The Role of Social Media Algorithms in Amplifying Misinformation in Political Elections

Ruoheng Yang

Krieger School of Arts and Sciences, Johns Hopkins University, Washington, USA
ryang50@jh.edu

Abstract. Social media platforms curate political information through engagement-optimising algorithms that determine what voters see and share. Evidence from computational communication and behavioural science demonstrates that false and sensational content spreads faster and farther than verified information, propelled less by automated actors than by human attention dynamics that algorithms privilege. This paper synthesises findings from twenty-two peer-reviewed studies across political communication, psychology, and information science to situate election misinformation within an engagement-first platform ecology. It reviews mechanisms by which recommender systems and feed ranking elevate emotionally charged posts; examines consequences for opinion formation, polarisation, and institutional trust across the United States, the United Kingdom, Brazil, and India; and articulates a theoretically grounded agenda for empirical assessment. Building on this review, the paper specifies research questions that connect algorithmic amplification to voter-level exposure, belief change, and trust outcomes, and proposes a mixed-methods design integrating trace data, audit experiments, and voter surveys. The contribution clarifies how engagement metrics function as algorithmic signals that advantage misinformation at scale and outlines an evaluative framework for mitigation levers including transparency, friction, and literacy interventions.

Keywords: algorithmic amplification, election misinformation, engagement metrics

1. Introduction

Social platforms originally designed for interpersonal connection now mediate political attention through algorithmic curation [1,2]. During elections, ranking systems that optimise for engagement systematically elevate content that provokes anger, fear, or surprise, irrespective of accuracy [3,4]. Empirical diffusion studies reveal that falsehoods travel farther and faster than truths because human reactions—likes, comments, and shares—serve as quality signals to algorithms [5,6]. These mechanisms have acute implications for democratic practice: algorithmically elevated misinformation can distort candidate perception, deepen polarisation, and erode confidence in institutions [7,8]. This paper consolidates current knowledge on algorithmic amplification, derives testable research questions, and specifies a feasible empirical design to evaluate both harms and remedies.

2. Literature review and theoretical framework

Research on diffusion dynamics demonstrates that engagement-optimised systems amplify emotionally charged misinformation. False stories spread six times faster than verified ones on Twitter because users preferentially share novel or affective content [5,9]. Emotional contagion studies confirm that outrage and fear predict comment and reshare behaviour, thereby reinforcing algorithmic ranking loops [4]. Large-scale audits of YouTube's recommendation system find that users starting with moderate political content are gradually steered toward more extreme material [10].

Cross-platform analyses suggest that these mechanisms stem from algorithmic optimisation itself: engagement metrics act as proxies for relevance, making affective intensity a structural advantage [11,12]. Bakshy et al. show that Facebook's News Feed reduces exposure to ideologically diverse information [2], contributing to echo chambers [3]. The combined effect of emotional resonance and network homophily forms what Lazer et al. call an "attention economy distortion," privileging virality over veracity [1].

Despite these insights, limited research directly compares how engagement-prioritising algorithms influence users' exposure to political misinformation relative to verified information during election periods—contexts where accuracy is most consequential.

RQ1: How do engagement-prioritising algorithms shape individual exposure to political misinformation during election periods relative to verified information?

Empirical evidence from multiple countries links algorithmic amplification to electoral misinformation. In the U.S. 2016 presidential election, engagement with fake news sites rivalled mainstream outlets, particularly among low-information voters [7,13]. Similar amplification shaped the Brexit referendum, where emotionally provocative claims received greater reach than factual corrections [14]. In Brazil's 2018 general election, disinformation circulated through encrypted WhatsApp groups beyond the reach of moderation [15]. In India's 2019 election, a field experiment showed that targeted media-literacy training reduced belief in false claims by 25 percent [16]. Collectively, these cases reveal how platform architecture and emotional incentives interact to affect voter cognition and participation.

Broda and Strömbäck summarise these cross-disciplinary findings: algorithmic ranking amplifies misinformation when affective engagement outweighs epistemic quality [17]. Pennycook and Rand further note that human cognitive shortcuts—limited reasoning and motivated attention—amplify these effects [18]. Consequently, misinformation becomes a socio-technical phenomenon, shaped by both algorithmic design and user psychology.

While prior studies document exposure patterns and cognitive biases, less is known about their cumulative impact on democratic attitudes. Specifically, it remains unclear how algorithmically amplified misinformation influences voters' beliefs, perceived polarisation, and institutional trust.

