

COUNTER-PROPAGANDA

對抗宣傳

Concepts, History, Psychology, Strategy, Law, and Resilient Information
Ecosystems



The American Newspaper | <https://americannewspaper.org>

AmericanTV | <https://americantv.org>

Research Report • August 12, 2026

This report distinguishes documented historical facts, evidence-supported interpretations, and contested claims. Government self-descriptions are treated as sources to be evaluated, not as conclusive proof.

Executive Summary

Counter-propaganda is best understood not as a single act of rebuttal but as a layered defensive and persuasive system designed to reduce the influence, credibility, reach, and durability of hostile or manipulative communication. It can operate at the level of a claim, source, frame, narrative, social network, algorithm, platform, or institution. Its most defensible democratic form does not seek to monopolize truth by state decree; rather, it increases the cost of deception, improves verification, supplies context and credible alternatives, protects independent journalism, and strengthens the audience's ability to resist manipulation.

The core analytical mistake is to equate counter-propaganda with “saying the opposite.” Propaganda often works through identity, emotion, repetition, agenda-setting, social proof, selective framing, and control of distribution. A factual correction may therefore leave the deeper narrative intact. Effective response requires diagnosing what the adversarial communication is doing: Is it introducing a false fact, undermining trust, creating ambiguity, provoking fear, polarizing groups, laundering a narrative through proxies, or overwhelming attention with contradictory content? Different mechanisms require different interventions.

The research literature also complicates several popular assumptions. Repetition can increase familiarity and perceived truth even when a claim is false, so careless corrections can keep a falsehood cognitively available. The much-discussed “backfire effect,” in which corrections reliably strengthen false beliefs, appears less general and less frequent than early discussion suggested; later experimental work generally finds that corrections improve factual accuracy, though effects can be modest, uneven, short-lived, and filtered by identity and source trust. Pre-bunking and inoculation can improve resistance to manipulation techniques, but their effects decay and often require reinforcement.

Historically, counter-propaganda evolved from wartime rebuttal and morale operations into a broader architecture of broadcasting, public diplomacy, psychological operations, independent or state-funded journalism, exposure of covert influence, platform interventions, and societal resilience. The First and Second World Wars institutionalized large propaganda bureaucracies; the Cold War added transnational radio, cultural diplomacy, defectors, leaflets, and “active measures” competition; the post-9/11 era emphasized counterterrorism narratives; and the social-media/AI era has shifted the unit of competition from the message alone to networks, recommender systems, synthetic media, and computational targeting.

The democratic challenge is structural. A state may have legitimate reasons to expose foreign deception, warn citizens, protect elections, or counter incitement, yet the same institutional capacity can become a tool for censorship, partisan narrative control, covert propaganda, or pressure on platforms and journalists. The boundary between democratic counter-propaganda and authoritarian information control therefore depends less on what an institution calls itself than on transparency, accuracy, source disclosure, independent verification, proportionality, legal authority, judicial or legislative oversight, press freedom, contestability, and citizen autonomy.

Bottom line: The strongest long-term form of counter-propaganda is not “better propaganda.” It is an information environment in which deception is easier to detect, corrections travel through trusted channels, institutions disclose

their evidence, journalists can investigate independently, citizens can challenge official claims, and audiences develop durable habits of verification.

How to Read the Evidence

Label	Meaning
Established / well documented	Supported by archival records, primary documents, or broad scholarly agreement.
Strongly supported	Supported by multiple empirical studies, but effect size or boundary conditions vary.
Contested / context dependent	Credible scholarship disagrees, evidence is mixed, or causal attribution is difficult.
Official claim	A government or organization's stated description of its own activities; useful as evidence of intent but not treated as independently verified fact.



Figure 1. Generated visual overview of counter-propaganda: conceptual boundaries, historical evolution, mechanisms, legal safeguards, and evaluation. The figure is illustrative; the report text is authoritative where details differ.

Contents

- 1. Definition, Etymology, and Conceptual Origins
- 2. Conceptual Boundaries: Counter-Propaganda and Adjacent Fields
- 3. Strategic Objectives of Counter-Propaganda
- 4. Major Forms: Advantages, Limits, and Unintended Effects
- 5. Psychology of Success and Failure
- 6. Historical Development: From World War I to Generative AI
- 7. Comparative Systems: United States, United Kingdom, USSR/Russia, China, Nazi Germany, North Korea
- 8. Major Thinkers: Propaganda, Persuasion, Cognition, and Social Influence
- 9. Units of Analysis and Variables of Effectiveness
- 10. Digital Transformation: Bots, Trolls, Deepfakes, Algorithms, Influencers, Microtargeting, and AI
- 11. Democracy, Law, Human Rights, Journalism, and the Risk of State Abuse
- 12. Evaluating Effectiveness
- 13. Comparative Case Studies: Successes, Failures, and Mixed Outcomes
- 14. Integrated Propaganda–Counter-Propaganda Model
- 15. Design Principles for a Resilient Information Ecosystem
- Conclusion
- Representative Bibliography and Source Guide

1. Definition, Etymology, and Conceptual Origins

1.1 What 對抗宣傳 Means

The Chinese expression 對抗宣傳 is a transparent compound. 對抗 (duìkàng) denotes opposition, resistance, confrontation, or counteraction; 宣傳 (xuānchuán) denotes propagation, publicity, advocacy, or propaganda. In a literal sense, 對抗宣傳 therefore means communication undertaken to resist, neutralize, answer, or counter another propaganda effort. The expression is semantically broader than “fact-checking,” because it can target not only factual propositions but also emotional appeals, frames, narratives, sources, distribution networks, and the social conditions that make propaganda persuasive.

In modern Chinese and East Asian usage, 宣傳 does not map perfectly onto the pejorative English sense of “propaganda.” The term can also refer neutrally to publicity, promotion, or dissemination. Accordingly, 對抗宣傳 can describe a range of practices from overt rebuttal to organized information response. The analytical task is to identify function and method rather than to infer legitimacy from the label.

