

DIGITAL PLATFORMS AND POLITICAL POLARIZATION: A NARRATIVE REVIEW OF ALGORITHMS, MISINFORMATION, AND DEMOCRATIC RESILIENCE

Zubair Ahmad¹, Abdul Haseeb², Hafiz Muhammad Javed³, Arsalan Maqsood⁴,
Muhammad Amin⁵, Taimoor Ameen⁶

¹Researcher at Università degli Studi del Sannio di Benevento, Italy

^{2,4,5,6}Department of Engineering, Yaseen Nexus Institute, Peshawar, Pakistan

³Qurtuba University D.I. Khan

¹zubair.phdstudent@lum.it, ²ah909375@gmail.com, ³Javedhafiz232@gmail.com

¹ORCID ID: 0000-0003-3754-0396

Corresponding Author: *

Abdul Haseeb

DOI: <https://doi.org/10.5281/zenodo.22055635>

Received	Accepted	Published
26 January 2026	11 March 2026	25 March 2026

ABSTRACT

The political impact of digital platforms has shifted from the periphery of academic discussion to the centre of public debates, but it is not as clear-cut as is popularly assumed. The paper is a critical narrative review of the linkage between digital platforms and political polarisation, based on a synthesis of the evidence and scholarship that have emerged in political science, communication studies, computational social science and platform governance research since 2015 and prior to 2025. The review does not view polarisation as a single phenomenon but rather as a combination of ideological, affective, perceived, and network-based forms of polarisation and poses the question about whether it is algorithmic curation, filter bubbles, or misinformation that causes polarisation, exacerbates it, or simply mirrors the polarisation that already exists. Algorithmic amplification does not seem to be a uniform cause but an interaction with existing partisan identities and selective user behaviour; false political content, though measurable and consequential in terms of the trust of institutions, is not as voluminous and not as widespread as is suggested by their public alarm; the formation of echo chambers is real for some users but not widespread; and public exposure to this false political content is smaller in volume and more demographically concentrated than public alarm implies. Affective polarisation, by contrast, as well as a decrease in institutional trust and platform business incentives, is more consistent than ideology per se. The review looks at the various platform-level, institutional and individual interventions for strengthening democratic resilience and proposes that none of these alone is enough to make a difference and that the impact of these interventions is strongly dependent on the national media systems and institutional strength. Prioritised research needs include gathering evidence from outside North America and Western Europe. While under certain circumstances, digital platforms can exacerbate polarisation, polarisation is more properly viewed as a socio-technical phenomenon that is a product of platform design, political culture and political institutional context, rather than a direct technological effect.

Keywords: Digital platforms; political polarization; affective polarization; algorithms; misinformation; disinformation; echo chambers; democratic resilience; platform governance

Introduction

Much of the world's population now gets its political information and engages in civic activity, in however loose a way, via digital platforms. The platforms of Facebook, X, YouTube, TikTok, WhatsApp and even AI-mediated platforms select not only which political information content is shown but also the way in which it is shown, as well as the order. This is an issue that has become one of the most hotly debated in modern social science: whether platforms are driving polarisation—in fact, whether they are driving it—in many democracies. The intuition doesn't sound far-fetched. Commercial platforms have an interest in engaging as many people as they can, and when political content is highly charged, emotionally it is more likely to be distributed widely and quickly than non-political content (Vosoughi et al., 2018). Cross-national data indicate that affective polarisation has increased significantly in some democracies and only slightly in others over the same time period as the penetration of social media has increased almost everywhere (Boxell et al., 2024), while large-scale platform data suggest that algorithmic curation plays a smaller role in ideological segregation than user choice does (Bakshy et al., 2015; González-Bailón et al., 2023). All this makes a simple causal narrative more complicated.

