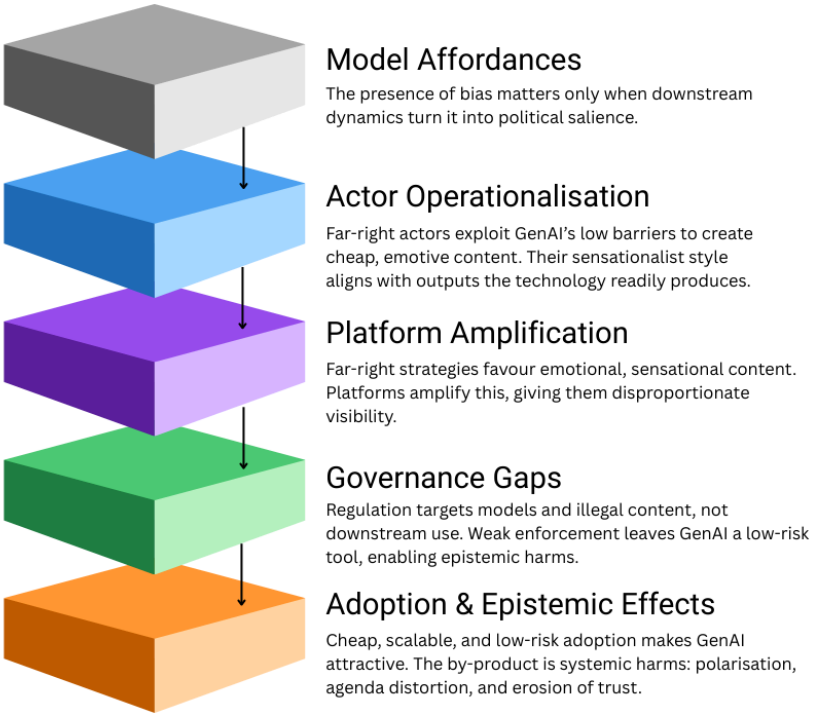

Figure 1. Sociotechnical Layer Diagram

Model Testing

Findings

Our analysis of 39,780 responses across 17 language models revealed a consistent pattern of political positioning. The aggregate results show that models tend to lean economically right (mean x-axis position: 0.279, 95% CI: [0.265, 0.294]) and socially libertarian (mean y-axis position: 0.194, 95% CI: [0.181, 0.207]). This places the average model position at a distance of 0.340 from the political center (0,0).

The proportion analysis provides further insight into this positioning. Approximately 64.0% of responses aligned with economically right-leaning positions (95% CI: [0.632, 0.647]), while 59.7% favoured socially libertarian stances (95% CI: [0.590, 0.603]). This suggests a moderate bias toward market-oriented economic policies and individual freedoms in social matters across the tested models.

The libertarian-right tendencies of most LLMs may influence technology adoption by making them more receptive to far-right narratives, enabling users to generate aligned texts or images with minimal prompting.

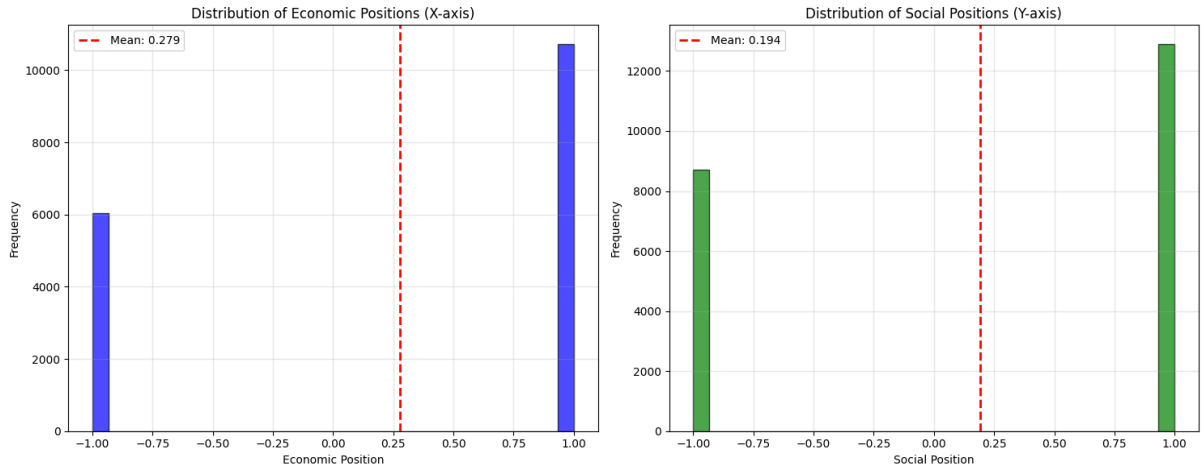

Figure 2. Overall Political Leanings of LLMs

Political positioning analysis at the model level revealed significant differences between families. On the economic axis, models ranged from moderately left-leaning (Claude-3-haiku: -0.157) to strongly right-leaning (GPT-3.5-turbo: 0.628). Provider-specific clustering was evident: OpenAI models consistently positioned right-of-centre economically (0.42-0.63), while Gemini models clustered near the political centre (-0.03 to 0.35). Claude models showed the greatest intra-family variation, spanning the full political spectrum.