1.2 English Term: counter-propaganda / counterpropaganda

The English compound “counter-propaganda” emerged from the vocabulary of mass politics, wartime information operations, and early twentieth-century propaganda studies. During the First World War, belligerents did not merely broadcast their own narratives; they also organized efforts to rebut, discredit, expose, or pre-empt enemy claims. Interwar scholarship already discussed wartime material as propaganda and counter-propaganda, and the term became more explicit in military, diplomatic, and psychological-warfare literature during the Second World War and Cold War. By the mid-twentieth century, counter-propaganda was treated as a recognizable component of psychological warfare: analysis of adversary claims, selection of vulnerabilities, and communication designed to reduce their effect.

The term later partially receded from some official U.S. joint doctrine, where vocabulary such as “countering adversary information,” “military information support operations,” “strategic communication,” “information operations,” and “counter-disinformation” became more common. This change in terminology did not eliminate the underlying function. Contemporary practice often disperses counter-propaganda across public affairs, intelligence, diplomatic communication, cyber operations, platform policy, and civil-society fact-checking rather than locating it in a single bureau.

Historical caution: There is no single founding document that “invented” counter-propaganda. The practice predates the term. The concept crystallized as states professionalized mass persuasion, intelligence, and psychological warfare in the twentieth century.

1.3 A Working Definition

Counter-propaganda is the systematic use of truthful or evidentially defensible communication, source analysis, exposure, narrative competition, distribution strategy, and audience-resilience measures to reduce the persuasive, behavioral, or diffusion effects of adversarial propaganda or manipulative information activity. In democratic settings, the strongest normative version of this definition adds

procedural constraints: transparency about official involvement, respect for rights, independent verification, and avoidance of coercive information monopoly.

1.4 Why “Refuting the Adversary” Is Inadequate

1. Propaganda is rarely only a claim. It can be a frame that determines which facts are salient and which are ignored.
2. It often links information to identity, emotion, grievance, status, fear, belonging, or moral judgment. A factual rebuttal may not alter those attachments.
3. It can exploit source confusion and proxy networks. Correcting the visible post does not necessarily expose the hidden sponsor, amplification system, or laundering chain.
4. It can function through repetition and attention capture. Rebutting every claim may inadvertently repeat and circulate the adversary’s agenda.
5. It can aim to create uncertainty rather than belief. Flooding the environment with mutually inconsistent stories can reduce trust even if none is fully believed.
6. It can be embedded in memes, entertainment, rumors, influencers, edited video, or authentic documents presented misleadingly. A “true/false” format may miss the manipulation.
7. It can target behavior rather than opinion: discouraging voting, increasing panic, provoking protest, inducing defection, or making institutions appear incapable.
8. Long-run defense requires resilience. If the audience remains dependent on weak institutions and opaque sources, the next narrative can recreate the same vulnerability.

2. Conceptual Boundaries: Counter-Propaganda and Adjacent Fields

The following concepts overlap but are not interchangeable. The most useful distinction is functional: what is being targeted, by whom, at what scope, and with what legal or institutional authority?

Concept	Core function	Typical scope	Typical actors
Propaganda	Systematic communication intended to shape perceptions, beliefs, emotions, identities, or behavior in support of an actor or cause.	Broad: messages, symbols, narratives, institutions.	States, parties, movements, firms, activists.
Counter-propaganda	Systematic effort to weaken, expose, neutralize, reframe, or replace hostile propaganda and strengthen audience resilience.	Claims, sources, frames, narratives, networks, institutions.	States, media, civil society, platforms, communities.
Counter-messaging	Targeted messages that directly challenge extremist, adversarial, or misleading content.	Usually message/campaign level.	Government, NGOs, community groups.
Strategic communication	Coordinated use of communication, actions, policy signals, and institutional behavior to achieve strategic objectives.	Whole-of-organization or whole-of-government.	Governments, militaries, organizations.
Public diplomacy	Engagement with foreign publics to build relationships, explain policy, and influence perceptions.	Foreign audiences; long-term relationships.	Foreign ministries, cultural institutions, broadcasters.
PSYOP / MISO	Planned military activities directed at selected foreign audiences to influence emotions, motives, reasoning, and behavior in support of operations.	Operational/tactical, mission-oriented.	Military organizations.
Information operations	Coordinated employment of	Systems, processes, networks,	Military, intelligence, cyber

Concept	Core function	Typical scope	Typical actors
	information-related capabilities to influence, disrupt, corrupt, or protect decision-making.	decision environments.	organizations.
Information warfare	Strategic contest for advantage across information creation, control, denial, exploitation, and protection.	Broad information environment.	States and armed actors.
Influence operations	Overt or covert efforts to shape attitudes, choices, or behavior, often through networks, proxies, or covert sponsorship.	Audiences and social systems.	State and non-state actors.
Cognitive warfare	An emerging umbrella concept focused on influencing perception, judgment, identity, and decision processes at scale.	Human cognition and social meaning systems.	Military/strategic communities; concept remains contested.
Fact-checking	Verification of specific factual claims against available evidence.	Claim level.	Journalists, fact-checkers, researchers.
Debunking	Post-exposure correction that explains why a false or misleading claim is wrong and ideally supplies an alternative account.	Claim/frame level after exposure.	Media, experts, institutions, platforms.
Pre-bunking	Pre-exposure warning or inoculation against manipulation techniques, narrative patterns, or likely false claims.	Preventive, audience-resilience level.	Educators, platforms, governments, civil society.
Media literacy	Skills and habits for evaluating sources, evidence, context, ownership, incentives, and manipulation techniques.	Long-term individual and community capacity.	Schools, media, libraries, civil society.
Counter-disinformation	Broader ecosystem response to deliberately deceptive information, including detection, attribution, exposure, disruption, regulation, and resilience.	Content + networks + platforms + institutions.	Multi-stakeholder.

Key boundary: Fact-checking asks “Is this claim accurate?” Counter-propaganda asks the broader question “How does this information operation work, why is it persuasive, how is it distributed, and what combination of responses will reduce its effect without causing greater harm?”