This review assumes that neither of these two dominant lines of public discourse – that platforms are causally responsible for a polarised and misinformed public nor that platforms are merely a passive reflection of offline reality – is tenable. Digital platforms should be viewed as one among a number of institutions, located within specific political economies and media systems, that can reinforce, shift, and sometimes hinder deeper institutional and social forms of division. The evidentiary basis is largely drawn from the United States, which leads to a lack of confidence in claims made for the entire world. This critical review of algorithmic curation, filter bubbles and selective exposure, misinformation, platform business models, affective and identity-based polarisation, emerging algorithmic technologies, and the mechanisms that might support democratic resilience goes beyond the question of "platforms: good or bad for democracy" to offer a

more nuanced picture of which mechanisms are likely to work under which conditions.

Narrative Review Method and Approach

This paper is a critical narrative review, not a systematic review, and is not a fully protocol-driven synthesis, as the literature base is too wide-ranging to define 'polarisation' in a consistent manner and is mostly incommensurable in methodology. The literature was selected by conducting structured searches in Google Scholar, Web of Science and Scopus with the following search terms: "political polarisation", "affective polarisation", "algorithm", "filter bubble", "echo chamber", "misinformation", "disinformation", "computational propaganda", and "democratic resilience"; and by citation searching on key reviews (Kubin & von Sikorski, 2021; Tucker et al., 2018; Iyengar et al., 2019) and from institutional sources such as the Reuters Institute, the Pew Research Center, the OECD, UNESCO, and the European Commission. Priority was given to peer-reviewed studies and major university-press books from 2015 onward using large-scale behavioural data, cross-national designs, or pre-registered experiments, since these offer the strongest basis for distinguishing correlation from causation; earlier foundational work (Sunstein, 2017; McPherson et al., 2001) was retained where conceptually indispensable. The approach identifies influential, credible, recent literature capable of supporting a coherent synthesis rather than exhaustive coverage, and is weighted toward English-language, US/European scholarship, a limitation returned to below.

Conceptualising Political Polarization

Public debate about platforms often conflates distinct phenomena. Ideological polarization refers to divergence in citizens' policy preferences toward the extremes of a political spectrum; evidence on mass-level ideological polarization, as distinct from elite polarization, is considerably more mixed than commonly assumed. Affective polarization concerns not what citizens believe but how they feel toward opposing partisans, rooted in partisanship as a social identity rather than a summary of policy commitments (Iyengar et al., 2019); much platform-related concern actually

targets tone and contempt rather than belief extremity. Perceived polarisation is the sense that society is more polarised than the survey results suggest and is a real problem for compromise and institutional trust, whether or not the sentiment is accurate. 'Political sorting' refers to the process by which partisan identity becomes correlated with other social identities (religion, race, and geography) and is largely offline, with the process being brought to light but not created by platforms. Online or networked polarisation is the tendency of people to form ideologically similar groups of interactions, as judged by network structure, not by individual beliefs, and is thus most easily captured in platform data. Distinctions are important because what is found at one level (e.g., network segregation) is often described as if it were found at another (e.g., belief extremity or the increase in hostility), but it is actually a different empirical question.

Algorithms and Platform Curation

Recommendation algorithms are often cited as a key factor driving polarisation, based on the premise that optimising for engagement will lead to the more extreme and emotionally provocative content being recommended, which is supported in part by evidence that content that is morally and emotionally charged tends to be shared and engaged with more (Vosoughi et al., 2018). Whether algorithms can be said to "cultivate" polarisation or to reflect existing polarisation is a more difficult issue to resolve than popular discussions make it appear. However, analysis of more than 10 million Facebook users revealed that ideologically cross-cutting exposure was modestly lower due to algorithmic ranking than to what users' networks would provide, but the click of individual users resulted in significantly lower levels of ideologically cross-cutting exposure than did algorithmic ranking (Bakshy et al., 2015). Moreover, a study of 208 million US Facebook users that was conducted in 2020 found segregation to be greater at the stage of user engagement than at the stage of algorithmic curation, and interestingly, the asymmetry is that conservative users see more segregation than liberal users (González-Bailón et al., 2023). A similar study identified like-minded sources in

feeds but did not identify a clear relationship with polarisation outcomes across their experimental test (Nyhan et al., 2023).