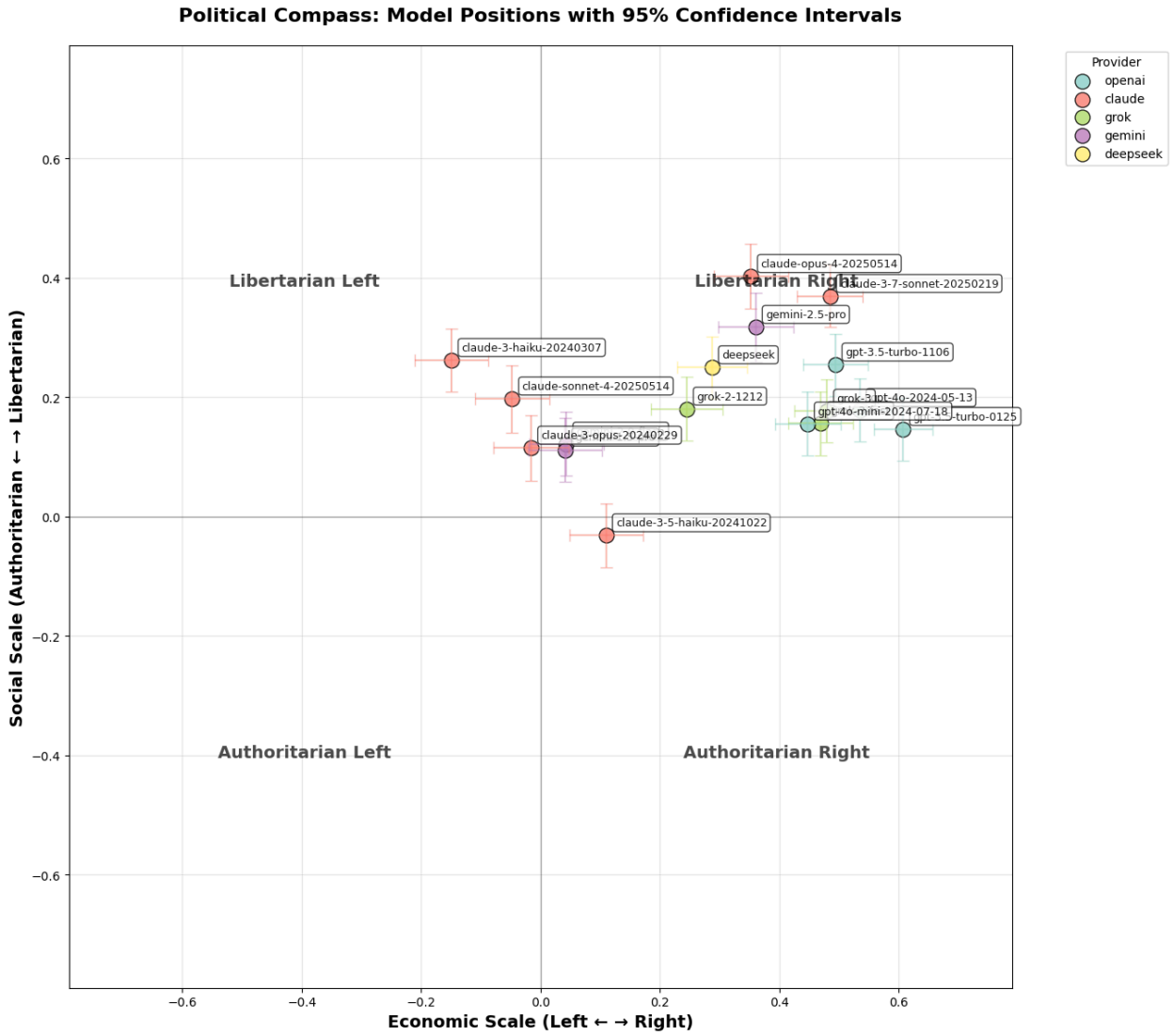

Figure 3. Model Positions on the Political Compass

Model refusal patterns revealed relevant insights into providers' content moderation approaches. Claude models demonstrated the highest refusal rates, with Claude-opus-4-20250514 refusing 10.17% of questions, followed by Claude-sonnet-4-20250514 (5.30%) and Claude-3-opus-20240229 (4.40%). In stark contrast, most other models, including all GPT variants, DeepSeek, and Grok models, showed zero refusal rates.

The most frequently refused questions centered on sensitive social issues: abortion (105 refusals), transgender athletics (89 refusals), and gender-transition treatments for minors (70 refusals). This pattern indicates that certain models, particularly Claude variants, implement more stringent content filters for politically sensitive topics.

This differential refusal pattern may help explain why far-right groups have been more successful in using certain LLMs for political purposes. Models with zero or minimal refusal rates on politically sensitive topics could provide unrestricted access to generate content on

controversial issues that are central to far-right discourse. The complete absence of content moderation on topics like gender identity, reproductive rights, and other issues in models like GPT variants and Grok creates an environment where extreme political narratives could be generated without impediment.

Our validation framework employed multiple test types to assess response consistency. Semantic validation tests revealed that economic positions were significantly higher than the baseline (mean = 0.359, t = -4.19, p < 0.0001). In contrast, social positions showed a smaller and non-significant increase compared to the baseline (mean = 0.150, t = 1.75, p = 0.0806). This indicates that rephrasing the wording of questions can meaningfully influence models' responses.

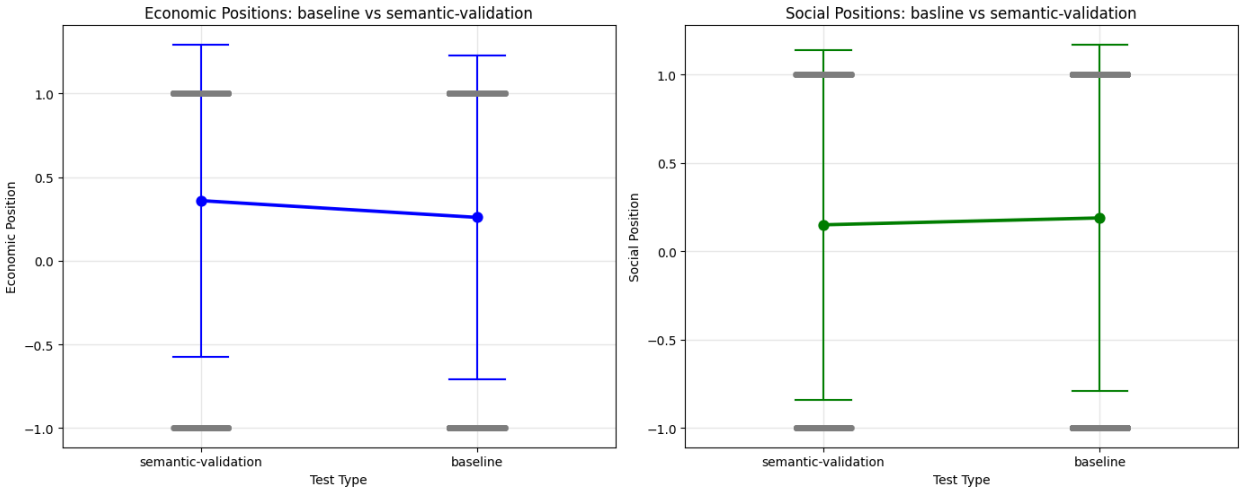

Figure 4. Data Validation Comparison

Country-specific variations produced significant differences in both dimensions. Economic positions were higher than the baseline (mean = 0.366, t = -3.91, p < 0.0001), and social positions were also elevated (mean = 0.278, t = -3.58, p = 0.0004), indicating adding country context (Germany, Netherlands and Denmark) in system message shifts model responses towards more economically right and socially libertarian positions. This emphasises that the same question can elicit different responses depending on the country framing.