RQ2: To what extent does algorithmically amplified misinformation influence voters' beliefs, perceived polarisation, and trust in democratic institutions?

Beyond behavioural outcomes, the amplification of false content undermines trust in democratic institutions. De Ridder argues that repeated exposure to misleading claims erodes citizens' sense of epistemic security [8]. Tsfati and Cappella show that even sceptical audiences continue consuming news they distrust, complicating attempts to rebuild credibility [19]. Reisach and Livingston & Bennett emphasise platform accountability and the ethical obligation to audit engagement-driven systems [20,21]. Deepfakes and synthetic video escalate these concerns by blurring perceptual boundaries between fact and fabrication [22].

Taken together, the literature supports three propositions: (1) engagement optimisation structurally advantages sensational misinformation; (2) the consequences vary with user habits, prior beliefs, and platform affordances; and (3) mitigation levers—transparency, friction, and literacy—offer partial remedies but must be context-specific. Yet empirical evidence comparing the relative efficacy of these mitigation strategies across platform environments remains scarce.

RQ3: How effective are mitigation strategies—such as algorithmic transparency, sharing frictions, and media-literacy prompts—in reducing belief in misinformation and curbing its further spread across different platform environments?

3. Method

A mixed-methods design triangulates exposure, belief, and behaviour across data sources. The first component involves a large-scale trace analysis of public posts and reshares during defined pre-election windows, in which verified and false claims are coded using fact-checking databases [12]. Virality is modelled as a function of emotional content and engagement, enabling comparison of diffusion patterns for true versus false items [5,3].

The second component is an audit experiment using controlled accounts that follow balanced political seed sets to record recommended content. This procedure quantifies how often neutral viewing paths lead to extreme or misleading material, thereby identifying algorithmic pathways and the frequency of corrective exposure [10].

The third component is a voter survey designed to measure media diets, platform use, institutional trust, and belief accuracy [13,19]. Within the survey, embedded experimental manipulations vary transparency notices and sharing frictions to test their effects on belief correction and sharing intentions [20].

Finally, qualitative interviews with a purposive subset of participants explore how voters interpret feed signals and weigh accuracy against identity-affirming narratives [6,5]. Together, these interconnected components enable cross-validation of algorithmic exposure, cognitive outcomes, and intervention efficacy while maintaining coherence across methodological levels.

4. Synthesis and discussion

The integrated design addresses methodological gaps noted by Broda and Strömbäck [17]: diffusion studies capture scale and velocity but not cognition or belief formation; surveys capture attitudes and trust but cannot verify actual exposure; and audit experiments identify algorithmic pathways without connecting them to individual-level psychological outcomes. By nesting these approaches within a unified framework, the study makes it possible to trace how algorithmic choices at the platform level translate into belief change, polarisation, and institutional trust at the voter level.

Findings from Allcott and Gentzkow, Guess et al., and Höller converge on the notion that engagement metrics function as noisy proxies for relevance, thereby producing a structural bias towards emotionally intense or identity-congruent content [7,13,14]. This bias is not uniform: exposure effects are moderated by prior media diets, partisanship, and cognitive style [11]. Users who primarily consume partisan or algorithmically filtered feeds exhibit stronger reinforcement effects, while those with diverse information diets demonstrate greater resilience. Closed networks such as WhatsApp, which lack algorithmic ranking but enable high-speed forwarding, shift the feasible locus of intervention from algorithmic tuning to community-level literacy and peer moderation [15].

Emotional amplification, as articulated by Metzler and Garcia, transforms attention into a currency of visibility, rewarding outrage and novelty [4]. At the same time, cognitive shortcuts such as heuristic reasoning and reduced deliberation—what Pennycook and Rand term "cognitive laziness" [6]—sustain the viral diffusion of misinformation even in the absence of deliberate bias. These dual processes—algorithmic optimisation and human cognition—jointly perpetuate misinformation cycles that are difficult to disrupt through a single-channel solution.

Transparency interventions alone rarely change beliefs, yet when combined with structural frictions that slow impulsive sharing, they can significantly dampen onward diffusion [16,20]. Embedding literacy programmes within the everyday communication ecosystems that voters already use—schools, civic groups, and local media—can strengthen critical evaluation skills, though their long-term durability and scalability remain uncertain [22]. Ethical frameworks further emphasise that platforms bear a normative duty to disclose their ranking logics and to allow independent auditing of algorithmic systems [1,21]. Together, these principles define a research and governance agenda that aligns empirical evidence with practical mechanisms for restoring informational integrity in democratic environments.

