

Cyberpsychology and Strategic Communication: An Integrated Cognitive Defense Framework for Countering Foreign Digital Propaganda

Mükremin T. Rahman

Researcher, Global Research & Analysis Lab (GRAL), Romania

Abstract

Foreign digital propaganda exploits the same cognitive architecture that strategic communication was designed to persuade, making the two fields inseparable partners in contemporary information defense. This article synthesizes theory and empirical evidence from strategic communication and cyberpsychology to propose an integrated, four-layer cognitive defense framework for countering state-sponsored and state-aligned disinformation online. Drawing on the epidemiological model of misinformation susceptibility, inoculation theory, and evidence from randomized controlled trials of prebunking interventions, alongside computational studies of coordinated inauthentic behavior on platforms such as Twitter/X and Telegram, the article argues that effective counter-propaganda strategy must operate simultaneously at the perceptual, cognitive, social, and institutional levels. The proposed Layered Cognitive Defense Model (LCDM) integrates psychological inoculation, source-credibility signaling, network-level detection, and coordinated strategic messaging into a single analytic architecture. The article concludes that durable resilience against foreign influence operations depends less on debunking individual falsehoods after the fact and more on building anticipatory psychological and institutional immunity before exposure occurs. Implications for policy, platform governance, and future interdisciplinary research are discussed.

Keywords: strategic communication, cyberpsychology, disinformation, foreign propaganda, inoculation theory, cognitive security, information warfare

1. Introduction

Foreign propaganda has migrated from radio transmitters and printed leaflets to algorithmically amplified social media feeds, and in doing so it has changed from a problem of message delivery into a problem of mind. Contemporary influence operations no longer merely broadcast a narrative; they exploit the architecture of human cognition, the design of recommendation systems, and the social dynamics of online communities to make falsehood feel true (Vosoughi et al., 2018). Vosoughi et al. (2018) demonstrated in a landmark analysis of roughly 126,000 cascades on Twitter that false stories were 70% more likely to be retweeted than true stories and reached 1,500 people six times faster, an asymmetry the authors attributed to the novelty and emotional intensity of falsehood rather than to bot amplification alone. This single finding reframes the counter-propaganda problem: if falsehood spreads faster because it is psychologically more

compelling, then strategic communication cannot rely on correction after the fact. It must engage the same cognitive terrain that propagandists already occupy.

Two academic traditions converge on this terrain but have historically developed in relative isolation from one another. Strategic communication, defined by Hallahan et al. (2007) as "the purposeful use of communication by an organization to fulfill its mission" (p. 3), supplies the institutional logic of goal-directed, audience-centered messaging across public relations, political communication, and public diplomacy. Cyberpsychology, an interdisciplinary field examining how digital environments shape cognition, emotion, and behavior, supplies the mechanistic account of why particular messages take hold in particular minds under particular platform conditions. The central argument of this article is that neither tradition alone is sufficient to counter foreign digital propaganda, and that a genuinely effective response requires their deliberate integration into a layered framework that treats the individual mind, the social network, and the communicating institution as a single defensible system.

This article proceeds in four stages. First, it reviews the theoretical foundations of strategic communication and situates foreign propaganda within that literature. Second, it examines the cyberpsychological mechanisms, cognitive, affective, and social, that render individuals susceptible to online manipulation. Third, it analyzes the operational anatomy of contemporary foreign influence campaigns, drawing on computational and discourse-analytic studies of coordinated inauthentic behavior. Fourth, it synthesizes this evidence into an original analytic contribution, the Layered Cognitive Defense Model, and discusses its implications for policy, platform design, and future research.

2. Theoretical Foundations: Strategic Communication as Purposeful Influence

Strategic communication emerged as an umbrella construct unifying six previously fragmented disciplines, management communication, marketing, public relations, technical communication, political communication, and social marketing, under the shared premise that organizations act as social agents whose survival depends on purposeful, audience-directed communication (Hallahan et al., 2007). Hallahan et al. (2007) argued that what distinguishes strategic communication from communication in general is not its form or channel but its intentionality: communication becomes strategic when it is deliberately designed to advance an organization's mission by shaping the attitudes, beliefs, or behaviors of a defined public. This definition is analytically useful precisely because it is agnostic about the moral valence of the communicating actor. A humanitarian organization coordinating disaster relief messaging and a hostile state security service running a coordinated inauthentic network are both, in the Hallahan et al. (2007) sense, practicing strategic communication; they differ only in mission, transparency, and truthfulness.