The results are not politically neutral algorithms, but rather they show that there is a curation effect and interaction with prior homophilous network structure, one that is far older than any recommender system (McPherson et al., 2001). Where there is a small effect per session, feedback loops are still possible, as small nudges over several years could add up, but this is a concern about the dynamics of the systems, which has not yet been extensively studied with access to data only for a single year, as is more typical. Modest short-term deactivation effects on some measures were observed, alongside a decrease in factual political knowledge, in large-scale deactivation experiments (Allcott et al., 2020). The evidence suggests a cautious middle ground: algorithms make emotionally charged content more visible, and they combine with other factors of user agency and homophily to have measurable, but mostly small, additional effects and don't cause mass polarisation as an independent and dominant force.

Filter Bubbles, Echo Chambers, and Selective Exposure

The filter bubble thesis (Pariser, 2011) and the related concept of echo chambers (interpersonal groups that tend to interact with like-minded individuals and thus reinforce shared views) build on Sunstein's (2017) earlier warnings about online group polarisation. Public discourse is more reliable than the empirical record. Other network studies report that while there is a high degree of homophily in political elites' networks and among highly partisan users of Twitter/X on the platform, the networks of ordinary users are far from an echo chamber, as they also consist of family members, colleagues and mostly strangers of mixed or unknown political views (Barberá et al., 2015). This literature synthesis reveals that, if they do indeed exist, echo chambers tend to be engaged by a small minority rather than forming the majority of users' experience (Cinelli et al., 2021). The pattern is consistent throughout the field and would be more likely to appeal to specific at-risk audiences than to be cast as generalities, which

could lead to concern about filter bubbles. The other surprising result is that breaking the echo chamber actually makes people feel more hostile, especially against their own partisans, at least in a few seemingly counterintuitive field experiments in which partisans were exposed to content contradictory to their own views, especially from an elite, visible source. It contradicts the simplistic policy premise that more exposure across the board leads to a healthier democracy, implying that it is not only that more exposure to a good is better for a good democracy but also that how and to whom it's exposed are also important.

Misinformation, Disinformation, and Political Division

Whether through "misinformation", "disinformation", "coordinated inauthentic activity through automated accounts", or, more recently, "synthetic media", generated by artificial intelligence, there is concern about the spread of false information – and there is no one to blame but the sender. A more influential account of diffusion, which analysed about 126,000 confirmed news cascades on Twitter (2006-2017), revealed false political news propagates more widely and quickly than credible news, largely because it is more novel and emotional than factual news (Vosoughi et al., 2018). However, research from the fields of survey and web traffic suggests that ordinary users' information diet is not all that much composed of visits to unreliable websites, as the vast bulk of users are exposed to these websites only by a small margin, and certain claims about high rates of false-news exposure have proved to be exaggerated. But survey and web-traffic research suggests that the mass of users' information diet is not particularly full of visits to unreliable websites, for a number of reasons: first, the rate of such visits is relatively low on the average user's diet; second, the websites visited are heavily pro-attitudinal; and third, claims about widespread false-news exposure have been exaggerated. Fake news exposure and sharing were found to be focused in a small subset of users, who were older and more conservative (Guess, Nagler, & Tucker, 2019; Grinberg et al., 2019).

Together, these strands paint a picture of false content which, when viewed as a whole, moves

more than is usual in terms of distance and, when shared, seems to travel farther than is usual in terms of volume, taking advantage of emotional engagement mechanisms that are common across most platforms, but the number of people who are consuming and sharing high volumes of such content is smaller and more concentrated than the term "misinformation epidemic" suggests, and the intensity of the problem should not diminish for the minority affected or for the institutions they attack. How to explain susceptibility to misinformation is better accounted for by the general lack of careful reasoning (Pennycook & Rand, 2021) than by a directional partisan motivation, as individuals are more often more, rather than less, accurate when assessing politically congenial true headlines than when assessing less politically congenial ones. A second layer is coordinated inauthentic behaviour, as found in the cross-national studies of automated and paid commentary operations in dozens of countries with different types of political regimes—with different forms of this behaviour in different countries depending on the nature of political competition (Woolley & Howard, 2019).