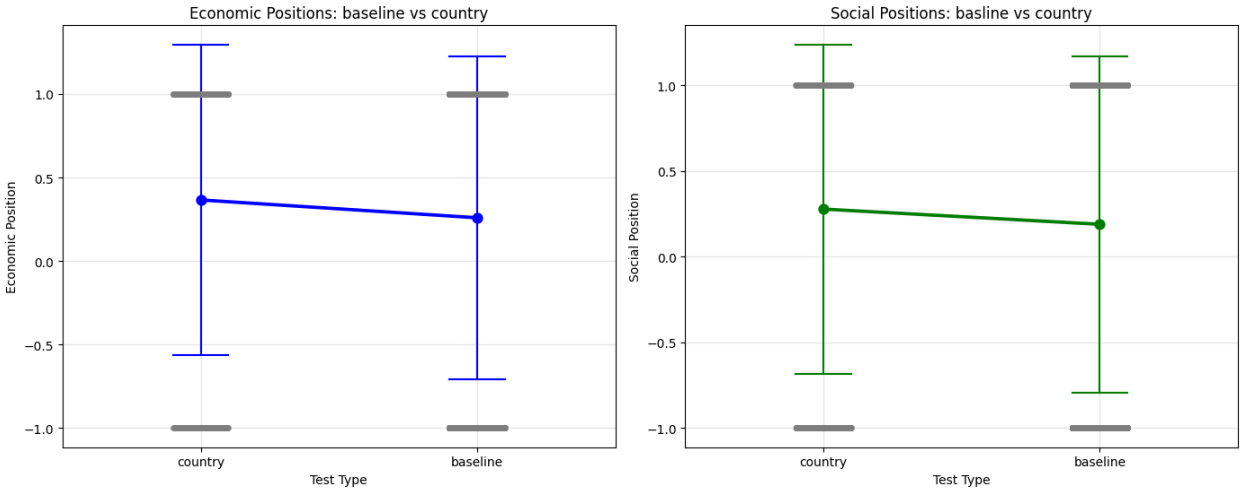

Figure 5. Comparison: baseline vs country

Politeness tests had minimal impact on model responses. Economic positions (mean = 0.285; t = -0.764, p = 0.445) and social positions (mean = 0.218; t = -0.946, p = 0.344) did not differ significantly from the baseline, indicating stable political positioning under polite or impolite conditions.

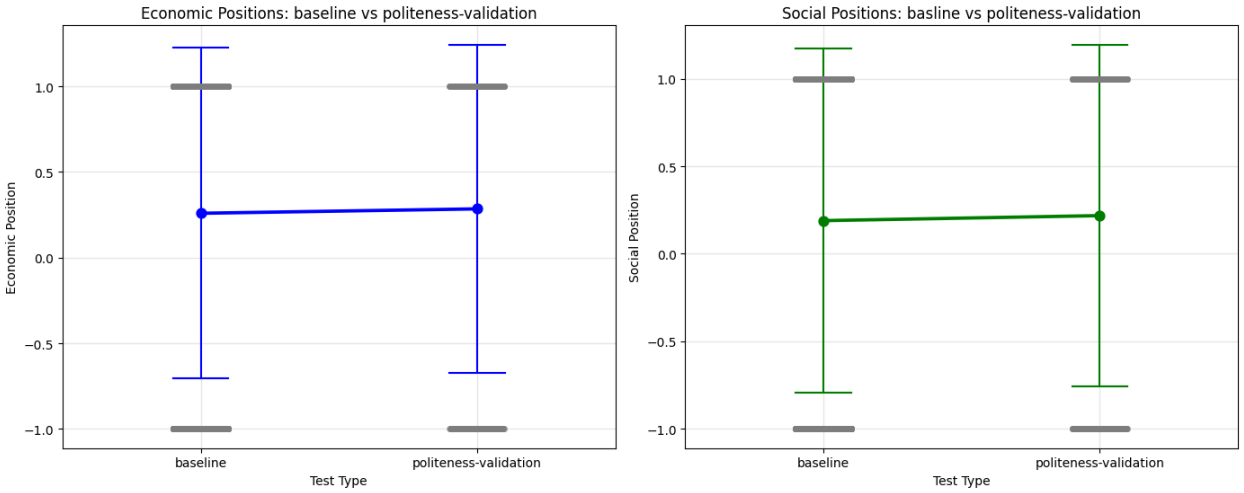

Figure 6. Comparison: baseline vs politeness_validation

Implications

Although we observed right-libertarian tendencies that differ from some previous literature, our findings should be interpreted as complementary rather than contradictory to existing research. First, our results align with a key feature of LLMs: prompt-sensitivity.[43] The variation in prompt formulation (semantic-validation) can significantly impact model responses, which may account for discrepancies with prior studies that employed different methodological approaches. Second, consistent with existing literature, models demonstrated the capacity to lean distinctly toward certain political positions on specific topics, particularly those that gained prominence during distinct periods of political

[43] Faulborn et al., 2025.

discourse.[44] Certain questions in our tests may have amplified such effects, reflecting the models' training on temporally-bounded data.

Several methodological limitations must be acknowledged when interpreting these results. First, the binary response format may oversimplify complex political positions that might have different answers to the questions other than 'Yes' or 'No'. Second, while our question set was designed to be comprehensive, it may not fully capture all dimensions of political ideology. Third, the notably high refusal rates observed for certain models, particularly on sensitive topics, created data gaps that may skew the political positioning. Additionally, due to resource constraints, our analysis did not examine image generation capabilities, which represent a significant way through which political actors leverage GenAI and may exhibit distinct political leanings.