This conceptual neutrality matters for counter-propaganda scholarship because it clarifies that foreign propaganda is not a communicative aberration but a mirror-image application of the same strategic logic that democratic institutions use for public diplomacy, national branding, and crisis communication. Foreign information manipulation and interference, the term increasingly used by European institutions to describe state-directed, coordinated, and deceptive influence activity, is strategic communication practiced by an adversarial organization whose mission is destabilization rather than persuasion toward a disclosed, legitimate goal (Bateman & Jackson, 2024). Recognizing this symmetry has an important implication: the discipline best equipped to diagnose the tactics of a strategic communicator, whether benign or hostile, is

strategic communication itself. Defensive strategic communication, therefore, is not simply louder or more frequent messaging; it is the disciplined application of audience research, message testing, channel selection, and evaluation to the specific task of preserving a population's capacity for independent judgment. Bateman and Jackson (2024), in a comprehensive Carnegie Endowment policy review of ten commonly proposed counter-disinformation interventions, found that no single intervention performed well simultaneously on the criteria of evidentiary strength, effect size, and scalability. Media literacy education and support for local journalism showed the strongest evidence of meaningful impact but were the most difficult to scale, while platform-level interventions such as algorithmic downranking were easier to scale but had thinner evidentiary support. This finding is consistent with the strategic communication premise that no single channel or tactic substitutes for an integrated campaign; it also anticipates the argument developed later in this article that durable resilience requires layered, mutually reinforcing interventions rather than a single silver-bullet countermeasure.

3. Cyberpsychological Mechanisms of Susceptibility to Online Propaganda

Where strategic communication explains the intentional structure of an influence campaign, cyberpsychology explains why that structure succeeds against the human mind operating inside a digital environment. Van der Linden (2022), synthesizing over a decade of research on misinformation susceptibility, proposed that belief in false or misleading content arises from a predictable interaction of individual, social, and technological factors: limited analytical reasoning under time pressure, reliance on heuristic cues such as source familiarity and repetition, motivated reasoning that protects identity-congruent beliefs from revision, and platform architectures that reward emotionally arousing and novel content with greater algorithmic reach. Critically, van der Linden (2022) reframed misinformation susceptibility using an epidemiological metaphor: false beliefs spread through populations in a manner structurally analogous to viral contagion, with individuals functioning as vectors of transmission whenever they share unverified content within their social networks.

This epidemiological framing carries a specific and consequential implication for cyberpsychology, namely that immunity, not merely correction, is the appropriate unit of intervention. Ziemer and Rothmund (2024), in a systematic scoping review of 176 psychological studies of misinformation countermeasures, distinguished between interventions grounded in classical reasoning accounts, which assume that susceptibility results from insufficient analytical scrutiny and can be remedied by improving reasoning quality, and interventions grounded in motivated reasoning accounts, which assume that susceptibility results from identity-protective cognition and cannot be remedied by information alone. Their review found that the large majority of existing interventions, particularly fact-checking, have been designed around the classical reasoning account and are poorly suited to counter identity-driven acceptance of propaganda, precisely the mechanism that foreign influence operations most often exploit when they target polarized political, ethnic, or religious identities.

A second cyberpsychological mechanism of particular relevance to foreign propaganda is astroturfing, the fabrication of an artificial impression of grassroots public support through coordinated fake accounts, sock puppets, or social bots. Zerback et al. (2021), in an experimental study of 2,353 participants exposed to pro-Russian astroturfing comments, found that such comments measurably shifted political opinions and

increased opinion uncertainty even among participants who had received prior inoculation messaging, although inoculation still reduced the magnitude of the effect relative to no intervention at all. This finding is theoretically important because it demonstrates that astroturfing operates through a social-normative mechanism, the perception of consensus, rather than solely through the persuasive content of the message itself; it manipulates the cyberpsychological heuristic that widely shared opinions are more likely to be valid, a heuristic that functions efficiently in ordinary social life but becomes a vulnerability when public consensus can be cheaply fabricated at scale.

