

# FRAMING THE SYNTHETIC: A REVIEW OF HOW AI-GENERATED AND FAKE-NEWS NARRATIVES ARE CONSTRUCTED AND CONTESTED ON SOCIAL MEDIA

Umar Nawaz<sup>\*1</sup>, Areeba Tariq<sup>2</sup>, Pawan Masud<sup>3</sup>

<sup>\*1</sup>MPhil in Journalism and Media Studies

<sup>2,3</sup>In-Charge Department of Media and Communication Studies

DOI: <https://doi.org/10.5281/zenodo.21351829>

## Keywords

framing; generative AI; deepfakes; misinformation; fake news; media literacy; social media; spiral of silence; Pakistan Framing the Synthetic

## Article History

Received: 17 April 2026

Accepted: 07 June 2026

Published: 21 June 2026

Copyright @Author

Corresponding Author: \*

Umar Nawaz

## Abstract

Generative artificial intelligence has revived old anxieties about deception and manipulation online, and most existing scholarship still approaches synthetic and fake-news content mainly as a detection problem – asking whether audiences can be fooled and how false material might be screened out. This paper takes a different, complementary route by applying the lens of framing. Building on Entman's (1993) account of framing as the selective emphasis given to certain problem definitions, causal stories, moral judgements, and proposed remedies, it kind of draws together existing research on how AI-generated and fake-news narratives are built spread interpreted and pushed back against on social media in that messy, ongoing way. The scattered literature is brought into one conceptual model spanning five linked stages – construction, algorithmic amplification, audience processing, resulting outcomes, and counter-framing – shaped throughout by the prevailing opinion climate (spiral of silence) and by levels of media and digital literacy. A central debate running through the review is whether the dangers of generative-AI misinformation have been overstated: supply-side accounts that stress the threat of fabricated content are weighed against evidence suggesting that demand and distribution, not production capacity, are the real bottleneck. The review also considers how synthetic media may harm public discourse less by convincing people of particular falsehoods than by spreading generalised doubt and weakening trust in news, and it traces how counter-framing efforts – fact-checking, debunking, and inoculation – meet with mixed success against manipulative narratives. Three gaps emerge from this synthesis: a heavy skew toward Global-North evidence, a near-total absence of frame-analytic work on synthetic content specifically, and very little indexed research from South Asia, Pakistan included. The paper closes by sketching a framing-centred, computationally informed, and comparative research agenda suited to communication scholarship that takes platform dynamics and the Global South seriously.

## 1. INTRODUCTION

Worry about false and manipulated information is hardly new, but generative AI has given it a sharper edge. Large language models can churn out fluent political text on demand, while image and video

generators can produce convincing depictions of events that simply did not happen (Goldstein et al., 2023; Kreps et al., 2022). Since platforms circulate this material quickly and unevenly, the way citizens meet or run into political information

has shifted, in ways communication scholars are still trying to map, properly. Much of the public conversation reduces this to a question of authenticity and detection – can people spot what is fake, and can platforms remove it?

This review starts from a related but different premise. Regardless of whether a post is factually accurate, it is always framed in some way: certain aspects of reality are foregrounded, made to feel important, and steered toward a particular reading (Entman, 1993). Seen in this light, synthetic and fake-news content is not just a bundle of claims to be fact-checked but an act of framing to be unpacked, every piece names a problem, points to a cause, carries a moral charge, and then quietly implies some kind of response. Approaching it this way helps explain how it works, why correcting it is so often unsuccessful, and where intervention might actually land and make a difference. In the end the paper sets out to pull together a literature that feels fragmented yet expanding rapidly, organizing it into one coherent model of how AI-generated and fake-news narratives take shape, then get contested, on social media.