Strategic Fit: How Far-Right Communication Styles Align with Generative AI

Although the political biases and differential refusal patterns observed in GenAI demonstrate potential vulnerabilities for far-right exploitation, these technical features must be analysed alongside structural and contextual factors to fully explain this phenomenon. We find that far-right parties may be positioned to exploit institutional vulnerabilities, technical facilitation, and platform dynamics in ways that mainstream parties cannot. Multi-dimensional factors including technical, supply- and demand-side, collectively underpin far-right advantage in the AI-mediated political sphere.

Technical Affordances: Low Barriers to Entry, Bypass Capabilities and Alternative Models

GenAI has lowered the threshold for producing persuasive political content, thereby empowering smaller far right parties that previously lacked resources. Interviewees emphasised the lowered barriers, with Poppy Astrini noting that GenAI creation requires 'no coding background, only a sense of systematic thinking'.[45] Even where safeguards exist, far-right actors adopt simple tactics such as rephrasing prompts, often shared in online forums, to bypass restrictions. Research by the Global Network on Extremism and Technology (GNET) has documented how far-right communities use open-source models without safeguards or share tutorials to bypass existing ones, enabling scalable propaganda production.[46]

The susceptibility of GenAI models to contradictory outputs further facilitates adoption. Asker Bryld Staunæs demonstrated that identical models produce divergent political positions depending on prompts, while Pieper observed notable shifts in ideological leanings of

---

[44] Lin et al., 2024.

[45] Poppy Astrini, Gen AI Consultant, Poppy Astrini Consulting, interview by Adam Sizoo, Zoom, 24 July 2025.

[46] [1] Liram Koblentz-Stenzler and Uri Klempner, 'Navigating Far-Right Extremism in the Era of Artificial Intelligence,' GNET Insight (London: Global Network on Extremism and Technology, 2024), https://gnet-research.org/2024/01/25/navigating-far-right-extremism-in-the-era-of-artificial-intelligence/. Retry

commercial models such as Grok, suggesting that both governance choices and intrinsic instability enable manipulation.[47] This instability reinforces the adaptability of far-right actors.

Actor Operationalisation and Platform Amplification: GenAI Affordances Mediated Through Ideological Fit, Institutional Norms and Algorithms

Beyond technical enablers, far-right advantage also stems from supply-side dynamics that shape communication styles and content strategies. GenAI's stylistic affordances align closely with far-right communication strategies, which prioritise emotional storytelling, symbolic dichotomies of 'good' versus 'evil', and the construction of threatening 'Others'.[48] As Dr. Bharath Ganesh explained how this maps onto platform logics: algorithms privilege 'easier, flashier, more outrageous' content, effectively rewarding far-right messaging strategies[49].

Feitsma argues the disproportionate alignment between AI capabilities and far right messaging becomes clear when we consider Dr. Byung-Chul Han's 'infocracy' which refers to an attention economy where algorithmic engagement replaces democratic deliberation.[50] Pieper uses the label 'epistemic crisis' tracing a shift from a public consensus mediated by journalists, politicians and academics largely agreed on macroeconomic truths to a truth economy where the public are less accepting of the elite consensus as algorithms 'benefit easier, often more outrageous explanations' of public issues.[51] In such an attention economy, the determining factor is willingness to abandon truthfulness, a technique far right parties who excel at 'half-truths' that are immune to fact-checking are poised to exploit.[52]

Conversely, mainstream parties do not adopt these messaging tactics. While not bound by technical capacities, institutional and democratic norms of factual accuracy limit their uptake of genAI to produce controversial content to avoid alienating their core voting constituencies. While established institutions are 'risk averse' and 'institutionally lagging',[53] right-wing parties have a higher 'willingness to use illegitimate tactics'.[54] Pieper highlighted the role far right's use of digital surrogacies have in furthering this as labour division can be seen where influencers spread right deviated content while official party channels maintain a symbol of legality.[55] This trap highlights a structural disadvantage of truthfulness - abandoning these norms alienates core supporters but maintaining them ensures an algorithmic disadvantage.

---

[47] Asker Bryld Staunæs, PhD in artistic research, Aarhus University and Kunsthal Aarhus, interview by Sibo Kang, Zoom, 29 July 2025

Sebastian Pieper, Project Manager, Das Progressive Zentrum, interview by Liha Leung, Zoom, 31 July 2025.

[48] Mads Fuglsang Hove, Postdoc at Department of Political Science at Aarhus University, interview by Adam Sizoo, Copenhagen, 23 July 2025

Joram Feitsma, Assistant Professor, Utrecht University School of Governance, interview by Adam Sizoo, Utrecht, 18 July 2025.

[49] Bharath Ganesh, Assistant Professor, University of Amsterdam, interview by Adam Sizoo, Utrecht, 17 July 2025.

[50] Joram Feitsma, Assistant Professor, Utrecht University School of Governance, interview by Adam Sizoo, Utrecht, 18 July 2025, Han, Byung-Chul, Infocracy: Digitalization and the Crisis of Democracy (Cambridge: Polity Press, 2022).

[51] Sebastien Pieper, Project Manager, Das Progressive Zentrum, interview by Liha Leung, Zoom, 18 July 2025.

[52] Bharath Ganesh, Assistant Professor, University of Amsterdam, interview by Adam Sizoo, Utrecht, 17 July 2025.

[53] Poppy Astrini, Gen AI Consultant, Poppy Astrini Consulting, interview by Adam Sizoo, Zoom, 24 July 2025.

[54] Oliver Marsh, Head of Tech Research, AlgorithmWatch, interview by Liha Leung, Zoom, 1 August 2025.

[55] Sebastien Pieper, Project Manager, Das Progressive Zentrum, interview by Liha Leung, Zoom, 18 July 2025.

Through the truthfulness trap we can understand that the intrinsic bias of the model matters less than the actor's relationship to truthfulness in determining who benefits from GenAI.