The Political Economy of Platforms and Affective Polarization

In the study of algorithms, much of the work assumes platform behaviour is a neutral technical artefact, whereas it is more like an outcome of an attention-based business model that prioritises attention and behavioural data extraction, driving structural incentives to maximise attention rather than civic value (Zuboff, 2019). This framing should be used judiciously: while there is reasonably solid evidence that engagement-optimised systems increase emotionality (Vosoughi et al., 2018), there is much less evidence that platforms actively set out to polarise as a strategy, as opposed to platforms being polarising merely as a side effect of engagement optimisation coupled with negativity bias and outgroup-threat sensitivity that predate the advent of digital media. The strongest, most consistent evidence, platform-linked, is related to affect and identity, not belief content. Partisan identity is like other salient social identities, as it affects how even seemingly neutral information is processed (Van Bavel &

Pereira, 2018), which helps to explain the difficulty in simply informing people about factual correction. Platforms plausibly exacerbate identity dynamics by making the approbation metrics of their platforms visible, and because the platforms are associated with combative and morally charged partisanship, they exacerbate the reputational toll of appearing to compromise more visible; Settle (2018) discusses this argument in the American context. To overestimate what processes add when they have deep offline roots: affective polarisation was increasing in the US prior to mass use of social media, and cross-national comparison reveals that some countries with high penetration rates are seeing relatively small gains in polarisation, while others with low penetration rates are seeing larger gains, consistent with country-specific political and media-system forces rather than a uniform technology effect (Boxell et al., 2024).

AI and Emerging Technologies

While concerns about misinformation are not novel to generative AI, they are different, and the literature should distinguish between known and unknown misinformation risks. Well established: Synthesised audio, video, and image production now enable realistic depictions of real public figures with minimal effort and little expense, and legal commentary has emerged outlining concerns with privacy, elections and national security (Chesney & Citron, 2019). Platform features, like AI-driven recommender systems and micro-targeting at the level of the political party, are also well documented. While public concerns about deepfakes are widespread, there is still limited population-level evidence that they have an impact on elections, and their effects may be even more damaging when people are aware that they exist—in addition to the impact on trust in authentic evidence, which has been called the "liar's dividend" (Chesney & Citron, 2019). Yet, claims of generative AI's dramatic impact on mass persuasion or qualitatively new kinds of polarisation beyond those already described in the literature require a level of caution and should not be made with the same level of evidence as more established claims.

Digital Platforms and Democratic Resilience

Democratic resilience is the ability of a political system to withstand polarisation, misinformation and distrust in the democratic institutions without disruption of fundamental democratic practices like peaceful transfers of power and contested elections. There are multiple layers of resilience at work.

Institutional, Platform, Individual, and Societal Levels

Structural buffers, such as independent electoral administration, judiciaries, and a professional press, operate as buffers—polarisation is most likely to have harmful consequences in periods of weak or captured institutions, but not so when institutions are strong (McCoy et al., 2018). This implies the same platform dynamics documented largely in the US literature may carry substantially different democratic risk depending on institutional context. Platform-level interventions, content moderation, transparency and data-access programmes, and "accuracy nudge" prompts, show mixed but not discouraging evidence: brief prompts to consider accuracy before sharing produce modest, measurable reductions in false-content sharing with the practical advantage of being largely content-agnostic (Pennycook & Rand, 2021). Individual-level media literacy interventions are comparatively encouraging: a pre-registered field experiment across the United States and India found a brief digital literacy intervention measurably improved discernment between credible and false news for several weeks, with broadly similar effects across two very different national contexts (Guess et al., 2020). Societal-level mechanisms, fact-checking, professional journalism, and cross-partisan dialogue, broaden the information ecosystem but face funding pressure from the same advertising-revenue shift that helped drive engagement-optimised design, and offline cross-partisan contact appears more robust against affective polarization than online cross-cutting exposure alone. Across all four levels, no single intervention suffices, effectiveness is highly conditional on context, and democratic resilience is best conceptualised as a layered, redundant system rather than a single lever.