Four objectives structure the review. The first is conceptual: clarifying terms such as fake news, information disorder, and synthetic media, and making the case for reading them through framing. The second traces how such narratives are put together and amplified, and how audiences then process them. The third looks at how they are challenged – via fact-checking, debunking, and inoculation – and at the role media literacy plays in shaping outcomes. The fourth pulls out the major debates and gaps in the field and sketches an agenda for future work. A persistent imbalance runs through all of this: most of the evidence comes from the Global North, while regions such as South Asia – and Pakistan especially – remain comparatively understudied despite being heavily exposed to online misinformation (Ali & Qazi, 2023; Javed et al., 2022).

## 2. Scope and Approach

This review is narrative and integrative rather than systematic. Its goal is conceptual rather than exhaustive: to draw together strands of research that have largely developed on their own tracks –

framing, generative AI and deepfakes, fake-news diffusion, correction and inoculation, media literacy, and platform manipulation – and read them through one theoretical lens. Sources were located through searches of major databases and communication journals, with priority given to peer-reviewed work from 2018 to 2025, alongside earlier foundational studies that still carry authority in the field. Reports and preprints from established research centres are used where they are well cited and clearly identified as such. Particular care was taken to surface the limited indexed scholarship coming out of South Asia – partly to draw on it, and partly to make its scarcity visible as a gap in its own right.

## 3. Clarifying the Terrain: Fake News, Information Disorder, and Synthetic Media

The phrase “fake news” is itself contested ground, heavily loaded with political meaning. In a review of scholarly definitions, Tandoc, Lim, and Ling (2018) sorted the concept into six recurring forms – satire, parody, outright fabrication, manipulated content, advertising, and propaganda – separated along two axes: how factual the content is, and whether the author intends to deceive. Their typology makes clear that “fake news” as commonly used lumps together quite different communicative purposes, not all of which involve deliberate deception. Partly for this reason, many researchers now favour the broader idea of information disorder suggested by Wardle and Derakhshan (2017), sort of splitting misinformation (false content spread without intent to cause harm) from disinformation (false content pushed deliberately to cause harm) and mal-information (true content weaponised to cause harm). They then look at each stream in terms of who makes it, what the content is, and who ends up receiving it, which sounds a bit obvious but still matters.

In that same conceptual space, Lazer et al. (2018) framed fake news as fabricated material that takes on the surface feel of journalism yet misses the editorial process or the underlying intent, and they asked for a coordinated response, not just from researchers but from platforms too. Vosoughi, Roy, and Aral (2018) then added empirical weight

to these worries: by tracking around 126,000 stories moving on Twitter, they observed that false ones went further, faster, and deeper than true stories in almost every category, and false political stories were the quickest to spread of all. What's more, they traced this so-called advantage to human behaviour rather than bot activity, arguing that the novelty and emotional pull of false content is what drives it in practice. If you read it through a framing lens, the result is telling: falsehoods don't always travel because they're just wrong; they often get traction because they're cast, or framed, in ways that are more striking and emotionally loaded than the truth.

Generative AI does not really invent a whole new situation so much as tack on an extra production layer to an older pattern. AI made, or AI tweaked, text images audio and video – what this review calls synthetic media – makes fabricated content cheaper to generate and also easier to pass off as convincing, at least visually (Goldstein et al., 2023). Yet, as the sections that follow argue, what makes synthetic media communicatively significant is less its technical novelty than how it gets framed, how it travels, and how audiences and institutions respond to it.

#### 4. Theoretical Framework

Framing theory anchors this review, supported by two secondary lenses – the elaboration likelihood model and the spiral of silence – that together help explain how synthetic and false narratives are constructed, how they manage to persuade, and how they both shape and are shaped by the surrounding climate of opinion.

**Framing.** Entman (1993) described the field of framing research as a “fractured paradigm,” yet proposed a definition that has held up well: to frame something is to pick out particular aspects of reality and give them prominence, in a way that pushes a given problem definition, causal account, moral judgement, and recommended fix. De Vreese (2005) built on this with a process model that separates frame-building (how frames originate), frame-setting (how frames meet audience predispositions), and framing's downstream consequences, while also distinguishing generic frames that recur across

topics from frames tied to one specific issue. Applying this lens here, AI-generated and fake-news content can be treated as framing acts in their own right – examined for the problems they name, the parties they blame, the values they appeal to, and the responses they push toward.