As highlighted in the literature review, the far right's controversial and emotive communication style is naturally amplified by platform and structural features. Right wing users have been found to engage 'like' or 'comment' more than others, while controversial posts receive more engagement from across the spectrum. Platforms, meanwhile, algorithmic reward posts that receive higher engagement, meaning that far-right content is amplified due to its epistemically controversial nature, regardless of the intrinsic appeal of their messaging. Taken together, these technical, supply-side, and demand-side dynamics create a reinforcing ecosystem in which far-right actors are uniquely positioned to exploit GenAI for political gain.

Weak Governance and Regulations: Structural Enablers of Exploitation

This section aims to tie together our sociotechnical conceptual model to trace the causal chain from technical bias to political effects by demonstrating how a permissive regulatory environment and misalignment between the object of harm and regulation can explain far-right adoption and the subsequent consequences arising from the use of GenAI in political contexts.

At the model layer, the AI Act's (AIA) focus on pre-deployment obligations and allocative bias results in epistemic harms that emerge in downstream political use to be overlooked by regulations. As the previous sections demonstrate, embedded bias exist within models but are operationalised by far-right communication strategies, which experts such as Feitsma argue is because regulation often 'miss[es] the more underlying move from democracy to infocracy',[56] highlighting how political harms emerge as by-products of technology use. Therefore, experts have questioned whether pre-deployment obligations like safety testing[57] are sufficient to regulate downstream harms, with Dormuth noting that even with technical guardrails, they can be easily bypassed through techniques such as prompt-injection.[58] Additionally, Steinacker emphasised that 'AI is not just a technical object' but embedded in human choices and incentives,[59] suggesting that technical regulation alone overlooks the way that actors choose to operationalise these technologies. These insights underscore that focusing on foreseeability in pre-deployment cannot anticipate downstream use. The AIA therefore regulates the model as object but not the communicative uses where harms arise.

Moreover, the AIA focuses on regulating the technical properties of the model layer due to its broader orientation towards ensuring allocative fairness—how benefits or harms are distributed—to mitigate algorithmic discrimination.[60] However, as our model testing results

---

[56] Joram Feitsma, Assistant Professor, Utrecht University School of Governance, interview by Adam Sizoo, Utrecht, 18 July 2025.
[57] See Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (AI Act) and amending certain Union legislative acts, OJ L 202, 12.7.2024, p. 1, art. 55 (on model evaluation, systemic risk assessment and mitigation, and documentation obligations).
[58] Ina Dormuth, Research Assistant, TU Dortmund University, interview by Liha Leung, Zoom, 25 July 2025.
[59] Léa Steinacker, Social Scientist and Entrepreneur, interview by Liha Leung, Zoom, 30 July 2025.
[60] See Annex III of the Artificial Intelligence Act, which classifies high-risk use cases such as employment, credit scoring, and access to public services.

suggest, biased outputs go beyond discriminatory decision-making, but can potentially influence voting behaviour, public discourse, and individual opinions. The AIA's focus on mitigating *allocative* bias thus leaves *epistemic* bias[61] in LLMs largely unregulated. Pieper echoes these concerns, especially given that LLMs are increasingly the public's default interface for information, yet regulation does little to govern how models can answer politically-sensitive questions.[62] This corroborates our findings of biased outputs on political topics and low refusal rates, highlighting how governance gaps risks undermining public discourse and eroding trust in democratic discourse by overlooking how biased outputs could influence users.

At the platform layer, the Digital Services Act (DSA) imposes duties to moderate illegal content and assess systemic risks, including to democracy. However, enforcement is slow and cautious, especially for lawful-but-harmful political speech. As Ganesh observed, platforms already have 'robust' community guidelines, but 'they're just not enforced.'[63] This under-enforcement is exacerbated by legal ambiguities created by the DSA, which mandates that illegal content is removed from platforms, but leaves politically sensitive 'lawful-but-awful'[64] content that much of far-right content falls under the discretion of platforms. While risk assessment obligations also require platforms to mitigate systemic risks to electoral processes and civic discourse, the 'DSA Elections Toolkit' that provided further guidance on how regulators could engage with platforms to ensure they were upholding their obligations on monitoring election-related risks was only published recently in February 2025.[65] Moreover, given the attention economy business model of many platforms, polarising content often retains more engagement, which may also reduce incentive for platforms to remove lawful-but-awful content.

---

[61] For more information on this concept, see Eugenia Villa et al., 'The Epistemic Dimension of Algorithmic Fairness: Assessing Its Impact in Innovation Diffusion and Fair Policy Making,' arXiv.org, 2025.
[62] Sebastian Pieper, Project Manager, Das Progressive Zentrum, interview by Liha Leung, Zoom, 31 July 2025.
[63] Bharath Ganesh, Assistant Professor, University of Amsterdam, interview by Adam Sizoo, Utrecht, 17 July 2025.
[64] For more information on this concept, see Martin Husovec, 'Rising Above Liability: The Digital Services Act as a Blueprint for the Second Generation of Global Internet Rules,' Berkeley Technology Law Journal 38, no. 3 (forthcoming, October 10, 2023), 37 pages, SSRN, https://ssrn.com/abstract=4598426 or http://dx.doi.org/10.2139/ssrn.4598426
[65] European Commission, 'DSA Elections Toolkit for Digital Services Coordinators | Shaping Europe's Digital Future,' Shaping Europe's Digital Future, 2025, https://digital-strategy.ec.europa.eu/en/library/dsa-elections-toolkit-digital-services-coordinators

Figure 7. Screenshot of Alternative für Deutschland (AfD) Instagram post, 12 June 2025

Figure 7 illustrates this regulatory tension, with an AfD slogan portraying Germany as an 'asylum paradise' and migrants as a 'welfare magnet'. Such rhetoric is stigmatising yet legal under EU free expression standards, and also drives engagement and outrage on platforms. Given that platforms may also fear backlash if they remove politically-sensitive content, content moderation is often slowly enforced, at which point potential harms such as polarisation or agenda distortion may already have materialised. Additionally, as GenAI enables mass production of similar slogans at low cost, this makes moderation delays even more consequential. Our findings suggest that it is precisely because of the polarising nature of such content, the platform affordances that amplifies its visibility, and the DSA's under-enforcement that explains far-right adoption of GenAI—a low-cost and low-risk campaigning tool, with lawful-but-awful content amplified in ways that drives polarisation and reinforces stigmatising narratives.