5. Policy and design implications

Election regulators need auditable visibility into how political content is curated during sensitive periods such as pre-election windows and candidate debates [20]. Regulatory oversight should include access to platform-level transparency reports that document the ranking and recommendation criteria used for political content, as well as aggregate data on amplification patterns. Such reports can disclose the ratio of verified to debunked claim reach, the velocity of false information diffusion, and the visibility of corrective or fact-checked materials. Public release of these privacy-preserving indicators would not only enable independent monitoring but also allow comparative evaluation across platforms and election cycles.

Social media companies, for their part, can assume a proactive role by integrating friction and transparency mechanisms directly into the user interface. One approach is to embed lightweight prompts that encourage users to pause and read before sharing politically charged or emotionally arousing content [18]. These "think-before-you-share" notices, already tested in other behavioural domains, operate as micro-interventions that slow impulsive diffusion without restricting speech. Designers can also include cognitive "speed bumps"—brief interruptions that require minimal confirmation clicks—to delay the automatic reshare of content identified as highly emotive or lacking external verification [4]. Over time, such design elements can recalibrate the pace of information flow, discouraging the viral spread of unverified material while maintaining users' sense of autonomy.

Beyond platform governance, civil-society organisations and educational institutions can play a complementary role by advancing digital literacy initiatives. Modular media-literacy programmes modelled after the Indian field experiment [16] can be deployed through schools, civic campaigns, and community centres to train citizens in identifying emotional manipulation, distinguishing between credible and dubious sources, and understanding how algorithms prioritise engagement. Tailoring these programmes to local media ecologies—using examples drawn from domestic political discourse—enhances both cultural relevance and retention.

These interventions, however, should not be treated as substitutes for one another. Transparency without friction may inform but not deter, while literacy without structural change places the burden solely on users. Likewise, algorithmic redesign absent public accountability risks reproducing opaque hierarchies of information power. None of these strategies suffice in isolation; only when combined do they constitute a pragmatic, multi-layered agenda that mirrors how amplification actually occurs—through the interaction of technological systems, human psychology, and institutional oversight. Such coordination among regulators, designers, and educators is essential to restore trust, reduce exposure inequality, and align platform incentives with democratic integrity.

6. Limitations and future research

The framework remains quasi-experimental and partly observational for public-feed dynamics, limiting the ability to make definitive causal claims about algorithmic effects [17]. Although the mixed-methods approach strengthens internal validity through triangulation, it still depends on digital trace data that may incompletely capture user exposure patterns. Self-report measures of belief and trust also risk underestimating susceptibility to misinformation due to social desirability bias or retrospective rationalisation [19].

Encrypted messaging platforms pose an additional challenge, as the privacy features that protect user communication also restrict external observation, thereby constraining visibility into private virality and group-level coordination [15]. To address these gaps, future research should examine the longitudinal persistence of belief change across multiple election cycles and employ controlled cross-platform comparisons that hold content constant while varying affordances such as visibility and feedback cues [3].

Moreover, collaborative, privacy-preserving audits involving researchers, regulators, and platforms could provide a path toward credible causal inference. Such partnerships would enable real-time evaluation of recommendation pathways and intervention effects under ecologically valid conditions [10,1]. Developing these infrastructures of transparency and cooperation remains essential to advancing the empirical study of algorithmic amplification and its democratic implications.

7. Conclusion

Engagement-first curation transforms emotional arousal into a currency of visibility. In electoral contexts, that logic systematically advantages misinformation over accuracy, not from platform malice but from algorithmic optimisation for virality [5,6]. The reviewed literature—spanning diffusion studies, audits, election analyses, and interventions—offers a coherent account of amplification and a blueprint for empirical testing. Strengthening democratic resilience requires calibrated platform design, timely literacy, and transparent evaluation tailored to both public feeds and encrypted groups [11,20].