Taken together, these findings indicate that cyberpsychological susceptibility to foreign propaganda is not a deficit of intelligence or education but a predictable byproduct of cognitive architecture operating under the specific conditions that digital platforms create: high information volume, low friction for sharing, algorithmic amplification of emotional content, and the ease of manufacturing artificial social proof. Effective countermeasures must therefore be designed with equal attention to the platform environment and to the psychological processes it activates.

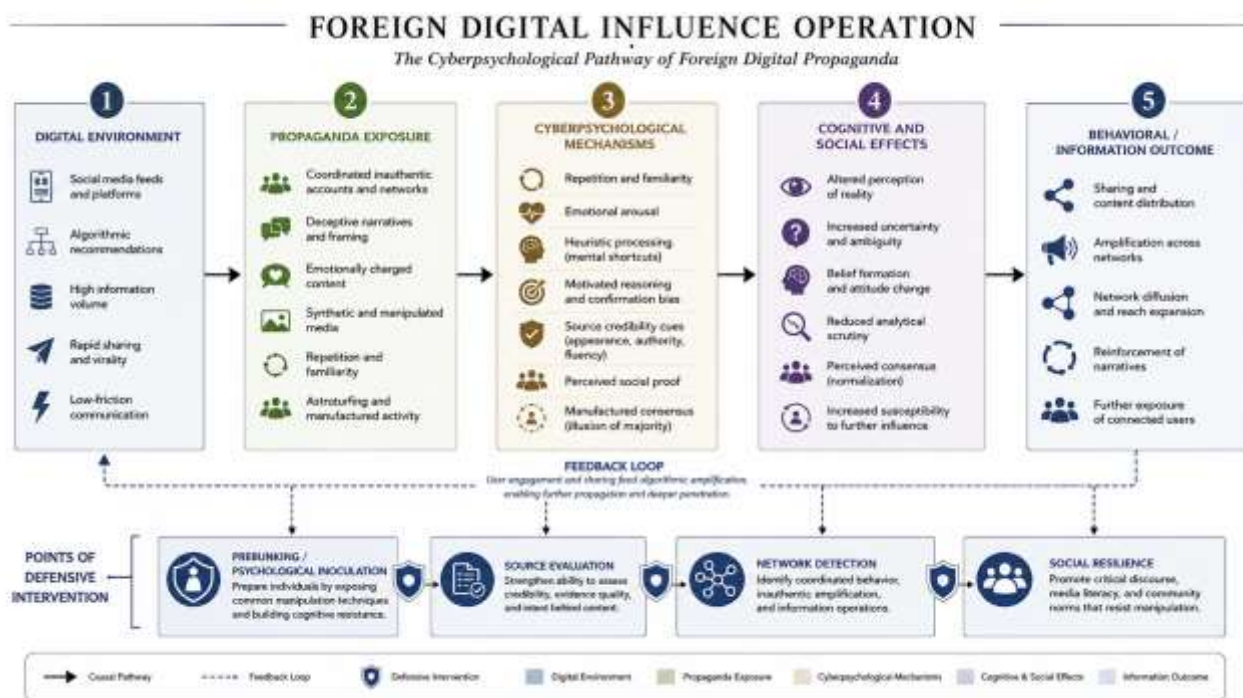


Figure 2. The Cyberpsychological Pathway of Foreign Digital Propaganda

4. The Operational Anatomy of Foreign Digital Propaganda

Understanding how foreign influence operations are technically and organizationally constructed sharpens the diagnostic value of the strategic communication and cyberpsychology frameworks reviewed above. Computational research has increasingly moved beyond describing propaganda content to detecting the coordinated account networks that distribute it. Bailey (2022) applied centering resonance analysis and community-detection algorithms to distinguish Russian disinformation accounts from ordinary Twitter users active during the 2016 United States presidential election, achieving a Matthews correlation coefficient of 0.907 for identifying coordinated bot accounts on the basis of discursive similarity alone, that is, the

accounts could be detected because they echoed one another's language patterns far more closely than authentic, independent users ever do. This finding illustrates a structural signature of foreign propaganda that is largely invisible to individual message recipients but highly detectable at the network level: coordination without independent variation.