3. Strategic Objectives of Counter-Propaganda

Objective	Operational meaning
Undermine hostile credibility	Raise justified doubt about deceptive sources, methods, incentives, hidden sponsorship, or track record.
Expose false or misleading information	Correct verifiable errors and make the evidentiary basis of the correction visible.
Dismantle frames	Show how selective facts, false dichotomies, scapegoating, conspiratorial structure, or loaded language shape interpretation.
Verify sources	Establish provenance, authenticity, chain of custody, ownership, and independence of the source.
Build cognitive immunity	Teach recurring manipulation techniques so audiences recognize them before accepting the next message.
Provide alternative explanations	Fill the explanatory vacuum with a coherent account that better fits the evidence and the audience’s lived experience.
Protect social trust	Avoid allowing adversarial narratives to convert institutional

Objective	Operational meaning
	errors into generalized nihilism or mutual suspicion.
Protect institutional legitimacy	Demonstrate competence, admit uncertainty or mistakes, and provide auditable evidence rather than demanding deference.
Reduce harmful diffusion	Slow virality, reduce algorithmic amplification, remove coordinated inauthentic behavior where rules permit, and increase friction.
Preserve decision autonomy	Ensure citizens and decision-makers can choose with access to reliable information rather than through covert manipulation.

These objectives can conflict. Rapid takedown may reduce reach but also create censorship claims; exposing a covert actor may improve attribution but reveal intelligence methods; repeating a false claim may improve correction among one audience while increasing familiarity among another. Counter-propaganda therefore requires optimization across multiple goals, not maximization of a single metric.

4. Major Forms: Advantages, Limits, and Unintended Effects

Form	What it does	Advantages	Limitations / risks	Best practice
Direct refutation	Explicitly state that a claim is false or misleading; supply evidence and a correction.	Clear and fast; useful when the audience has not deeply committed.	Can repeat the falsehood; may trigger source distrust or identity defense.	Use a truth-first structure; avoid sensational repetition.
Indirect refutation	Provide contrary evidence or context without foregrounding the false claim.	Reduces repetition; can reach less polarized audiences.	May be too subtle when the falsehood is viral.	Useful when salience is low or the claim is not yet dominant.
Fact-checking	Verify specific propositions using documented evidence.	Auditable; compatible with journalism and science.	Claim-by-claim approach may not change a broader narrative.	Best for discrete, testable claims.
Source exposure	Reveal sponsorship, coordination, forged provenance, bot networks, conflicts of interest, or prior deception.	Can collapse credibility across many messages at once.	Attribution errors can be costly; "who said it" can substitute for evidence.	Use confidence levels and disclose methodology.
Logical contradiction	Show mutually inconsistent claims, impossible timelines, or internal contradictions.	Strong when audiences value coherence.	People can rationalize contradictions when identity stakes are high.	Pair logic with concrete evidence and a replacement explanation.
Reframing	Change the interpretive question, causal model, moral frame, or reference point.	Targets the structure of meaning rather than only facts.	May be perceived as counter-spin.	Most effective when the existing frame is visible and contestable.
Alternative narrative	Offer a coherent, credible story that explains events without relying on the adversary's frame.	Fills explanatory and emotional vacuum; can restore agency.	Requires time, messengers, and real-world policy consistency.	Use when audiences need meaning, identity, or future orientation.
Pre-bunking / inoculation	Warn about likely manipulation techniques or weakened examples before exposure.	Builds generalized resistance; avoids chasing every falsehood.	Effects decay; may not transfer across topics or contexts.	Reinforce periodically; teach techniques, not just claims.
Humor and satire	Use ridicule, parody, memes, or cultural references to puncture status and emotional power.	Highly shareable; can weaken intimidation and pretension.	May alienate, trivialize harm, or strengthen in-group mockery.	Use with cultural fluency and clear factual anchor.
Platform-level response	Labels, provenance signals, ranking changes, friction, takedowns for rule violations, demonetization, coordinated-behavior disruption.	Can reduce reach at scale and target network behavior.	Governance, bias, transparency, and speech concerns; migration to other channels.	Prefer transparent rules, appeals, and behavior-based enforcement.
Education / media literacy	Teach verification, source evaluation, lateral reading, statistical literacy, emotional	Long-term resilience; reduces dependence on centralized correction.	Slow, uneven, resource intensive; sophisticated users can still be biased.	Treat as a civic infrastructure, not a one-time campaign.

Form	What it does	Advantages	Limitations / risks	Best practice
	self-monitoring.			

4.1 The Truth-Sandwich and “Truth First” Structure

A widely used correction design is to begin with the verified fact, briefly identify the myth only when necessary, explain why it is wrong, and end by restating the accurate account. The logic is not that merely mentioning a falsehood is always harmful; rather, the accurate model should remain the most cognitively available explanation. Corrections work better when they provide an alternative causal story instead of simply removing the false one.

4.2 When Not to Respond

Response itself creates attention. If a fringe claim has little reach, a high-status rebuttal can “scale” the adversary’s message by making it newsworthy. Decision-makers should ask: How much organic reach does the claim have? Is the target audience likely to encounter it anyway? Does the correction require repeating vivid false details? Can the response be targeted to affected communities rather than broadcast universally? Non-response is not passivity when paired with monitoring, pre-bunking, quiet outreach to journalists, or network-level disruption.

5. Psychology of Success and Failure

5.1 Familiarity and the Illusory-Truth Effect

Repeated statements become easier to process, and processing fluency can increase perceived truth. This is one reason propaganda benefits from repetition and why counter-propaganda should avoid needlessly turning a false slogan into a memorable refrain. Importantly, the lesson is not “never repeat the myth.” Clear corrections often need to identify what they correct. The practical rule is to minimize unnecessary repetition, emphasize the verified account, and give the audience an explanatory model that can replace the false one.

5.2 The Backfire Effect: Important but Often Overstated

Early political-communication research raised concern that corrective information could sometimes strengthen false beliefs among highly committed groups. Later studies across larger numbers of issues have generally found backfire to be uncommon rather than a routine response. Corrections typically improve factual accuracy, although the improvement may be small and may not translate into policy preference or behavior. A useful distinction is therefore between factual correction and identity or attitude change: people can update a factual belief without changing what they want, whom they trust, or how they vote.

Evidence status: The claim “fact-checking usually backfires” is not supported by the balance of current evidence. The more defensible conclusion is conditional: corrections usually help factual accuracy, but their effects vary by source, identity, prior belief, message design, and time.