**Elaboration likelihood.** Petty and Cacioppo's (1986) elaboration likelihood model separates a central route to persuasion – driven by careful evaluation of a message's content – from a peripheral route, driven instead by surface cues like source credibility or apparent realism. This distinction proves useful for synthetic media specifically: a convincing deepfake can persuade through the peripheral route, on the strength of how authentic it looks, even when its underlying claims would collapse under closer inspection. Audiences with little motivation or capacity to scrutinise what they see are especially vulnerable to this kind of peripheral persuasion.

**Spiral of silence.** Noelle-Neumann (1974) proposed that people who believe their opinions sit in the minority tend to stay quiet about them out of fear of social isolation, setting off a spiral in which minority views grow steadily less visible. This held up empirically: a meta-analysis by Matthes, Knoll, and von Sikorski (2018), drawing on 66 studies and over 27,000 participants, found a modest but consistent link between the perceived climate of opinion and people's willingness to speak up – a link that held just as strongly online as offline. Stoycheff (2016) added an experimental dimension, showing that simply being aware of online surveillance further discourages people from voicing minority political views. This theoretical strand connects the climates created by synthetic and false narratives – along with any accompanying surveillance or censorship – to citizens' willingness to be heard at all.

These three lenses kind of come together in the model shown in Figure 1, where Synthetic and false narratives start out as framing acts. They're often stitched up with the help of AI generation tools, then they get pushed along via algorithmic diffusion and some coordinated manipulation, and audiences end up processing them through central or peripheral routes. After that, the

outcomes can range from outright belief to more uncertainty, eroded trust, and even shifts in what people will actually say. Two moderators shape this entire sequence — the opinion climate and associated self-censorship dynamics on one side,

media and digital literacy on the other — while a feedback loop of contestation (fact-checking, debunking, inoculation) functions as counter-framing that pushes back against the original narratives.