Policy and Regulatory Responses

Policy responses fall into a few categories with their own set of trade-offs between harm reduction and the protection of democratic freedoms. Transparency and requirements for data access are meant to ensure the algorithm and moderation practices are auditable, as the majority of empirical debates in this literature continue to be as a result of a lack of reliable platform data access, as demonstrated by the collaborative arrangements of the most rigorous recent studies (González-Bailón et al., 2023; Nyhan et al., 2023). The European Union's Digital Services Act is the most extensive proposal to give such access and impose systemic risk-assessment duties on very large platforms (European Commission, 2022). While the problem of political advertising disclosure is smaller and more manageable than the problem of platform algorithms typically, the need for AI governance frameworks is nascent for synthetic-media disclosure. The whole book has been organised around this balancing act: the need to minimise manipulation versus the need to prevent the loss of expressions, privacy and participation — excessive moderation can result in the elimination of valid speech and could centralise discretionary power in private businesses, while minimal moderation can leave unchecked amplification systems. In democracies, this balance is made differently, owing to different legal traditions; and a few non-European or non-American countries have opted for alternative methods that are tailored to their context and concerns about the misuse of the powers of "anti-disinformation".

Integrated Synthesis and Conceptual Framework

A tentative outline may be drawn, representing the tendencies that interact rather than a definite order. These incentives for design, the algorithmic curation that they enable, the existing homophilous network structure, and selective users' behaviour combine to amplify the exposure of emotionally charged content, which in turn interacts with partisan identity to increase affective engagement, which can under certain circumstances lead to growing affective polarisation, diminished trust in institutions and increased resistance to misinformation; all of these

processes are mediated by the simultaneous action of institutional, platform, individual, and societal mechanisms of resilience (see Figure 1). This is not a linear relationship: higher affective polarisation can lead to a greater need for content that affirms identity and thus greater demand for platforms, and existing polarisation can influence the platforms that citizens engage with, as well as how. Factors that regularly moderate the relationship are the strength of the national institutions and media systems; existing polarised society; disposition towards free thinking and computer literacy; platform and content format; and demographic factors like age. The most defensible overarching argument is that platforms are amplifying and redirecting infrastructure: they do not create division out of whole cloth, but they engage with psychological tendencies and social structures that can shape their ability to measurably increase affective and perceived polarisation in predictable ways, while comparatively affecting mass ideological extremity less than public discourse claims.

Research Gaps and Future Directions

One of the biggest weaknesses of the literature is the geographic focus on the United States and Western Europe, where messaging-app-centred communication is uncommon and the political party system is different than in many other parts of the world, such as Latin America, South Asia and Africa, where these dynamics may be more complex than current feed-based theory can explain and where the bulk of existing work on misinformation and media trust is based on Africa (Wasserman & Madrid-Morales, 2019) but remains a single starting point. Causal identification remains difficult, since most observational studies cannot fully rule out reverse causation or confounding, and the rare large field experiments that achieve strong identification (Allcott et al., 2020) may not generalise to non-participating populations. Proprietary data access remains a structural constraint, leaving the field's evidentiary base contingent on voluntary corporate cooperation. Rapid platform change limits generalisability of findings based on mid-2010s architectures to newer, video-centred, AI-recommendation-driven environments.

Definitional inconsistency across polarization's dimensions continues to hamper cross-study comparison, long-term panel research remains rare relative to cross-sectional work, and research on protective and resilience factors remains considerably less developed than research on risks and harms, an asymmetry future work should correct.

Strengths and Limitations of the Review

As a narrative rather than systematic review, this paper does not follow a fully reproducible, protocol-driven search process and does not claim to capture every relevant study; selection and synthesis necessarily involved interpretive judgement. These limitations are offset by the narrative format's capacity to integrate literatures, political psychology, computational social science, platform governance law, and communication studies, that rarely speak to one another directly despite addressing closely related questions, a cross-disciplinary connection a narrowly systematic review would likely miss.