5.3 Motivated Reasoning and Identity-Protective Cognition

People do not process political information as detached statisticians. Beliefs can signal group loyalty, moral commitments, status, or worldview. When a correction is interpreted as an attack on a valued identity, the receiver may reject the messenger, reinterpret evidence, or shift from factual disagreement to a contest over legitimacy. Counter-propaganda therefore often works better through trusted in-group messengers, procedural fairness, acknowledgment of legitimate grievances, and separation of factual correction from contempt for the audience.

5.4 Confirmation Bias, Selective Exposure, Echo Chambers, Group Polarization

Audiences disproportionately seek information consistent with prior beliefs, but the popular image of perfectly sealed “echo chambers” can be too simple. People often encounter cross-cutting content incidentally, especially on large platforms. The more important issue is that network structure, homophily, recommendation systems, and group status can reward partisan interpretation. Group discussion among like-minded participants may intensify positions, while social endorsement cues can make a claim appear more normal or credible than its evidence warrants.

5.5 Source Credibility

Hovland’s persuasion research made source credibility central to communication effects, and modern misinformation research reinforces the point. Expertise, trustworthiness, independence, transparency, familiarity, and group identity all matter. A highly accurate correction from a distrusted institution can fail where a less technically expert but trusted local messenger succeeds. Credibility is cumulative capital: it is built before a crisis and can be destroyed by visible inconsistency, concealment, exaggerated certainty, or failure to correct one’s own errors.

5.6 Social Proof and Network Signals

Likes, shares, trending labels, reposts, comments, elite endorsements, and influencer repetition create informational and normative cues. The individual may infer that “many people believe this,” “people like me believe this,” or “this topic must be important.” Modern counter-propaganda must therefore analyze not only content but also apparent consensus. Coordinated inauthentic behavior can simulate social proof; conversely, transparent signals of expert consensus or cross-partisan agreement can sometimes reduce uncertainty if they are credible and not coercive.

Mechanism	Why it matters	Counter-propaganda implication
Familiarity / illusory truth	Repetition increases fluency and can raise perceived truth.	Use truth-first corrections; avoid slogan repetition; repeat the accurate alternative.
Motivated reasoning	Reasoning can defend goals, identity, or desired conclusions.	Use trusted messengers; separate correction from identity threat.
Confirmation bias	Congruent information is easier to accept and retrieve.	Provide evidence in forms that do not require total worldview surrender.
Identity protection	Beliefs can function as group membership signals.	Use in-group validators; preserve dignity and agency.
Selective exposure	Audiences choose congenial sources.	Reach people through channels they already use; do not assume official channels are seen.
Group polarization	Like-minded discussion can intensify	Encourage cross-cutting credible voices;

Mechanism	Why it matters	Counter-propaganda implication
	positions.	avoid public shaming that consolidates groups.
Source credibility	Trustworthiness and expertise shape acceptance.	Disclose evidence, uncertainty, affiliations, corrections, and methods.
Social proof	Perceived consensus and engagement alter credibility and salience.	Expose fake amplification; contextualize engagement metrics.
Emotion / arousal	Fear, anger, disgust, humiliation, and hope influence attention and memory.	Correct facts while addressing the emotional function of the narrative.

6. Historical Development: From World War I to Generative AI

6.1 World War I: Mass Propaganda Becomes a State Function

The First World War demonstrated the political power of mass newspapers, posters, film, rumor, censorship, atrocity stories, and centralized information offices. Britain, Germany, France, the United States, and other belligerents built organizations to mobilize domestic populations and influence foreign audiences. The U.S. Committee on Public Information (CPI), created in 1917 and chaired by George Creel, coordinated speeches, films, posters, publications, and foreign information work. Its records show how wartime governments institutionalized persuasion on a scale that later scholars would treat as a model—and warning—of modern propaganda.

Counter-propaganda in this era was often reactive: answering atrocity allegations, challenging enemy claims, contesting responsibility for the war, and competing for neutral opinion. A major lesson was that poorly designed rebuttal could itself become evidence against the rebutting side. German attempts to answer Allied atrocity propaganda sometimes appeared defensive or implausible to external audiences, illustrating a recurring problem: credibility cannot be manufactured merely by issuing denials.

6.2 World War II: Propaganda, Psychological Warfare, and Specialized Counter-Propaganda

World War II produced mature propaganda bureaucracies. Nazi Germany centralized much of domestic communication under Joseph Goebbels's Ministry of Public Enlightenment and Propaganda while maintaining overlapping party and state structures. The regime used censorship, radio, film, press directives, spectacle, antisemitic conspiracy narratives, and deception; its system illustrates how propaganda can become an instrument of authoritarian social control rather than merely foreign persuasion.

The United States created the Office of War Information in 1942, while military and intelligence organizations conducted psychological warfare. Britain created the Political Warfare Executive in 1941 to undermine enemy morale and support resistance through overt and covert propaganda. British practitioners such as Sefton Delmer became associated with “black” radio that imitated insider German voices while attacking corruption, hypocrisy, and morale. Such methods could be tactically ingenious, but they also complicate contemporary ethical claims that successful counter-propaganda must always be

fully transparent. Historically, covert deception was part of Allied wartime practice; whether it is legitimate in peacetime democratic communication is a separate normative question.

The key innovation was systems thinking: propaganda intelligence, audience research, radio monitoring, leaflet design, rumor analysis, and measurement through prisoner interrogation, defection rates, or resistance reports. Counter-propaganda became an operational cycle: collect enemy messaging, analyze themes and vulnerabilities, select a response, distribute through an appropriate channel, and assess effect.

6.3 Cold War: Broadcasting, Cultural Diplomacy, and “Active Measures”

The Cold War shifted the contest from total war mobilization to permanent ideological competition. Radio became a central instrument because it could cross borders and bypass censorship. Radio Free Europe began broadcasting in 1950 to Eastern Europe; Radio Liberty followed in 1953 for the Soviet Union. These organizations were initially covertly funded in significant part through the CIA before later becoming openly congressionally funded. Their historical importance lies not simply in anti-communist messaging but in providing information unavailable in censored media environments, which could make factual reporting itself a form of strategic competition.