More recent research extends this diagnostic approach beyond open platforms such as Twitter/X to encrypted or semi-closed messaging environments. Kireev et al. (2024), analyzing more than seventeen million Telegram messages during the Russia-Ukraine war, identified two independent coordinated networks distributing pro-Russian and pro-Ukrainian propaganda respectively, and developed a detection method exploiting the relationship between authentic user replies and propaganda messages that achieved a 97.6% detection rate after observing only a single message from an account, outperforming human moderators by 11.6 percentage points. This case demonstrates that foreign propaganda infrastructure has adapted to platform migration, moving into spaces with weaker content moderation, while simultaneously confirming that its structural signatures, coordination, replication, and reply-network asymmetries, remain empirically detectable regardless of platform.

Bakirov and Suleimenov (2025), in a broad interdisciplinary review of information warfare synthesizing 592 publications, catalogued the expanding technical toolkit of contemporary influence operations, including botnets, memetic strategies, and increasingly, generative artificial intelligence used to produce synthetic text, images, and deepfake video at low marginal cost. The authors emphasized that the shift from labor-intensive human troll farms toward generative-AI-assisted content production compresses the cost and time required to produce large volumes of tailored propaganda, a development that plausibly outpaces the capacity of manual fact-checking to respond and strengthens the case, developed further below, for anticipatory rather than purely reactive countermeasures. Consistent with this concern, Huntsman (2025) situates the emerging field of cognitive security at the intersection of information integrity and artificial intelligence, arguing that state and non-state actors increasingly compete not merely to control information but to shape the underlying belief-formation processes of target populations, a contest she terms, following classical rhetorical framing, a war for reality.

5. Toward Integration: Psychological Inoculation as the Bridge Between Fields

If foreign propaganda operates simultaneously through message design (a strategic communication phenomenon) and cognitive exploitation (a cyberpsychological phenomenon), then the countermeasure most likely to succeed is one native to both fields. Inoculation theory, originally developed by McGuire in 1964 as a model of resistance to persuasion and revived over the past decade as the theoretical basis for "prebunking," occupies precisely this bridging position (van der Linden, 2022). The theory proposes that exposing individuals to a weakened, forewarned dose of a manipulation technique, accompanied by refutational content explaining why the technique is misleading, confers resistance to subsequent, full-strength exposure to that same technique, in the same way that a vaccine confers immunological resistance to a pathogen (Traberg et al., 2022).

The empirical record supporting inoculation as a countermeasure against foreign propaganda specifically, rather than misinformation in general, is unusually strong for a behavioral intervention. Roozenbeek et al. (2022), in a set of preregistered randomized controlled trials conducted in partnership with a major

technology platform, developed five short inoculation videos targeting common manipulation techniques, emotionally manipulative language, incoherence, false dichotomies, scapegoating, and ad hominem attacks, and found that exposure significantly improved participants' ability to discern manipulative content and reduced their intention to share it, with effects that generalized across the ideological spectrum. Traberg et al. (2022) further distinguished issue-based inoculation, which targets a specific false claim, from technique-based inoculation, which targets the underlying rhetorical strategy and therefore generalizes across topics, a distinction of direct strategic relevance because foreign propaganda campaigns routinely recycle a small set of manipulation techniques across many different narrative topics and target audiences. Two features of inoculation research make it uniquely suited to synthesis with strategic communication theory. First, its efficacy depends on message design choices, framing, narrative structure, refutation strength, that fall squarely within the professional competence of strategic communicators rather than psychologists alone (Traberg et al., 2022). Second, because inoculation operates prophylactically, before exposure to the harmful message, it aligns with the strategic communication principle of proactive, planned campaign design rather than reactive crisis response, allowing institutions to build psychological resistance into public information campaigns on a predictable schedule rather than scrambling to debunk each new falsehood as it emerges (Bateman & Jackson, 2024). At the same time, existing scholarship candidly acknowledges the limits of inoculation: Ziemer and Rothmund (2024) note that psychological inoculation, while effective at the level of the individual recipient, does not by itself address the weak media environments or intergroup dynamics that motivate people to seek out and voluntarily engage with propaganda in the first place, underscoring the need for the multi-level framework proposed below.