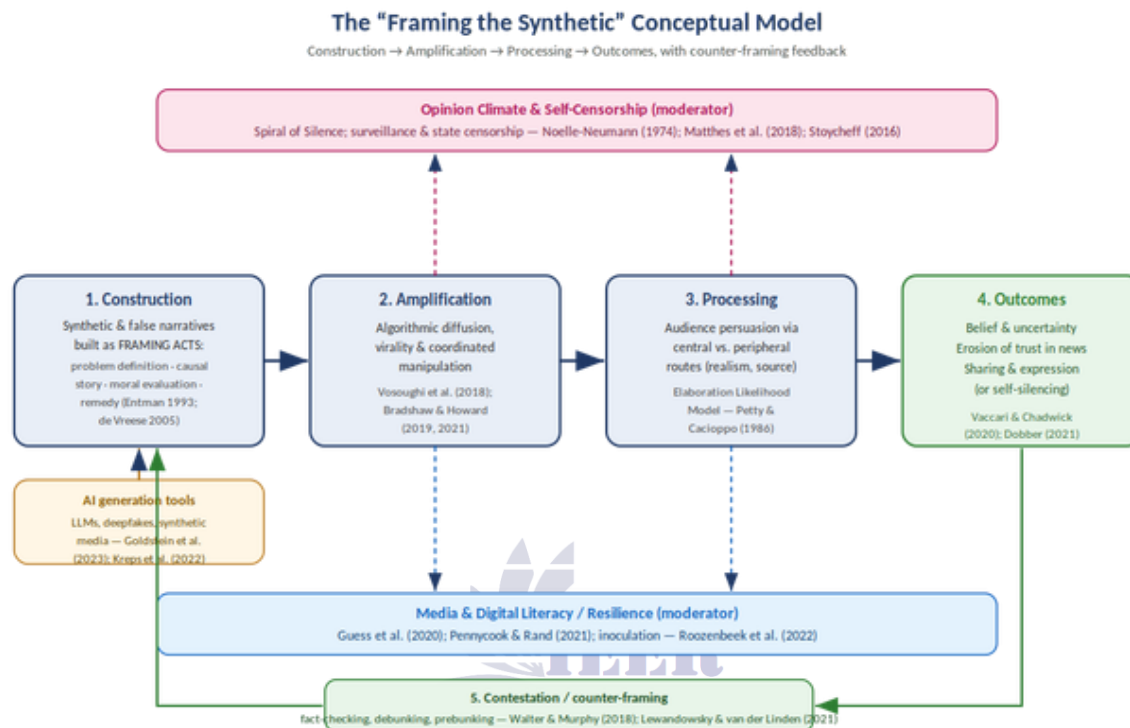


Figure 1. The “Framing the Synthetic” conceptual model.

## 5. Constructing Synthetic and False Narratives

What goes into building an AI-generated or fake-news narrative, and what makes it land? Three findings are worth pointing out. First, machine-written text is genuinely hard to separate, like it’s difficult to tell apart from human writing. In multiple experiments, Kreps, McCain, and Brundage (2022) observed that participants usually could not distinguish AI-generated news stories from human-written ones. They also found partisan leanings predicted which stories people regarded as credible. And importantly, more advanced models only improve perceived credibility a bit, not dramatically. The takeaway is that synthetic text persuades less because it’s technically polished, and more because it’s framed well enough to fit what readers already believe. Second, synthetic video is capable of shifting attitudes outright. Dobber, Metoui, Trilling,

Helberger, and de Vreese (2021) built a political deepfake and found that exposure to it measurably worsened viewers’ opinions of the politician depicted, with the effect growing stronger when the video was microtargeted at receptive audiences. Framing and targeting reinforce each other here — a frame that already fits an audience’s leanings, presented in a realistic synthetic format, has real power to move opinion.

Third — and this is arguably the most theoretically important finding — synthetic media may damage public discourse less by getting people to believe specific lies than by sowing generalised doubt. In a large UK-based experiment, Vaccari and Chadwick (2020) found that political deepfakes left viewers uncertain far more often than they left them genuinely deceived, and that this uncertainty itself eroded trust in news shared on social media. This suggests a particular kind of framing effect:

merely knowing that any photo or video could be fabricated recasts the whole information environment as suspect – a dynamic often labelled the liar’s dividend, where even genuine evidence can be waved away as potentially fake.

Taken as a whole, this body of work kind of supports treating synthetic content as an act of framing rather than just a question of truth, or falsehood. Whether a given piece of content is real, matters less than what issue it points to, who it assigns blame to, and how its design, plus targeting make one specific reading feel convincing, in practice.

#### **6. Amplification: Algorithms, Virality, and Coordinated Manipulation**

How a narrative is built matters less than how far it travels. Because of the diffusion advantage documented by Vosoughi et al. (2018), emotionally charged falsehoods can simply outrun any correction aimed at them. Beyond organic spread, there are actors who deliberately produce and push manipulative content. The Oxford Internet Institute’s successive inventories of organised social-media manipulation tracked such operations expanding from a small handful of countries to dozens within a matter of years, documenting state and party actors deploying human, automated, and hybrid accounts to steer public opinion (Bradshaw & Howard, 2019). A subsequent inventory found this had grown into an increasingly professionalised – even industrial-scale – disinformation industry active across dozens of countries (Bradshaw et al., 2021).

What stands out for debates about synthetic media specifically is that these same inventories found deepfakes were, at the time of reporting, still rarely used in actual political operations compared with cruder tactics – doctored images, decontextualised “cheapfakes,” and straightforward coordinated amplification. That finding usefully corrects any assumption that sophisticated synthetic media already dominates the manipulation landscape. When you see things through a framing lens, amplification is really the stage where a frame either starts getting traction, or it quietly fades away ; and the platform design alongside the coordinated labour behind these networks,

together decide which version of reality ends up visible to mass audiences.