Voice of America began wartime broadcasting in 1942. Over time it developed a charter emphasizing accurate, objective, and comprehensive news while also presenting U.S. policies. That institutional duality—journalistic credibility inside a state-funded strategic mission—captures a central problem of democratic counter-propaganda. Credibility increases when audiences believe the broadcaster will report inconvenient facts; strategic authorities, however, may prefer message discipline. The tension cannot be eliminated, only governed.

The Soviet Union maintained an extensive international propaganda system combining official media, party networks, front organizations, diplomatic channels, cultural organizations, and covert “active measures.” Western governments used the term active measures for influence activities including forgeries, planted stories, covert support, and disinformation. One prominent late-Cold-War example was the false claim that AIDS had been created by the U.S. military. U.S. counter-efforts increasingly emphasized public attribution, documentation of forgeries, diplomatic exposure, and interagency coordination rather than simple denial.

The BBC World Service, Western cultural diplomacy, exchanges, books, exhibitions, music, and intellectual networks also mattered. These channels show that counter-propaganda can work indirectly: rather than debating every adversarial claim, a state or society can increase access to credible information and alternative cultural experience. The audience then compares systems through lived evidence rather than official slogans alone.

6.4 Korean War and Vietnam War

The Korean War used leaflets, loudspeakers, radio, surrender appeals, and psychological warfare on a large scale. Messages sought tactical effects—defection, surrender, reduced morale—as well as broader political legitimacy. The conflict also exposed the difficulty of communicating across language, culture, ideology, and rapidly changing front lines. Tactical success could often be measured more clearly than long-term political persuasion.

Vietnam intensified the credibility problem. U.S. and South Vietnamese psychological operations distributed enormous volumes of leaflets and broadcasts, but the wider “credibility gap” created by discrepancies between official claims and observable events limited persuasive power. Counter-propaganda cannot compensate indefinitely for policy realities that contradict the message. The same principle applies today: strategic communication is constrained by strategic behavior.

6.5 Post-Cold War and Post-9/11 Counterterrorism Communication

After 1991, Western institutions reduced some Cold War structures while public diplomacy and information operations adapted to regional conflicts, satellite television, and then the internet. After the attacks of September 11, 2001, counterterrorism communication became a major policy field. Governments attempted to contest al-Qaeda and later ISIS narratives through official messaging, partnerships, online engagement, and support for community voices.

The results were mixed. Government-branded counter-messaging often faced a source-credibility problem: audiences suspicious of U.S. foreign policy were unlikely to accept U.S. statements simply because they contained accurate facts. Research on counter-ISIS narratives suggests some messages can reduce support, but effects depend heavily on messenger, audience segment, emotional content, and perceived authenticity. The more durable lesson was to distinguish “counter-message” from “counter-network”: weakening recruitment may require community resilience, credible defectors, platform enforcement, disruption of financing and territorial claims, and policy outcomes that undercut the movement’s narrative.

6.6 Russia, Ukraine, and the Networked Information Environment

From the 2000s onward, Russian information activity has been widely studied as a combination of state media, diplomatic messaging, intelligence-linked influence, covert online personas, proxy outlets, hack-and-leak operations, and opportunistic amplification of existing social divisions. A recurring analytical insight is that some campaigns aim less to make audiences believe one consistent doctrine than to increase cynicism, confusion, polarization, or distrust of institutions. This shifts counter-propaganda from “prove this one claim false” toward attribution, network mapping, pre-bunking, rapid evidence release, and reinforcement of trusted local institutions.

The 2014 seizure of Crimea and the full-scale invasion of Ukraine in 2022 accelerated Western and Ukrainian counter-disinformation practices. Ukraine, EU institutions, NATO, independent investigators, journalists, and open-source communities have used geolocation, satellite imagery, forensic verification, intercepted or captured materials, platform analysis, and rapid debunking. Since 2025–2026, NATO and EU work increasingly emphasizes attribution frameworks and Foreign Information Manipulation and Interference (FIMI), reflecting a move from content-only debunking toward evidence about actors, infrastructure, behavior, and legal thresholds.

6.7 China’s International Propaganda and External Communication

The People’s Republic of China has expanded international broadcasting, state-media distribution, diplomatic social-media activity, partnerships, content-sharing agreements, and efforts to shape global narratives about governance, development, human rights, Taiwan, Xinjiang, Hong Kong, and China’s

international role. The slogan commonly translated as “tell China’s story well” captures an official emphasis on narrative power and international communication. Independent research has also examined censorship, pressure on information intermediaries, coordinated amplification, and the blending of overt public diplomacy with more manipulative practices. Counter-response therefore requires differentiating legitimate state advocacy from covert or deceptive influence rather than treating all Chinese communication as disinformation.

6.8 Social Media, Generative AI, and the Next Stage

Social media changed the speed and structure of propaganda by making distribution endogenous to users and algorithms. A message no longer needs a centralized broadcaster: anonymous accounts, influencers, meme pages, recommendation systems, coordinated networks, advertising tools, and microtargeting can create fragmented persuasion at scale. False news has been shown in major platform datasets to diffuse farther and faster than true news in some contexts, although this is not a universal law for all platforms and topics.

Generative AI further lowers the cost of producing persuasive text, images, audio, and video. It enables rapid localization, persona creation, A/B testing, synthetic testimonials, fake experts, conversational persuasion, and multilingual adaptation. At the same time, AI can assist verification, provenance analysis, clustering, translation, and personalized corrections. The strategic problem is therefore not “AI versus truth” but an arms race in generation, detection, persuasion, attribution, and trust.

7. Comparative Systems: Major States and Regimes

The table below compares broad historical patterns. It does not imply moral equivalence among regimes; the key analytical distinction is whether information power is constrained by pluralism, law, independent media, and public contestation.