The misalignment between the object of regulation (technical risks, illegal content) and the object of harm (epistemic effects in discourse) thus creates a permissive environment for political use of GenAI. By focusing on pre-deployment testing and provider intent, the AIA leaves downstream communicative uses under-specified. Meanwhile, the DSA's systemic-risk provisions are weakened by legal ambiguity, delayed guidance, and under-enforcement, meaning polarising but lawful content remains highly visible. The interactions between technical bias, political communication, far-right strategy, and regulatory gaps make GenAI a low-cost, scalable, and effective at producing content that can be operationalised at scale to create polarising content with little regulatory friction. The resulting harms are systemic rather than intentional, but include polarisation, agenda distortion, and erosion of trust—democratic risks that remain under-captured because regulation governs the wrong object at the wrong point.

Layer	Model Affordances	Actor Operationalisation	Platform Amplification	Governance Gaps
	Presence of political bias in generative AI models	Strategic use of AI	Algorithms and recommender systems increase visibility of high-engagement content	Misalignment between the object of harm and regulation
Adoption-Driver	Attractive Affordances (eg. bias and flexibility)	Strategic fit (cheap content, low barrier to entry, mobilisation, truthfulness trap)	Amplification of content (algorithms reward outrage and emotional content)	Low risk (little regulatory friction)
Downstream Harms	Conditional political salience, harms only emerge in interaction with structural factors downstream	Manipulation of information quality, and mainstreaming of fringe narratives	Polarisation, agenda distortion, and trust erosion	Under-captured epistemic harms and potential political impacts

Table 3. Sociotechnical Stack of Generative AI in Politics: Layers, Drivers, and Harms

Conclusion

Our findings demonstrate that GenAI is reshaping political communication of far-right parties and how they interact with democratic processes. By enabling the large-scale production and amplification of polarising narratives, it risks distorting public discourse, eroding trust, and undermining the epistemic foundations of democracy. While embedded model biases demonstrate steerable tendencies and varying refusal rates to politically-sensitive questions, far-right adoption and the capabilities of these technologies to create downstream impacts in politics must be understood in a broader sociotechnical structure that encompasses actors, platforms, and regulations.

Our key contribution to broader debates on AI governance and democratic resilience is to connect these layers empirically and conceptually, demonstrating that far-right adoption is less about ideological bias in models and more about strategic fit within permissive sociotechnical conditions. Our quantitative findings show embedded biases may facilitate the generation of more extremist narratives, which maps onto the actor layer where resource-restricted fringe parties in the far-right may be more willing to use cost-effective technologies like GenAI than mainstream parties, resulting in what we have described as the truthfulness trap. At the platform layer, our qualitative findings show that the amplification of polarising narratives as a result of engagement-driven algorithms explains the value of GAIC to mobilise political participation through ‘lawful-but-awful’ content, which regulation is undercaptured due to current legislation’s focus on upstream technical governance and weak enforcement of content moderation obligations.

By tracing this pathway from model affordances to systemic epistemic harms, we show how current regulation leaves a permissive environment for far-right exploitation. Addressing this gap demands concrete policy/societal responses, which we set out below in Annex A.

While our paper shows how different layers of the sociotechnical stack interact, constraints of time and scope limit causal explanations at the user level and in-depth testing of multimodal outputs. This points to several priorities for further research. First, while we have

identified a right-leaning bias in our model testing, we were unable to identify what factors could have contributed to this type of bias. Future work should examine whether such tendencies are shaped by training data, process, or feedback loops from far-right usage, in order to assess the significance of these biases and how human-computer interaction may reinforce them. Second, although we demonstrate that GenAI usage in politics can result in misinformation and polarising narratives, further research is needed to understand how politically-biased outputs from LLMs or AI-generated images may affect users' trust, perceptions of credibility, and voting behaviour. Third, future studies should further explore the dynamics of the 'truthfulness trap' and clarify whether far-right parties' willingness to employ misleading confers a measurable advantage in algorithmic attention economies. Finally, while our research has been conducted with a static conception of political bias and misinformation, critical perspectives of these concepts should be explored i.e are bias and misinformation intrinsically problematic or are they instrumental articulations of alternative views on society which are not compatible with hegemonic conceptions of factuality.

Bibliography

Primary Sources

Astrini, Poppy. Gen AI Consultant, Poppy Astrini Consulting. Interview by Adam Sizoo. Zoom, July 24, 2025.

Dormuth, Ina. Research Assistant, TU Dortmund University. Interview by Liha Leung. Zoom, July 25, 2025.

European Commission. 2025. 'DSA Elections Toolkit for Digital Services Coordinators | Shaping Europe's Digital Future.' *Shaping Europe's Digital Future*. https://digital-strategy.ec.europa.eu/en/library/dsa-elections-toolkit-digital-services-coordinators.

European Union. 2022. 'Regulation (EU) 2022/2065 on a Single Market for Digital Services (Digital Services Act).' *EUR-Lex*. https://eur-lex.europa.eu/eli/reg/2022/2065/oj/eng.

———. 2024. 'Regulation (EU) 2024/1689 Laying Down Harmonised Rules on Artificial Intelligence (AI Act).' *EUR-Lex*. https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng.

Feitsma, Joram. Assistant Professor, Utrecht University School of Governance. Interview by Adam Sizoo. Utrecht, July 18, 2025.

Ganesh, Bharath. Assistant Professor, University of Amsterdam. Interview by Adam Sizoo. Utrecht, July 17, 2025.

Hove, Mads Fulsgang. Postdoctoral Researcher, Aarhus University. Interview by Adam Sizoo. Copenhagen, July 23, 2025.

Marsh, Oliver. Head of Tech Research, AlgorithmWatch. Interview by Liha Leung. Zoom, August 1, 2025.

Pieper, Sebastien. Project Manager, Das Progressive Zentrum. Interview by Liha Leung. Zoom, July 18, 2025.

Staunæs, Asker Bryld. Artistic Researcher, Aarhus University. Interview by Sibo Kang. Zoom, July 29, 2025.

Steinacker, Léa. Social Scientist and Entrepreneur. Interview by Liha Leung. Zoom, July 30, 2025.