#### **7. Audience Processing and Outcomes**

Why do some people fall for manipulative frames while others see through them? In their review of the psychology of fake news, Pennycook and Rand (2021) argued that susceptibility has more to do with inattention and a failure to think analytically than with partisan-motivated reasoning, and that there is frequently a wide gap between what people privately believe and what they are willing to share publicly. This lines up neatly with the elaboration likelihood model: when motivation or ability to scrutinise content is low, peripheral cues – the apparent realism of a piece of synthetic media, or simply who shared it – end up carrying outsized weight.

The outcomes that follow from exposure vary widely. They range from belief in false claims to uncertainty and a general erosion of trust in legitimate news (Vaccari & Chadwick, 2020), to shifts in political attitudes (Dobber et al., 2021), to changes in how people choose to express themselves. The spiral-of-silence literature adds, an important reminder: exposure can also produce silence. When a synthetic or manipulated narrative appears to dominate the opinion climate, or when surveillance and censorship increase the perceived price of speaking up, people with minority views might simply choose not to voice them (Matthes et al., 2018 Stoycheff, 2016). Any framing analysis that overlooked these expressive outcomes would be missing a real part of how synthetic narratives reshape public discourse.

#### **8. Contesting Synthetic and False Narratives**

If manipulative content amounts to framing, then any attempt to counter it is, by definition, counter-framing. The evidence on corrections is hopeful, but with real caveats. Walter and Murphy’s (2018) meta-analysis found that corrective messages do have a moderate overall effect on belief in misinformation, but that this effect is noticeably weaker in political contexts than in health-related ones, and that corrections built around a coherent alternative explanation outperform simple factual

rebuttals. Walter and Tukachinsky (2020) documented something more stubborn still – a continued-influence effect, in which misinformation keeps shaping beliefs even after being corrected, an effect that credible sources and coherent alternative accounts can soften but not fully remove.

Partly because of these limits, researchers have increasingly shifted attention from debunking after the fact toward prebunking in advance. Drawing on inoculation theory, several studies have shown that briefly exposing people to weakened versions of manipulation techniques can build up resistance before the real thing is encountered. Traberg, Roozenbeek, and van der Linden (2022) surveyed this body of evidence, while Roozenbeek et al. (2022) ran large online experiments – including on a major video-sharing platform – showing that short inoculation videos genuinely improved viewers' ability to spot manipulation techniques. Lewandowsky and van der Linden (2021) argued that this kind of technique-based prebunking could be scaled up into a real frontline defense, sort of. In framing terms, inoculation works by having people learn to recognise how manipulative frames are put together in the first place, so the effect gets blunted before anything actually shows up in the wild.

### 9. Media Literacy as Resilience

Media and digital literacy is by far the most commonly proposed structural defence against misinformation, and the evidence does suggest it works to some degree. Guess et al. (2020) tested a brief “tips to spot false news” intervention in both the United States and India and found it improved participants' ability to tell mainstream news from false news in both settings. What matters for this review is that the effect was noticeably smaller in the Indian sample than in the American one – a clear sign that interventions designed for, and in, the Global North do not transfer cleanly into South Asian settings. This gap between contexts is itself a finding relevant to framing: literacy interventions need to be reframed and adapted to local information ecologies, languages, and patterns of trust if they are going to work.

Media literacy also feeds back into the psychological mechanisms discussed earlier. Now, if inattention really is a big driver of susceptibility (Pennycook & Rand, 2021), then literacy programmes that train people to engage analytically – not just handing out facts – are probably going to matter more in practice. And since both literacy and inoculation share the same aim of helping people see how content is composed, they fit pretty naturally next to the counter-framing strategies discussed in Section 8, more or less.