Actor	Historical/system features	Strengths as an information system	Limits / democratic concerns
United States	CPI (WWI); OWI (WWII); VOA; RFE/RL funding; USIA (1953–1999); military PSYOP/MISO; post-9/11 counterterrorism programs; GEC (2016–2024).	High global reach; strong private media and research ecosystem; capacity for open-source exposure and partnerships.	Credibility harmed when policy and messaging diverge; history of covert propaganda; recurring domestic free-speech and politicization controversies.
United Kingdom	WWI propaganda bodies; BBC external broadcasting; Political Warfare Executive (WWII); diplomatic and security communication.	Early institutional expertise; broadcasting credibility; sophisticated audience analysis.	WWII black propaganda raises ethical questions; state-security secrecy can conflict with transparency.
USSR / Russia	Agitprop, international broadcasting, party networks, front groups, active measures; later state media, intelligence-linked influence, proxies, online networks.	Centralized coordination; long experience with covert influence and narrative exploitation.	Credibility problems in pluralistic environments; exposed covert activity can trigger sanctions and reputational costs; heavy domestic control.
China / PRC	CCP propaganda system; state media; United Front-related influence; diplomatic communication; international content partnerships; digital platforms and external messaging.	Scale, resources, integrated party-state communication, long-term narrative strategy.	Domestic censorship and political control weaken claims of open information; foreign audiences vary greatly in receptivity; attribution can be contested.
Nazi Germany	Propaganda Ministry, party propaganda offices, press	High domestic penetration under dictatorship; integration of	Dependent on repression and monopoly; deeply tied to racist

Actor	Historical/system features	Strengths as an information system	Limits / democratic concerns
	directives, radio, film, rallies, censorship, deception.	modern media and political spectacle.	ideology, war, persecution, and genocide; catastrophic collapse of legitimacy with defeat.
North Korea	Party-state monopoly over domestic media, ideological education, censorship, controlled external broadcasting, restrictions on foreign information.	Extremely high internal control where alternative information is blocked.	Low international credibility; information monopoly depends on coercion and isolation; contradicts free-expression and citizen-autonomy standards.

7.1 United States: A Plural System with Internal Tensions

The U.S. case is best understood as a federation of institutions rather than a single propaganda ministry. Wartime bodies such as the CPI and OWI were temporary. The postwar period created lasting international broadcasting and public-diplomacy institutions, while the military maintained psychological-operations capabilities. The U.S. Information Agency existed from 1953 to 1999. In the 2010s the State Department’s Global Engagement Center coordinated efforts against foreign propaganda and disinformation; its statutory authorization ended in December 2024. Subsequent structures and policy frameworks have changed, illustrating how politically contested this field remains in the United States.

The comparative advantage of the U.S. system is not central control but an ecosystem: independent press, universities, think tanks, civil society, private technology firms, courts, congressional oversight, and open debate can collectively investigate foreign influence. The disadvantage is fragmentation and partisan distrust. A government message can be more easily challenged—which is democratically healthy—but also more easily dismissed when institutional trust is low.

7.2 United Kingdom: Broadcasting Credibility and Wartime Political Warfare

Britain’s two most influential traditions are credible international broadcasting and specialized wartime political warfare. The BBC developed global recognition partly because it could report unfavorable news and maintain professional standards, while wartime agencies also used covert deception. The combination demonstrates that “British counter-propaganda” was not a single ethical model. Overt journalism, public diplomacy, gray propaganda, and black propaganda coexisted under different legal and wartime conditions.

7.3 Soviet Union and Russia: From Ideological Propaganda to Adaptive Influence

The Soviet model combined ideological consistency in official doctrine with covert tactical flexibility. Foreign Communist parties, front organizations, friendly media, diplomacy, and intelligence activity could reinforce one another. Contemporary Russian practice is often described as more eclectic and less doctrinal: narratives may be contradictory but convergent in their effect of undermining trust, amplifying grievances, or weakening adversary cohesion. That interpretation is influential but should not become a universal explanation; campaign objectives vary by target, event, and sponsor.

7.4 China: Narrative Power, State Media, and Party-State Coordination

China’s international information system operates through overt state media, diplomatic channels, commercial distribution, cultural and educational networks, and a wider party-state influence apparatus.

Independent research emphasizes both persuasion and information control. Counter-propaganda toward China is most credible when it avoids civilizational or ethnic framing and instead focuses on verifiable conduct: source disclosure, media ownership, hidden sponsorship, censorship, coordinated behavior, document authenticity, and empirical claims.

7.5 Nazi Germany and North Korea: Information Control as Regime Structure

Nazi Germany and North Korea demonstrate why propaganda and counter-propaganda cannot be evaluated only by message effectiveness. In authoritarian or totalitarian systems, the state can increase apparent persuasive success by suppressing competing speech, punishing dissent, restricting foreign information, and controlling institutions of socialization. The relevant measure is therefore not merely “did the audience repeat the message?” but whether citizens had meaningful alternatives, access to evidence, and freedom to reject the official narrative.

8. Major Thinkers: Propaganda, Persuasion, Cognition, and Social Influence

Scholar	Core contribution	Implication for counter-propaganda
Jacques Ellul	Propaganda as a social and technological system that integrates individuals into organized action, not merely isolated lies.	Counter-propaganda must address the social environment and need for belonging, not only false facts.
Harold Lasswell	Communication can be analyzed by “who says what, in which channel, to whom, with what effect”; propaganda is tied to symbols, power, and wartime mobilization.	Provides a basic architecture for source-message-channel-audience-effect analysis.
Edward Bernays	Public relations and “engineering consent”; intermediaries and symbolic association shape mass opinion.	Highlights opinion leaders, third-party validation, agenda creation, and the ethical problem of invisible persuasion.
Walter Lippmann	Citizens act on “pictures in our heads,” stereotypes, and mediated pseudo-environments.	Counter-propaganda must correct the mental model and context, not just a data point.
Paul Lazarsfeld	Two-step flow and importance of interpersonal influence and opinion leaders.	Trusted peers and local influencers may matter more than mass official messaging.
Carl Hovland	Experimental study of source credibility, message order, audience factors, and persuasion.	Makes counter-propaganda an empirical design problem; credibility and audience predispositions are variables.
Leon Festinger	Cognitive dissonance: people are motivated to reduce inconsistency among beliefs, identity, and behavior.	Corrections that create dissonance can motivate updating—or rationalization—depending on available pathways.
William McGuire	Inoculation theory: exposure to weakened attacks plus refutation can build resistance to stronger persuasion.	Foundational theory for pre-bunking and manipulation-technique training.
Daniel Kahneman	Dual-process perspective, heuristics, salience, availability, and judgment under uncertainty.	Explains why vivid, simple, repeated narratives can beat complex corrections unless response design reduces cognitive burden.
Amos Tversky	Framing, heuristics, and decision biases; equivalent information can produce different choices depending on presentation.	Shows that changing a frame can change judgment even when facts remain constant.