6. The Layered Cognitive Defense Model

Building on the theoretical and empirical review above, this article proposes an original integrative framework, the Layered Cognitive Defense Model (LCDM), which organizes counter-propaganda strategy into four interdependent layers: the perceptual layer, the cognitive layer, the social layer, and the institutional layer. The central analytic claim of the LCDM is that foreign propaganda succeeds by exploiting weaknesses across all four layers simultaneously, and that a countermeasure operating at only one layer will be systematically outflanked at the others; durable resilience therefore requires coordinated investment across the full stack rather than a single flagship intervention.

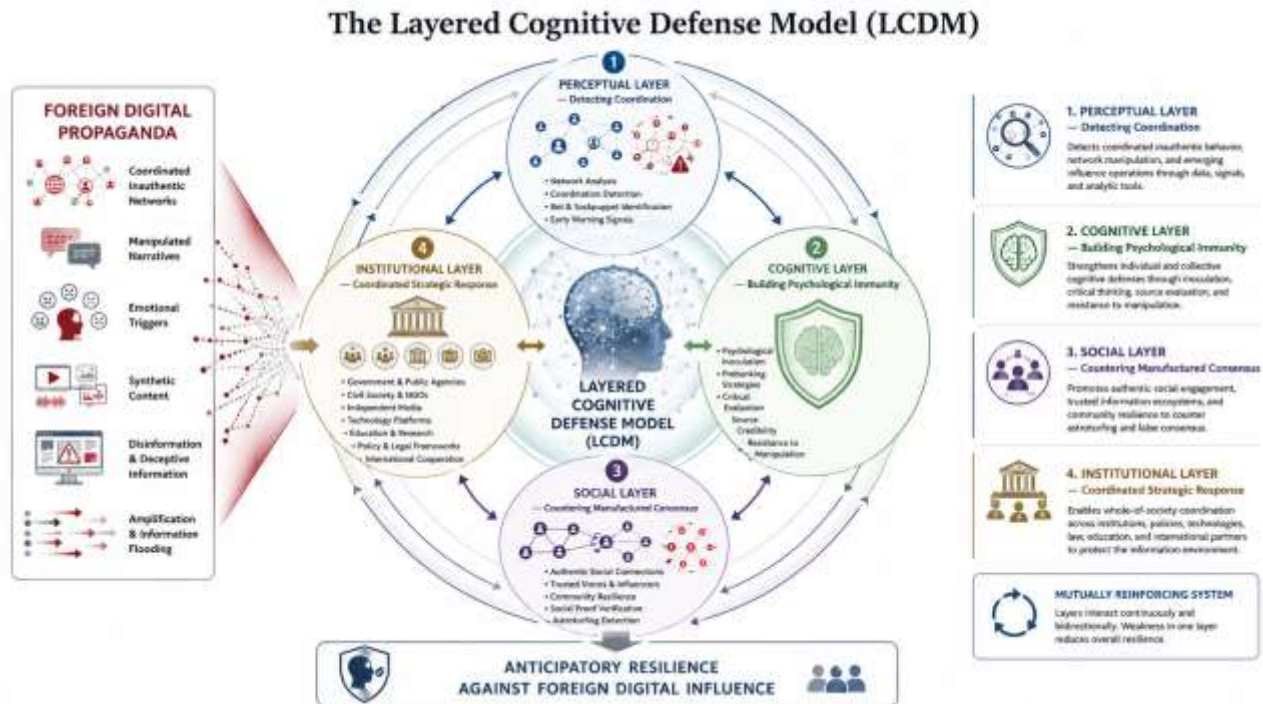


Figure 1. The Layered Cognitive Defense Model (LCDM)

6.1. The Perceptual Layer: Detecting Coordination

The perceptual layer corresponds to the technical detection of inauthentic coordination before content reaches human audiences at scale, drawing directly on the computational methods of Bailey (2022) and Kireev et al. (2024). Because coordinated networks exhibit measurable discursive and network-structural signatures distinct from organic public discourse, platform-level and government-level monitoring capacity focused on network topology rather than message content offers an early-warning function that operates upstream of individual cognition entirely. This layer is necessary but not sufficient: detection without downstream psychological or institutional response merely identifies the threat without neutralizing it.