### 10. The Global South and the Pakistani Context

Also, most of what we reviewed so far comes from North America and Western Europe. Still, some of the most consequential misinformation dynamics happen in the Global South, where messaging apps are pretty central to political life, and where institutional and journalistic safeguards are often weaker. There is evidence emerging from South Asia, but it is still fairly sparse. Garimella and Eckles (2020), for instance, showed that image-based misinformation was widespread within Indian political groups on a major messaging platform, which really points to how important visual and synthetic content is in this region. Then Javed et al. (2022) did one of the earlier studies of public messaging-app groups in Pakistan, and they found that during the COVID-19 period a large share of the messages counted as misinformation, and that politically affiliated groups had an outsized role in spreading it. This happened often well before the same material turned up elsewhere.

On the defensive side, Ali and Qazi (2023) ran a randomised field experiment in urban Pakistan and found that generic literacy messaging barely moved the needle, while messages that included personalised feedback meaningfully improved people's ability to spot false news – particularly among those with lower digital literacy to begin with. This complicates the more optimistic literacy findings reported elsewhere, and it also kind of reinforces the need for context specific design. The wider media environment shapes these dynamics too: research on Pakistani journalism suggests pressures around press freedom and

safety, plus uneven adoption of AI and automation in newsrooms (Jamil, 2021, 2023) and those conditions influence how forcefully false narratives can end up being challenged, or not.

Two things follow from this. First, indexed, peer-reviewed research on misinformation in Pakistan clusters around a handful of topics – health misinformation, messaging-app diffusion, and journalism studies – and remains thin given the actual scale of the problem. Second, that scarcity is itself an opening. A framing-centred, computationally grounded study of how synthetic and false narratives are built and contested within Pakistani social media would fill a genuine gap, while still speaking directly to the international debates surveyed in this review.

### 11. Synthesis: Debates and Gaps

Reading these strands of literature together through a framing lens brings several debates and gaps into focus – ones that mark out the current frontier of the field.

**Are generative-AI misinformation effects overstated?** This is arguably the field's central tension. Supply-side accounts stress that AI dramatically cuts the cost and boosts the realism of fabricated content, expanding what influence operations are capable of (Goldstein et al., 2023). Against this, a prominent counter-argument holds that such fears are overblown, since the real bottleneck on misinformation is not the supply of content but demand for it and how it gets distributed – most people simply do not go looking for misinformation, and those who do already have plenty of non-AI sources to draw on (Simon et al., 2023). The evidence so far, including how rarely deepfakes actually turn up in documented operations (Bradshaw et al., 2021), leans somewhat toward this more cautious view – though that could shift as the tools improve and new election cycles generate fresh evidence.

**Deception versus uncertainty.** A second debate concerns exactly how synthetic media causes harm. Given that it more often breeds uncertainty and distrust than specific false beliefs (Vaccari & Chadwick, 2020), the most serious risk may turn out to be a broad erosion of trust in the information environment as a whole – something

framing research is well positioned to investigate, but has not yet examined in much depth.

**The framing gap.** The bulk of research on synthetic media remains either experimental-psychological or focused on detection. Relatively little of it actually analyses how AI-generated narratives are framed – the problems they construct, the parties they blame, and how counter-framing measures fare against them. This gap is the most direct opportunity this review identifies.

**The Global-North bias.** Most of the foundational evidence comes from a small set of wealthy democracies, and the interventions developed there don't transfer in a clean way to other settings (Ali & Qazi, 2023; Guess et al., 2020). The comparative lack of South Asian evidence, and Pakistani evidence in particular, is both a real limitation of the field, and also a fairly obvious path for future work.