Conclusion

The evidence reviewed does not support the claim that social media and its algorithms straightforwardly cause political polarization, nor the dismissive counter-claim that platform effects are negligible. There is certain proof that platforms have measurably changed the infrastructure of political communication, affecting how information is found and disseminated, how partisan identities are

performed and how emotionally charged information is communicated, including an increased, but clearly visible, presence of misinformation. Such impacts complement and amplify rather than supplant the influence of more fundamental social, economic and institutional factors that have historically influenced political cleavages. Platform effects, on the one hand, are context-specific and affective rather than ideological, and, on the other, misinformation's threat is not as much in the quantity as in the ability to create a lack of shared factual ground for a consequential minority. To build democratic resilience, it will take coordinated action at levels: platforms addressing civic externalities and research transparency; regulators who can take the difficult decisions about the legitimate balance between preventing manipulation and ensuring users remain free to express themselves; independent institutions that can fill the void of authoritative reference points; researchers who can create a more global, more longitudinal evidence base; and citizens, journalists and educators who can cultivate the reflective habits of assessment that evidence suggests are more protective than exposure to opposing views. Digital platforms are not to blame for political polarisation, but they have new power to accelerate it, make it visible and, in certain segments of the population, make it more intense. And that is likely to continue to be a very important and complex challenge for democratic governance, to be managed responsibly.

Tables and Figure

Table 1. Key Concepts and Definitions

Concept	Definition	Relevance to Digital Platforms
Ideological Polarization	Divergence in citizens' policy preferences toward the extremes of a political spectrum.	Least consistently linked to platform use; mass ideological extremity has shifted less than affective hostility.
Affective Polarization	Emotional hostility or distrust toward opposing partisans, rooted in partisan social identity.	Most consistently associated with platform dynamics, including identity-expressive sharing.

Concept	Definition	Relevance to Digital Platforms
Perceived Polarization	The belief that society is more divided than survey data indicate.	Plausibly inflated by algorithmic surfacing of the most extreme, most engaging expressions.
Political Sorting	Alignment of partisan identity with other social, religious, or geographic identities.	A structural, largely offline process; platforms may increase visibility of sorted communities.
Online / Networked Polarization	Clustering into ideologically homogeneous communities, measured through network structure.	Directly observable in platform data; strongest among highly engaged minorities, not the median user.

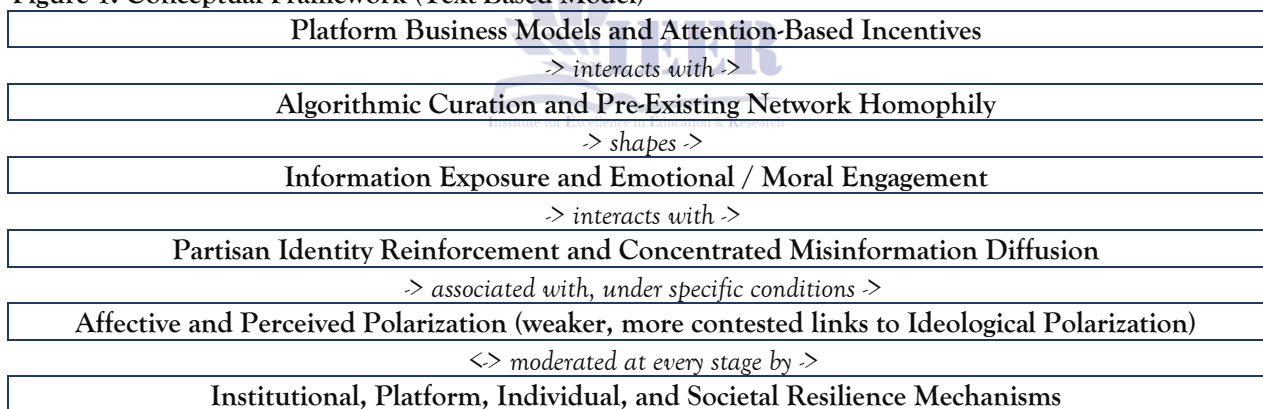
Table 2. Major Mechanisms Linking Digital Platforms and Political Polarization