Secondary Sources

Allchorn, Will. 2024. 'Global Far-Right Extremist Exploitation of Artificial Intelligence and Alt-Tech.' *Counter Terrorist Trends and Analyses* 16, no. 3: 13–18.

Bang, Yejin, Delong Chen, Nayeon Lee, and Pascale Fung. 2024. 'Measuring Political Bias in Large Language Models: What Is Said and How It Is Said.' January, 11142–59. https://doi.org/10.18653/v1/2024.acl-long.600.

Barberá, Pablo, John Jost, and Richard Bonneau. 2015. 'Tweeting from Left to Right: Is Online Political Communication More Than an Echo Chamber?' *Psychological Science* 26, no. 10: 1531–41. https://doi.org/10.1177/0956797615594620.

Black, Julia, and Andrew D. Murray. 2019. 'Regulating AI and Machine Learning: Setting the Regulatory Agenda.' *European Journal of Law and Technology* 10, no. 3 (December): 1–21.

Burris, Val, Emery Smith, and Ann Strahm. 2000. 'White Supremacist Networks on the Internet.' *Sociological Focus* 33, no. 2: 215–35. https://www.jstor.org/stable/20832076.

Chadwick, Andrew. 2017. *The Hybrid Media System: Politics and Power*. Oxford: Oxford University Press.

Chesterman, Simon. 2024. 'Lawful but Awful: Evolving Legislative Responses to Address Online Misinformation, Disinformation, and Mal-Information in the Age of Generative AI.' *The American Journal of Comparative Law* 72, no. 4 (Winter): 933–65.

Conway, Maura, Ryan Scrivens, and Logan Macnair. 2019. 'Right-Wing Extremists' Persistent Online Presence: History and Contemporary Trends.' ICCT Policy Brief. The Hague: International Centre for Counter-Terrorism. https://www.icct.nl/publication/right-wing-extremists-persistent-online-presence-history-and-contemporary-trends.

Cupać, Jelena, and Mitja Sienknecht. 2024. 'Regulate against the Machine: How the EU Mitigates AI Harm to Democracy.' *Democratization* 31, no. 5 (June): 1067–90.

Deck, Lisa, Markus-Philipp Forster, Niklas Kuhl, and Rainer Weidlich. 2025. *A Multi-Level Strategy for Deepfake Content Moderation under EU Regulation*. Bayreuth: University of Bayreuth; Fraunhofer FIT; FIM Research Center.

Demir, Mustafa. 2025. 'From Ideology to Algorithms, The New Face of Fascism: Digital Fascism.' *Journal of Social Sciences of Mus Alparslan University* 13, no. 1: 232–72. https://doi.org/10.18506/anemon.1619446.

Douek, Evelyn. 2022. 'Content Moderation as Systems Thinking.' *Harvard Law Review* 136, no. 2 (December): 528–606.

Errayes, Ilyass. 2025. *The Algorithmic Ballot: Fortifying Democratic Elections Against AI-Driven Disruption*. SSRN 5273244.

Feitsma, Joram. 2023. 'De PVV en het primaire verlangen naar eenvoud.' *Sociale Vraagstukken*, June 29. https://www.socialevraagstukken.nl/de-pvv-en-het-primaire-verlangen-naar-eenvoud/.

Fisher, Jillian, Ruth E. Appel, Chan Young Park, Yujin Potter, Liwei Jiang, Taylor Sorensen, Shangbin Feng, et al. 2025. 'Political Neutrality in AI Is Impossible—but Here Is How to Approximate It.' *arXiv*, 2025. https://arxiv.org/abs/2503.05728.

Fisher, Jillian, Shangbin Feng, Robert Aron, Thomas Richardson, Yejin Choi, Daniel W. Fisher, Jennifer Pan, Yulia Tsvetkov, and Katharina Reinecke. 2024. 'Biased AI Can Influence Political Decision-Making.' *arXiv*, 2024. https://arxiv.org/abs/2410.06415.

Freelon, Deen, Alice Marwick, and Daniel Kreiss. 2020. 'False Equivalencies: Online Activism from Left to Right.' *Science* 369, no. 6508: 1197–1201. https://doi.org/10.1126/science.abb2428.

Gerbaudo, Paolo, Ciro Clemente De Falco, and Federica Nunziata. 2023. 'Angry Posts Mobilise: Emotional Communication and Online Mobilization in the Facebook Pages of Western European Right-Wing Populist Leaders.' *Social Media + Society* 9, no. 1: 1–11. https://doi.org/10.1177/20563051231163327.

Gillespie, Tarleton. 2010. 'The Politics of ‘Platforms.'' *New Media & Society* 12, no. 3: 347–64.

Gonzalez-Bailon, Sandra, Valeria d’Andrea, Deen Freelon, and Manlio De Domenico. 2022. 'The Advantage of the Right in Social Media News Sharing.' *PNAS Nexus* 1, no. 3: 1–8. https://doi.org/10.1093/pnasnexus/pgac137.

Hacker, Philipp, Andreas Engel, and Theresa List. 2023. 'Understanding and Regulating ChatGPT, and Other Large Generative AI Models: With Input from ChatGPT.' *Verfassungsblog*. https://verfassungsblog.de/chatgpt/.

Han, Byung-Chul. 2022. *Infocracy: Digitalization and the Crisis of Democracy*. Cambridge: Polity Press.

Hersch, Gil, and Thomas Rowe. 2025. 'Allocative Fairness.' *Philosophy Compass* 20, no. 5 (May): 1–9.

Hiller, Anna, and Pablo Maristany de las Casas. 2025. *Generative AI and the German Far Right: Narratives, Tactics and Digital Strategies*. London: Institute for Strategic Dialogue, February 18. https://www.isdglobal.org/publications/generative-ai-and-the-german-far-right/.

Husovec, Martin. 2024. 'The Digital Services Act's Red Line: What the Commission Can and Cannot Do about Disinformation.' *Journal of Media Law* 16, no. 1 (January): 47–56.

Jonnala, Sridhar, Nisha Thomas, and Pramod Parida. 2024. 'EU AI Act: Underrepresented and Insufficient to Address the Risks and Vulnerabilities of Generative AI.' SSRN.