References

[1] Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J., & Zittrain, J. L. (2018). The science of fake news. Science, 359(6380), 1094–1096. https: //doi.org/10.1126/science.aao2998

[2] Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. Science, 348(6239), 1130–1132. https: //doi.org/10.1126/science.aaa1160

[3] Cinelli, M., De Francisci Morales, G., Galeazzi, A., Quattrociocchi, W., & Starnini, M. (2021). The echo chamber effect on social media. Proc. Natl. Acad. Sci. U.S.A., 118(9), e2023301118. https: //doi.org/10.1073/pnas.2023301118

[4] Metzler, H., & Garcia, D. (2023). Social drivers and algorithmic mechanisms on digital media. Perspect. Psychol. Sci., 19(5), 735–748. https: //doi.org/10.1177/17456916231185057 (Original work published 2024)

[5] Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. Science, 359(6380), 1146–1151. https: //doi.org/10.1126/science.aap9559
[6] Pennycook, G., & Rand, D. G. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. Cognition, 188, 39–50. https: //doi.org/10.1016/j.cognition.2018.06.011
[7] Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. J. Econ. Perspect., 31(2), 211–236. https: //doi.org/10.1257/jep.31.2.211
[8] de Ridder, J. (2021). What's so bad about misinformation? Inquiry, 67(9), 2956–2978. https: //doi.org/10.1080/0020174X.2021.2002187
[9] Langin, K. (2018, March 8). Fake news spreads faster than true news on Twitter—thanks to people, not bots. Science News: Social Sciences. https: //www.science.org/content/article/fake-news-spreads-faster-true-news-twitter-thanks-people-not-bots
[10] Ribeiro, M. H., Ottoni, R., West, R., Almeida, V. A. F., & Meira, W. (2020). Auditing radicalization pathways on YouTube. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT '20)* (pp. 131–141). Assoc. Comput. Mach.. https: //doi.org/10.1145/3351095.3372879
[11] Freelon, D., & Wells, C. (2020). Disinformation as Political Communication. Polit. Commun., 37(2), 145–156. https: //doi.org/10.1080/10584609.2020.1723755
[12] Shu, K., Bhattacharjee, A., Alatawi, F., Nazer, T. H., Ding, K., Karami, M., & Liu, H. (2020). Combating disinformation in a social media age. Wiley Interdiscip. Rev. Data Min. Knowl. Discov., 10(6), e1385. https: //doi.org/10.1002/widm.1385
[13] Guess, A. M., Nyhan, B., & Reifler, J. (2020). Exposure to untrustworthy websites in the 2016 U.S. election. Nat. Hum. Behav., 4(5), 472–480. https: //doi.org/10.1038/s41562-020-0833-x
[14] Höller, M. (2021). The human component in social media and fake news: the performance of UK opinion leaders on Twitter during the Brexit campaign. Eur. J. Engl. Stud., 25(1), 80–95. https: //doi.org/10.1080/13825577.2021.1918842
[15] Santos, G. F. (2020). Social media, disinformation, and regulation of the electoral process: A study based on 2018 Brazilian election experience. Rev. Inf. Comun., 7(2), 1–18. https: //doi.org/10.5380/rinc.v7i2.71057
[16] Badrinathan, S. (2021). Educative interventions to combat misinformation: Evidence from a field experiment in India. Am. Polit. Sci. Rev., 115(4), 1325–1341. https: //doi.org/10.1017/S0003055421000459
[17] Broda, E., & Strömbäck, J. (2024). Misinformation, disinformation, and fake news: Lessons from an interdisciplinary, systematic literature review. Ann. Int. Commun. Assoc., 48(2), 139–166. https: //doi.org/10.1080/23808985.2024.2323736
[18] Pennycook, G., & Rand, D.G. (2021). The psychology of fake news. Trends Cogn. Sci., 25(5), 388–402. https: //doi.org/10.1016/j.tics.2021.02.007
[19] Tsfati, Y., & Cappella, J. N. (2005). Why Do People Watch News They Do Not Trust? The Need for Cognition as a Moderator in the Association Between News Media Skepticism and Exposure. Media Psychol., 7(3), 251–271. https: //doi.org/10.1207/S1532785XMEP0703_2
[20] Reisach, U. (2021). The responsibility of social media in times of societal and political manipulation. Eur. J. Oper. Res., 291(3), 906–917. https: //doi.org/10.1016/j.ejor.2020.09.020
[21] Bennett, W. L., & Livingston, S. (Eds.). (2020). The disinformation age: Politics, technology, and disruptive communication in the United States. Cambridge University Press. https: //doi.org/10.1017/9781108914628
[22] Vaccari, C., & Chadwick, A. (2020). Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News. Soc. Media Soc., 6(1). https: //doi.org/10.1177/2056305120903408 (Original work published 2020)