Mechanism	Potential Pathway	Representative Evidence	Key Limitations
Algorithmic curation / engagement ranking	Ranking systems favour emotionally charged content over neutral content.	Bakshy et al. (2015); González-Bailón et al. (2023)	Effects modest relative to user self-selection; hard to isolate from network homophily.
Filter bubbles / echo chambers	Personalised feeds and homophilous networks reduce cross-cutting exposure.	Barberá et al. (2015); Cinelli et al. (2021)	Concentrated among a highly engaged minority; median user exposure often more mixed.
Misinformation diffusion	False content spreads faster, exploiting novelty and emotion.	Vosoughi et al. (2018); Grinberg et al. (2019)	Consumption concentrated in a narrow subgroup, not evenly distributed.
Affective / identity-based engagement	Visible social metrics reward combative, identity-affirming expression.	Iyengar et al. (2019); Van Bavel & Pereira (2018)	Hard to separate platform-specific effects from broader identity sorting.
Attention-economy business incentives	Advertising revenue models create engagement-maximising design incentives.	Zuboff (2019); Allcott et al. (2020)	Evidence for deliberate strategy weaker than evidence for byproduct.

Table 3. Democratic Resilience Strategies

Level	Strategy	Potential Benefit	Limitation
Institutional	Independent electoral administration, judiciary, and press	Absorbs partisan conflict without institutional collapse	Depends on pre-existing institutional strength, which varies widely
Platform	Transparency, researcher data access, accuracy-nudge interventions	Modest, measurable reductions in misinformation sharing	Voluntary and reversible; enforcement varies by company and jurisdiction
Individual	Media and digital literacy, prebunking / inoculation	Durable improvements in discerning credible from false content (Guess et al., 2020)	Scalability beyond controlled studies remains less established
Societal	Fact-checking, journalism, cross-partisan dialogue and contact	Broadens information ecosystem; offline contact shows stronger depolarizing effects	Funding and reach often smaller than the dynamics they aim to counter

Figure 1. Conceptual Framework (Text-Based Model)



Note. Arrows indicate plausible, empirically supported directions of influence rather than proven or exclusively linear causal pathways. Feedback loops operate throughout and are not fully captured by this single top-to-bottom sequence.

REFERENCES

Allcott, H., Braghieri, L., Eichmeyer, S., & Gentzkow, M. (2020). The welfare effects of social media. *American Economic Review*, 110(3), 629-676. <https://doi.org/10.1257/aer.20190658>