Kakavand, Azade. 2024. 'Far-Right Social Media Communication in the Light of Technology Affordances: A Systematic Literature Review.' *Annals of the International Communication Association* 48, no. 1: 37–56. https://doi.org/10.1080/23808985.2023.2280824.

Karpf, David. 2016. *Analytic Activism: Digital Listening and the New Political Strategy*. Oxford: Oxford University Press.

Knüpfer, Curd, and Ulrike Klinger. 2025. 'In-Groups and Outrage: How Narratives and Affect Shape Digital Surrogate Networks and Radicalize Right-Wing Parties.' In *Connective Action and the Rise of the Far-Right: Platforms, Politics, and the Crisis of Democracy*, edited by Steven Livingston and Michael Miller, 122–45. New York: Oxford University Press. https://doi.org/10.1093/oso/9780197794937.003.0005.

Koblentz-Stenzler, Liram, and Uri Klempner. 2024. 'Navigating Far-Right Extremism in the Era of Artificial Intelligence.' GNET Insight. London: Global Network on Extremism and Technology. https://gnet-research.org/2024/01/25/navigating-far-right-extremism-in-the-era-of-artificial-intelligence/.

Kusche, Isabel. 2024. 'Possible Harms of Artificial Intelligence and the EU AI Act: Fundamental Rights and Risk.' *Journal of Risk Research*, May: 1–14.

Lai, Samantha, Ben Nimmo, Alicia Wanless, Alexandre Alaphilippe, Samantha Bradshaw, and David A. Broiatowski. 2025. 'Measuring Changes Caused by Generative Artificial Intelligence: Setting the Foundations.' Carnegie Endowment for International Peace.

Leser, Julia, and Florian Spissinger. 2020. 'The Functionality of Affects: Conceptualising Far-Right Populist Politics beyond Negative Emotions.' *Global Discourse* 10, no. 2: 325–42. https://doi.org/10.1332/204378919X15762350844101.

Lin, Luyang, Lingzhi Wang, Jinsong Guo, and Kam-Fai Wong. 2024. 'Investigating Bias in LLM-Based Bias Detection: Disparities between LLMs and Human Perception.' *arXiv*, March. https://doi.org/10.48550/arxiv.2403.14896.

Livingston, Steven, and Michael Miller. 2025. 'Conclusion.' In *Connective Action and the Rise of the Far-Right: Platforms, Politics, and the Crisis of Democracy*, edited by Steven Livingston and Michael Miller, 286–302. New York: Oxford University Press. https://doi.org/10.1093/oso/9780197794937.003.0013.

Lopes Buarque, Beatriz. 2025. 'AI and the Far-Right Riots in the UK.' *LSE Media Blog*, June 9, 2025. https://blogs.lse.ac.uk/medialse/2025/06/09/ai-and-the-far-right-riots-in-the-uk/.

Momeni, Mina. 2025. 'Artificial Intelligence and Political Deepfakes: Shaping Citizen Perceptions through Misinformation.' *Journal of Creative Communications* 20, no. 1 (March): 41–56.

Quinn, Ben, and Dan Milmo. 2024. 'How the Far Right Is Weaponising AI-Generated Content in Europe.' *The Guardian*, November 26, 2024. https://www.theguardian.com/technology/2024/nov/26/far-right-weaponising-ai-generated-content-europe.

Rozado, David. 2023. 'The Political Biases of ChatGPT.' *Social Sciences* 12, no. 3: 148. https://doi.org/10.3390/socsci12030148.

Strick, Simon. 2022. 'Reflexive Fascism in the Age of History Memes.' *Journal of Modern European History* 20, no. 3: 335–51. https://doi.org/10.1177/16118944221110451.

The Economist. 2023. 'Europe: Elections to Watch in 2024.' *Economist Intelligence Unit*. https://www.eiu.com/n/europe-elections-to-watch-in-2024/.

Villa, Eugenia, Camilla Quaresmini, Valentina Breschi, Viola Schiaffonati, and Mara Tanelli. 2025. *The Epistemic Dimension of Algorithmic Fairness: Assessing Its Impact in Innovation Diffusion and Fair Policy Making*. *arXiv*. https://arxiv.org/abs/2504.02856.

Wang, Jennifer, Andrew Selbst, Solon Barocas, and Suresh Venkatasubramanian. 2025. 'Distinguishing Predictive and Generative AI in Regulation.' *arXiv*, July 2. https://arxiv.org/abs/2506.17347.

Wihbey, John. 2024. 'AI and Epistemic Risk for Democracy: A Coming Crisis of Public Knowledge?' *SSRN Electronic Journal*, January.

Yan, Harry Y., Garrett Morrow, Kai Yang, and John Wihbey. 2025. 'The Origin of Public Concerns over AI Supercharging Misinformation in the 2024 U.S. Presidential Election.' *HKS Misinformation Review*. https://misinforeview.hks.harvard.edu/article/the-origin-of-public-concerns-over-ai-supercharging-misinformation-in-the-2024-u-s-presidential-election/.

Annex A. Policy Recommendations

1. AI literacy should be integrated into public education to strengthen citizens' resilience against manipulative uses of generative AI in political communication. Public programs should equip individuals with critical skills to identify synthetic content, understand algorithmic amplification, and evaluate sources. Furthermore, policymakers should ensure accessible resources for vulnerable groups, who are often most exposed to disinformation. Building AI literacy across society may not eliminate risks completely, but it will narrow the gap between political actors, platform and application developers, regulators and individual users..
2. Rethinking campaign strategies - other parties could consider incorporating tactics that leverage virality-encouraging platform affordances without sacrificing their principled ideals they hold for society.
3. Regulations should focus on the object of harm—while liability is notoriously difficult to assign in the context of the complex AI value chain, assigning clear responsibilities and obligations based on the societal impact of the creation or dissemination of AI content would be crucial to mitigate epistemic harms of GenAI use in politics. Moreover, regulatory scrutiny of content moderation should extend beyond platforms to include the textual outputs of LLMs. This could involve broadening the scope of the DSA to cover generative AI systems, ensuring that their proven biases and role in amplifying polarising narratives are effectively governed.