These theorists come from different intellectual traditions and should not be collapsed into one model. Ellul emphasizes social integration; Lasswell emphasizes power and communication structure; Hovland

and McGuire emphasize experimental persuasion; Festinger, Kahneman, and Tversky illuminate cognitive processes. Taken together, they explain why counter-propaganda must simultaneously analyze institutions, messages, identity, heuristics, and social networks.

9. Units of Analysis and Variables of Effectiveness

Level	Key variables	Diagnostic question
Message	Accuracy, clarity, vividness, emotional intensity, narrative coherence, complexity, repetition, timing.	Does the content replace the false model or merely negate it?
Source	Expertise, trustworthiness, identity, transparency, track record, perceived independence.	Will the target audience accept evidence from this messenger?
Channel	Reach, accessibility, censorship resistance, format, algorithmic treatment, intimacy.	Is the channel actually used by the target population?
Audience	Prior belief, identity, grievance, media habits, literacy, age, language, risk perception.	What need or motive makes the propaganda attractive?
Frame	Problem definition, causal attribution, moral evaluation, proposed remedy, omitted context.	What question is the propaganda teaching the audience to ask?
Narrative	Characters, plot, victim/villain/hero roles, temporal arc, identity, meaning, future promise.	What larger story makes individual claims feel true?
Network	Community structure, influencers, bridges, bots, coordinated actors, homophily, cascade dynamics.	Who seeds, validates, and amplifies the message?
Algorithm	Ranking, recommendation, engagement incentives, ad targeting, personalization, moderation.	Which platform mechanisms increase exposure and repetition?
Institution	Transparency, competence, accountability, media independence, legal safeguards, correction practices.	Does the information environment generate durable trust or recurring vulnerability?

9.1 Cross-Level Interaction

Effectiveness rarely belongs to one level. A perfect fact-check can fail if the source is distrusted, the channel is not used, the audience sees it after identity has hardened, or the platform algorithm rewards the sensational falsehood. Conversely, a mediocre message from a trusted community figure can outperform a technically superior government release. Evaluation should therefore model interactions: message × source × audience × timing × network × institutional context.

10. Digital Transformation: Bots, Trolls, Deepfakes, Algorithms, Influencers, Microtargeting, and AI

10.1 Bots and Coordinated Accounts

Bots can automate posting, reposting, reply behavior, hashtag activity, and apparent consensus. The strategic effect is not only volume; it can manufacture social proof, manipulate trending systems, harass journalists, and create the impression that an opinion is spontaneous and widespread. Countermeasures should focus on coordinated behavior and provenance rather than ideology: network analysis, timing

patterns, shared infrastructure, account creation bursts, and repeated asset reuse can provide stronger evidence than content similarity alone.

10.2 Trolls and Human-Led Influence

Human operators are often more adaptive than bots. They can enter communities, imitate local language, provoke conflict, seed rumors, or exploit real grievances. “Troll” is therefore not a sufficient analytical category. Investigators should distinguish paid operators, ideological volunteers, opportunistic influencers, coordinated political activists, and ordinary users who unknowingly amplify a campaign.

10.3 Deepfakes and Synthetic Media

Deepfakes can fabricate apparent audiovisual evidence, but their impact is not limited to successful deception. The existence of deepfakes creates a “liar’s dividend”: authentic recordings can be dismissed as synthetic. Research increasingly shows that warnings and labels help but do not eliminate influence. Counter-propaganda therefore needs provenance systems, rapid forensic capability, trusted archival copies, chain-of-custody procedures, and public communication that explains uncertainty without creating generalized cynicism.

10.4 Algorithmic Recommendation

Recommender systems determine which content becomes salient, repeated, and socially validated. The algorithmic unit of analysis asks not only whether a claim is false but whether the platform’s objective function disproportionately rewards outrage, novelty, watch time, or polarizing interaction. Counter-response can include friction, downranking, reduced monetization, recommendation limits, context panels, or design changes. These measures raise legitimate questions about transparency, due process, and viewpoint discrimination; the defensible focus is on clearly defined behavior, safety risks, and platform rules rather than secret political blacklists.

10.5 Memes and Influencers

Memes compress complex political meaning into emotionally legible symbols. Influencers add parasocial trust and social identity. A government white paper may be factually stronger but socially weaker than a meme repeated by a culturally central creator. Effective counter-propaganda therefore needs translation across formats without sacrificing truthfulness: visual evidence, short explanatory video, community Q&A, trusted experts, and creators who can speak in the audience’s idiom.

10.6 Anonymous Accounts and Microtargeting

Anonymity can protect dissidents and vulnerable speakers, so anonymous speech is not inherently suspicious. The risk arises when anonymity is combined with deceptive persona construction, hidden sponsorship, coordinated networks, or targeted manipulation. Microtargeting can exploit distinct fears and identities with messages that are invisible to the broader public, reducing accountability. Countermeasures include ad libraries, sponsor disclosure, research access, and limits on sensitive targeting rather than blanket elimination of anonymous participation.

10.7 Generative AI

Generative AI changes propaganda economics. It reduces the marginal cost of producing thousands of localized messages, fake comments, synthetic images, personalized arguments, or conversational agents. Recent experiments show that AI-generated or AI-assisted persuasion can be highly effective under some conditions, particularly when systems are personalized or optimized for rhetoric. Other studies show that people can sometimes detect low-quality synthetic content and that labeling alone does not necessarily remove persuasive effects. The strategic implication is that authenticity signals, provenance, platform design, and trusted relationships become more important as content abundance increases.

AI also strengthens defense. Models can cluster narratives, translate foreign-language propaganda, detect coordinated patterns, compare claims against structured databases, generate draft rebuttals, and help analysts prioritize emerging stories. But automated counter-propaganda can itself hallucinate, overclaim attribution, or generate politically biased responses at scale. Human review, evidence traceability, red-teaming, and clear accountability are therefore essential.