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>

- Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A., & Bonneau, R. (2015). Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological Science*, 26(10), 1531-1542. <https://doi.org/10.1177/0956797615594620>
- Boxell, L., Gentzkow, M., & Shapiro, J. M. (2024). Cross-country trends in affective polarization. *Review of Economics and Statistics*, 106(2), 557-565. https://doi.org/10.1162/rest_a_01160
- Chesney, R., & Citron, D. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *California Law Review*, 107(6), 1753-1820. <https://doi.org/10.15779/Z38RV0D15J>
- Cinelli, M., De Francisci Morales, G., Galeazzi, A., Quattrociocchi, W., & Starnini, M. (2021). The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, 118(9), e2023301118. <https://doi.org/10.1073/pnas.2023301118>
- European Commission. (2022). *The Digital Services Act: Ensuring a safe and accountable online environment*. European Commission.
- González-Bailón, S., Lazer, D., Barberá, P., Zhang, M., Allcott, H., Brown, T., Crespo-Tenorio, A., Freelon, D., Gentzkow, M., Guess, A. M., Iyengar, S., Kim, Y. M., Malhotra, N., Moehler, D., Nyhan, B., Pan, J., Rivera, C. V., Settle, J., Thorson, E., Tromble, R., Wilkins, A., Wojcieszak, M., Kiewiet de Jonge, C., Franco, A., Mason, W., Stroud, N. J., & Tucker, J. A. (2023). Asymmetric ideological segregation in exposure to political news on Facebook. *Science*, 381(6656), 392-398. <https://doi.org/10.1126/science.ade7138>
- Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., & Lazer, D. (2019). Fake news on Twitter during the 2016 U.S. presidential election. *Science*, 363(6425), 374-378. <https://doi.org/10.1126/science.aau2706>
- Guess, A., Lerner, M., Lyons, B., Montgomery, J. M., Nyhan, B., Reifler, J., & Sircar, N. (2020). A digital media literacy intervention increases discernment between mainstream and false news in the United States and India. *Proceedings of the National Academy of Sciences*, 117(27), 15536-15545. <https://doi.org/10.1073/pnas.1920498117>
- Guess, A., Nagler, J., & Tucker, J. (2019). Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science Advances*, 5(1), eaau4586. <https://doi.org/10.1126/sciadv.aau4586>
- Guess, A. M., Nyhan, B., & Reifler, J. (2020). Exposure to untrustworthy websites in the 2016 US election. *Nature Human Behaviour*, 4(5), 472-480. <https://doi.org/10.1038/s41562-020-0833-x>
- Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., & Westwood, S. J. (2019). The origins and consequences of affective polarization in the United States. *Annual Review of Political Science*, 22, 129-146. <https://doi.org/10.1146/annurev-polisci-051117-073034>
- Kubin, E., & von Sikorski, C. (2021). The role of (social) media in political polarization: A systematic review. *Annals of the International Communication Association*, 45(3), 188-206. <https://doi.org/10.1080/23808985.2021.1976070>

- McCoy, J., Rahman, T., & Somer, M. (2018). Polarization and the global crisis of democracy: Common patterns, dynamics, and pernicious consequences for democratic polities. *American Behavioral Scientist*, 62(1), 16-42. <https://doi.org/10.1177/0002764218759576>
- McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27, 415-444. <https://doi.org/10.1146/annurev.soc.27.1.415>
- Nyhan, B., Settle, J., Thorson, E., Wojcieszak, M., Barberá, P., Chen, A. Y., Allcott, H., Brown, T., Crespo-Tenorio, A., Freelon, D., Gentzkow, M., González-Bailón, S., Guess, A. M., Lazer, D., Kim, Y. M., Malhotra, N., Moehler, D., Pan, J., Rivera, C. V., Tromble, R., Wilkins, A., Kiewiet de Jonge, C., Franco, A., Mason, W., Stroud, N. J., & Tucker, J. A. (2023). Like-minded sources on Facebook are prevalent but not polarizing. *Nature*, 620(7972), 137-144. <https://doi.org/10.1038/s41586-023-06297-w>
- Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. Penguin Press.
- Pennycook, G., & Rand, D. G. (2021). The psychology of fake news. *Trends in Cognitive Sciences*, 25(5), 388-402. <https://doi.org/10.1016/j.tics.2021.02.007>
- Settle, J. E. (2018). *Frenemies: How social media polarizes America*. Cambridge University Press.
- Sunstein, C. R. (2017). *#Republic: Divided democracy in the age of social media*. Princeton University Press.
- Van Bavel, J. J., & Pereira, A. (2018). The partisan brain: An identity-based model of political belief. *Trends in Cognitive Sciences*, 22(3), 213-224. <https://doi.org/10.1016/j.tics.2018.01.004>
- 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>
- Wasserman, H., & Madrid-Morales, D. (2019). An exploratory study of "fake news" and media trust in Kenya, Nigeria and South Africa. *African Journalism Studies*, 40(1), 107-123. <https://doi.org/10.1080/23743670.2019.1627230>
- Woolley, S. C., & Howard, P. N. (Eds.). (2019). *Computational propaganda: Political parties, politicians, and political manipulation on social media*. Oxford University Press.
- Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. PublicAffairs.