6.2. The Cognitive Layer: Building Psychological Immunity

The cognitive layer corresponds to individual-level psychological resistance built through technique-based inoculation, consistent with the evidence reviewed from Roozenbeek et al. (2022) and Traberg et al. (2022). Strategic communicators, applying the audience segmentation and message-testing competencies described by Hallahan et al. (2007), are positioned to design and disseminate inoculation content tailored to specific populations, languages, and manipulation techniques currently in circulation, effectively translating psychological theory into deployable public communication campaigns. This layer directly targets the classical-reasoning component of susceptibility identified by Ziemer and Rothmund (2024) but must be paired with the social layer to address motivated reasoning.

6.3. The Social Layer: Countering Manufactured Consensus

The social layer addresses the astroturfing and manufactured-consensus mechanisms documented by Zerback et al. (2021), in which propaganda exploits social proof heuristics rather than persuading through content alone. Countermeasures at this layer include transparency labeling of coordinated accounts once

detected at the perceptual layer, strategic amplification of authentic, trusted community voices, and public communication that explicitly names the astroturfing tactic, a technique consistent with technique-based inoculation but delivered through social rather than purely informational channels. This layer is where strategic communication's traditional expertise in stakeholder relations and community engagement becomes most directly operational.

6.4. The Institutional Layer: Coordinated Strategic Response

The institutional layer corresponds to the governance, policy, and cross-sector coordination mechanisms documented by Bateman and Jackson (2024) and by Rød et al. (2025), whose typological framework of societal resilience to disinformation spans legal and regulatory, educational, political and governance, psychological and social-psychological, and technological domains. Rød et al. (2025) argue that resilience indicators across these domains should ultimately be integrated into a measurable, comparable framework that allows policymakers to assess maturity and target investment where it is weakest, a recommendation directly compatible with the LCDM's premise that the four layers must be evaluated and resourced jointly rather than in isolation. Without institutional coordination, gains achieved at the perceptual, cognitive, and social layers remain fragmented, under-resourced, and vulnerable to the adversary simply shifting tactics or platforms, as observed in the migration of coordinated networks toward Telegram documented by Kireev et al. (2024).

The four layers of the LCDM are mutually reinforcing rather than sequential. Perceptual detection generates the threat intelligence that informs which manipulation techniques should be prioritized in cognitive-layer inoculation content; cognitive-layer resistance reduces the efficacy of social-layer astroturfing by making individuals less susceptible to manufactured consensus; and institutional-layer coordination ensures that gains at the other three layers are sustained, resourced, and adapted as adversarial tactics evolve. This architecture directly answers the central problem identified at the outset of this article: that strategic communication supplies the institutional and campaign-design competence while cyberpsychology supplies the mechanistic understanding of susceptibility, and that foreign propaganda cannot be durably countered by either discipline operating alone.

7. Discussion and Implications

The evidence reviewed here carries three implications for policy and practice. First, the strong evidentiary base for technique-based psychological inoculation (Roozenbeek et al., 2022; Traberg et al., 2022) suggests that governments and platforms should treat prebunking as a standing component of public communication infrastructure rather than an emergency response tool deployed only during elections or crises. Second, the demonstrated detectability of coordinated inauthentic networks (Bailey, 2022; Kireev et al., 2024) argues for sustained investment in network-level, rather than purely content-level, detection capacity, since manipulation techniques and narrative content evolve faster than any static keyword-based detection system can track. Third, the finding that no single intervention scores well on effectiveness, evidentiary strength, and scalability simultaneously (Bateman & Jackson, 2024) reinforces the LCDM's central architectural claim that resilience is an emergent property of coordinated, multi-layer investment rather than a property of any individual tool.

These implications also carry a normative caution. Because strategic communication is, as Hallahan et al. (2007) established, valence-neutral in its logic of purposeful influence, the same inoculation and detection capabilities recommended here could in principle be redirected toward domestic political manipulation rather than genuine defense against foreign interference. Democratic institutions adopting the LCDM should therefore pair its technical and psychological components with transparent governance safeguards, independent oversight of detection systems, and clear public disclosure of inoculation campaign content and funding, consistent with the emphasis Bateman and Jackson (2024) place on countering disinformation without undermining the civil liberties and institutional trust that such efforts are ultimately meant to protect.