### 12. A Framing-Centred Research Agenda

Pulling these threads together suggests a more coherent research agenda. Future work should treat synthetic and false narratives, pretty explicitly as framing acts, and examine how they're constructed empirically through content-analytic and computational text as data approaches, so they can identify the problem definitions, causal stories, moral judgements, and the remedies they carry (Entman, 1993; Grimmer & Stewart, 2013). This kind of analysis should be checked against human coding and, where feasible, scaled up using supervised or assisted classification of frames. It should also be comparative – testing whether patterns in how synthetic narratives are framed, and how well counter-framing and literacy interventions actually work, hold up across different media systems, languages, and understudied South Asian settings. Finally, it should also bring in the more expressive outcomes that the spiral of silence tradition really highlights, not just whether people believe synthetic narratives, but whether the climate that comes out of that makes them less willing, at all, to speak up. This kind of agenda is methodologically usable, set on solid theory, and it answers straight back to the

gaps that were already found, even if they were named kind of indirectly earlier.

### 13. Conclusion

Generative AI has sharpened an old problem far more than it has created a genuinely new one. False and manipulative content has always relied on framing – on selectively constructing reality so that one particular interpretation feels compelling – and synthetic media is best understood as a powerful new way of producing those framings, rather than as some force operating outside dynamics communication scholars already understand. Viewed through this lens, the construction, amplification, processing, and contestation of synthetic and fake-news narratives emerge as parts of one connected system, shaped throughout by the opinion climate and by media literacy. It also becomes clear where the field is still thin – in frame-analytic study of synthetic content specifically, and in evidence drawn from the Global South. A framing-centred, computationally grounded, comparative research programme, attentive to contexts like Pakistan, would advance both theoretical understanding and the public's grasp of how democracies might sustain a trustworthy, open information environment in an age of synthetic media.

### REFERENCES

- Ali, A., & Qazi, I. A. (2023). Countering misinformation on social media through educational interventions: Evidence from a randomized experiment in Pakistan. *Journal of Development Economics*, 163, 103108. <https://doi.org/10.1016/j.jdeveco.2023.103108>
- Bradshaw, S., & Howard, P. N. (2019). The global disinformation order: 2019 global inventory of organised social media manipulation (Working Paper 2019.2). Project on Computational Propaganda, Oxford Internet Institute.
- Bradshaw, S., Bailey, H., & Howard, P. N. (2021). Industrialized disinformation: 2020 global inventory of organized social media manipulation (Working Paper 2021.1). Project on Computational Propaganda, Oxford Internet Institute.
- de Vreese, C. H. (2005). News framing: Theory and typology. *Information Design Journal + Document Design*, 13(1), 51–62. <https://doi.org/10.1075/idjdd.13.1.06vre>
- Dobber, T., Metoui, N., Trilling, D., Helberger, N., & de Vreese, C. H. (2021). Do (microtargeted) deepfakes have real effects on political attitudes? *The International Journal of Press/Politics*, 26(1), 69–91. <https://doi.org/10.1177/1940161220944364>
- Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, 43(4), 51–58. <https://doi.org/10.1111/j.1460-2466.1993.tb01304.x>
- Garimella, K., & Eckles, D. (2020). Images and misinformation in political groups: Evidence from WhatsApp in India. *Harvard Kennedy School Misinformation Review*, 1(5). <https://doi.org/10.37016/mr-2020-030>
- Goldstein, J. A., Sastry, G., Musser, M., DiResta, R., Gentzel, M., & Sedova, K. (2023). Generative language models and automated influence operations: Emerging threats and potential mitigations (arXiv:2301.04246). arXiv. <https://doi.org/10.48550/arXiv.2301.04246>
- Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, 21(3), 267–297. <https://doi.org/10.1093/pan/mps028>