8. Limitations and Future Research

This synthesis is subject to several limitations. Most experimental inoculation research reviewed here has been conducted in North American, European, and other high-income, high-internet-penetration contexts, and evidence for its cross-cultural generalizability, particularly in the low- and middle-income information environments increasingly targeted by foreign influence operations, remains comparatively limited. Additionally, the rapid diffusion of generative artificial intelligence tools for propaganda production, noted by Bakirov and Suleimenov (2025) and Huntsman (2025), is likely to outpace the experimental literature on countermeasures, meaning that inoculation content calibrated to current manipulation techniques may require continual updating as synthetic media techniques evolve. Future research should prioritize longitudinal field experiments measuring the durability of inoculation effects over multi-month periods, cross-national replication of astroturfing and inoculation studies outside Western democracies, and the integration of real-time network detection data, of the kind produced by Bailey (2022) and Kireev et al. (2024), directly into the design and targeting of cognitive-layer interventions.

9. Conclusion

Foreign digital propaganda succeeds because it is engineered simultaneously as a strategic communication campaign and a cyberpsychological exploit, and it can only be reliably countered by a response that treats both dimensions as inseparable. This article has argued that strategic communication supplies the institutional discipline of purposeful, audience-centered messaging (Hallahan et al., 2007), while cyberpsychology supplies the empirical account of why particular manipulation techniques succeed against particular cognitive and social vulnerabilities (van der Linden, 2022; Ziemer & Rothmund, 2024). The Layered Cognitive Defense Model proposed here integrates network-level detection, technique-based psychological inoculation, social-layer counter-messaging against manufactured consensus, and institutional coordination into a single analytic architecture designed to build anticipatory resilience rather than perpetual, reactive correction. As generative artificial intelligence continues to lower the cost of producing convincing synthetic propaganda, the strategic imperative shifts decisively from debunking falsehood after it spreads to inoculating populations, networks, and institutions before exposure occurs.

References

1. Bailey, M. M. (2022). Detecting propagators of disinformation on Twitter using quantitative discursive

- analysis (arXiv:2210.05760). Journal of Intelligence Community Research and Development. <https://doi.org/10.48550/arXiv.2210.05760>
2. Bakirov, A., & Suleimenov, I. (2025). Theoretical bases of methods of counteraction to modern forms of information warfare. *Computers*, 14(10), Article 410. <https://doi.org/10.3390/computers14100410>
 3. Bateman, J., & Jackson, D. (2024). Countering disinformation effectively: An evidence-based policy guide. Carnegie Endowment for International Peace. <https://carnegieendowment.org/research/2024/01/countering-disinformation-effectively-an-evidence-based-policy-guide>
 4. Hallahan, K., Holtzhausen, D., van Ruler, B., Verčič, D., & Sriramesh, K. (2007). Defining strategic communication. *International Journal of Strategic Communication*, 1(1), 3–35. <https://doi.org/10.1080/15531180701285244>
 5. Huntsman, L. H. (2025). Editorial: Disinformation countermeasures and artificial intelligence. *Frontiers in Artificial Intelligence*, 8, Article 1750972. <https://doi.org/10.3389/frai.2025.1750972>
 6. Kireev, K., Mykhno, Y., Troncoso, C., & Overdorf, R. (2024). Characterizing and detecting propaganda-spreading accounts on Telegram (arXiv:2406.08084). arXiv. <https://doi.org/10.48550/arXiv.2406.08084>
 7. Rød, B., Pursiainen, C., & Eklund, N. (2025). Combatting disinformation: How do we create resilient societies? Literature review and analytical framework. *European Journal for Security Research*, 10, Article 105. <https://doi.org/10.1007/s41125-025-00105-4>
 8. 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), Article eabo6254. <https://doi.org/10.1126/sciadv.abo6254>
 9. 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>
 10. van der Linden, S. (2022). Misinformation: Susceptibility, spread, and interventions to immunize the public. *Nature Medicine*, 28(3), 460–467. <https://doi.org/10.1038/s41591-022-01713-6>
 11. 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>
 12. Zerback, T., Töpfl, F., & Knöpfle, M. (2021). The disconcerting potential of online disinformation: Persuasive effects of astroturfing comments and three strategies for inoculation against them. *New Media & Society*, 23(5), 1080–1098. <https://doi.org/10.1177/1461444820908530>
 13. Ziemer, C.-T., & Rothmund, T. (2024). Psychological underpinnings of misinformation countermeasures. *Journal of Media Psychology*, 36(6), 397–409. <https://doi.org/10.1027/1864-1105/a000415>