- Guess, A. M., 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>
- Jamil, S. (2021). Artificial intelligence and journalistic practice: The crossroads of obstacles and opportunities for the Pakistani journalists. *Journalism Practice*, 15(10), 1400–1422. <https://doi.org/10.1080/17512786.2020.1788412>
- Jamil, S. (2023). Automated journalism and the freedom of media: Understanding legal and ethical implications in competitive authoritarian regime. *Journalism Practice*, 17(6), 1115–1138. <https://doi.org/10.1080/17512786.2021.1981148>
- Javed, R. T., Usama, M., Iqbal, W., Qadir, J., Tyson, G., Castro, I., & Garimella, K. (2022). A deep dive into COVID-19-related messages on WhatsApp in Pakistan. *Social Network Analysis and Mining*, 12(1), 5. <https://doi.org/10.1007/s13278-021-00833-0>
- Kreps, S., McCain, R. M., & Brundage, M. (2022). All the news that's fit to fabricate: AI-generated text as a tool of media misinformation. *Journal of Experimental Political Science*, 9(1), 104–117. <https://doi.org/10.1017/XPS.2020.37>
- 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>
- Lewandowsky, S., & van der Linden, S. (2021). Countering misinformation and fake news through inoculation and prebunking. *European Review of Social Psychology*, 32(2), 348–384. <https://doi.org/10.1080/10463283.2021.1876983>
- Matthes, J., Knoll, J., & von Sikorski, C. (2018). The “spiral of silence” revisited: A meta-analysis on the relationship between perceptions of opinion support and political opinion expression. *Communication Research*, 45(1), 3–33. <https://doi.org/10.1177/0093650217745429>
- Noelle-Neumann, E. (1974). The spiral of silence: A theory of public opinion. *Journal of Communication*, 24(2), 43–51. <https://doi.org/10.1111/j.1460-2466.1974.tb00367.x>
- 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>
- Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. *Advances in Experimental Social Psychology*, 19, 123–205. [https://doi.org/10.1016/S0065-2601\(08\)60214-2](https://doi.org/10.1016/S0065-2601(08)60214-2)
- Roozenbeek, J., van der Linden, S., Goldberg, B., Rathje, S., & Lewandowsky, S. (2022). Psychological inoculation improves resilience against misinformation on social media. *Science Advances*, 8(34), eabo6254. <https://doi.org/10.1126/sciadv.abo6254>
- Simon, F. M., Altay, S., & Mercier, H. (2023). Misinformation reloaded? Fears about the impact of generative AI on misinformation are overblown. *Harvard Kennedy School Misinformation Review*, 4(5). <https://doi.org/10.37016/mr-2020-127>

- Stoycheff, E. (2016). Under surveillance: Examining Facebook's spiral of silence effects in the wake of NSA internet monitoring. *Journalism & Mass Communication Quarterly*, 93(2), 296–311.  
<https://doi.org/10.1177/1077699016630255>
- Tandoc, E. C., Lim, Z. W., & Ling, R. (2018). Defining “fake news”: A typology of scholarly definitions. *Digital Journalism*, 6(2), 137–153.  
<https://doi.org/10.1080/21670811.2017.1360143>
- Traberg, C. S., Roozenbeek, J., & van der Linden, S. (2022). Psychological inoculation against misinformation: Current evidence and future directions. *The ANNALS of the American Academy of Political and Social Science*, 700(1), 136–151.  
<https://doi.org/10.1177/00027162221087936>
- Vaccari, C., & Chadwick, A. (2020). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social Media + Society*, 6(1).  
<https://doi.org/10.1177/2056305120903408>
- 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>
- Walter, N., & Murphy, S. T. (2018). How to unring the bell: A meta-analytic approach to correction of misinformation. *Communication Monographs*, 85(3), 423–441.  
<https://doi.org/10.1080/03637751.2018.1467564>
- Walter, N., & Tukachinsky, R. (2020). A meta-analytic examination of the continued influence of misinformation in the face of correction: How powerful is it, why does it happen, and how to stop it? *Communication Research*, 47(2), 155–177.  
<https://doi.org/10.1177/0093650219854600>
- Wardle, C., & Derakhshan, H. (2017). Information disorder: Toward an interdisciplinary framework for research and policy making. Council of Europe. Framing the Synthetic – A. Tariq | Page 1