COUNTER-PROPAGANDA: CONCEPTS, STRATEGIES, HISTORY, AND PRACTICE

A Comprehensive Framework for Understanding
and Building Resilient Information Ecosystems

Authors:

The American Newspaper <https://americannewspaper.org>
AmericanTV <https://americantv.org>



TABLE OF CONTENTS

1. Definitions and Conceptual Origins
2. Key Concepts: Boundaries and Distinctions
3. Objectives of Counter-Propaganda
4. Forms of Counter-Propaganda
5. Psychological Mechanisms that Influence Effectiveness
6. Historical Development and Case Studies
7. Major States and Regimes: Approaches and Systems
8. Theories from Communication and Social Psychology
9. Analytical Levels and Variables of Effectiveness
10. Digital Era: New Actors, Tools, and Transformations
11. Law, Ethics, and Democratic Safeguards
12. Measuring Effectiveness
13. Successful and Failed Campaigns
14. Integrated Model: From Propaganda to Resilience
15. Conclusion and Key Recommendations
- References and Further Reading

1. DEFINITIONS AND CONCEPTUAL ORIGINS

對抗宣傳 (Duìkàng Xuānhuán)

“對抗宣傳”是指有意识地針對敵對或
不友善勢力的宣傳活動，採取系統性的
資訊、心理與溝通手段，以削弱、抵消
或抵禦其宣傳影響，並維護自身利益、
價值與社會穩定。

Counter-propaganda

is the systematic use of information,
communication, and psychological
means to neutralize, undermine, or
counteract the propaganda efforts of
an adversary while advancing one's
own legitimate interests, values, and
social stability.

Conceptual Origins

- Rooted in early 20th-century information warfare and wartime communication.
- Emerged from the need to respond to enemy propaganda in WWII.
- Developed systematically in WWII and the Cold War.
- In the digital age, expanded to include networked, psychological, and cognitive domains.

Why "Refuting Propaganda" is Not Enough

Counter-propaganda is more than
simple refutation. It includes proactive,
strategic, and multi-dimensional efforts
to shape perceptions, protect audiences,
and build resilience.

3. PRINCIPAL OBJECTIVES OF COUNTER-PROPAGANDA



- ✓ Undermine the credibility of hostile propaganda and sources.
- ✓ Expose false information and reveal manipulative tactics.
- ✓ Discredit false frames and narratives.
- ✓ Verify and enhance the credibility of information sources.
- ✓ Build cognitive immunity and resilience among audiences.
- ✓ Provide accurate, timely, and compelling alternative narratives.
- ✓ Protect social trust, cohesion, and institutional legitimacy.
- ✓ Reduce the spread and impact of disinformation.
- ✓ Support informed decision-making and democratic values.

4. FORMS OF COUNTER-PROPAGANDA: ADVANTAGES, LIMITATIONS, AND RISKS

Form	Description	Advantages	Limitations / Conditions	Risks / Unintended Effects
Direct Refutation	Explicitly refute false claims.	Clear, straightforward, effective with undecided audiences.	May not reach committed believers.	Familiarity effect, reinforcing the myth.
Indirect Refutation	Rebate implications without repeating the false claims.	Reduces repetition of falsehoods.	Requires skill and careful framing.	May be seen as vague or evasive.
Fact-checking	Verify claims against evidence.	Builds credibility, supports truth.	Slow, reactive, limited reach.	Perceived bias, information overload.
Source Exposure	Expose the source's bias, motives, or hidden agendas.	Undermines trust in hostile sources.	Needs solid evidence; preaching to choir.	Slander claims, credibility attacks.
Logical Contradictions	Complex internal contradictions or illogical claims.	Appeals to reason, effective with open minds.	Complex arguments can confuse.	Seen as overly technical.
Reframing	Change the frame through which the issue is viewed.	Can shift perceptions and priorities.	Requires deep understanding of audience.	Frame may be rejected as manipulation.
Alternative Narratives	Offer compelling new stories and interpretations.	Fills the narrative vacuum.	Hard to compete with emotional myths.	May be ignored or attacked.
Pre-bunking / Inoculation	Expose tactics in advance to build resistance.	Prevents belief, builds resilience.	Requires education and sustained effort.	Can increase awareness of myths.
Humor / Satire	Use humor to ridicule falsehoods.	Memorable, shareable, lowers resistance.	May offend, misinterpreted.	Backfire, trivialize serious issues.
Media / Platform Responses	Labeling, downranking, removal, algorithmic interventions.	Reduces reach of harmful content.	Free speech concerns, platform inconsistency.	Overreach, censorship accusations.
Education / Media Literacy	Teach critical thinking and media skills.	Long-term, builds societal resilience.	Slow impact, needs resources.	Underfunded, inequitable access.

No single form is sufficient. The best strategies combine multiple approaches tailored to context and audience.

6. HISTORICAL DEVELOPMENT OF COUNTER-PROPAGANDA

WWI (1914-1918)	WWII (1939-1945)	Cold War (1947-1991)	Korean & Vietnam Wars (1950s-1970s)	Post-Cold War (1990s-2000s)	Post-9/11 (2001-2010s)	2010s-Present: Information Age
• Committee on Public Information (CPI) • Posters, leaflets, newspapers • Early information control and censorship	• Extensive use of leaflets, radio broadcasts • BBC, Voice of America • Espionage, enemy attacks • War bonds, morale	• U.S.-Soviet propaganda competition • IFBUE, VOX • Cultural diplomacy, exchanges	• PSYOP leaflets, radio • Hearts and Minds • Credibility challenges	• Globalization of media • Internet begins • Narrative trust • Rise of social media and digital outreach	• Counterterrorism communication • Strategic messaging to counter extremism • Rise of social media and digital outreach	• Social media disinformation • Counter-disinformation • Russian info ops, IS propaganda • Ukraine War: China's influence • AI, deepfakes, generative AI

8. SCHOLARLY THEORIES THAT INFORM COUNTER-PROPAGANDA

Propaganda is the conscious manipulation of mass thought through direct communication.	"The way in which the world is seen is what counts." Direct communication matters.	Public relations pioneer. Propaganda operates through engineering consent.	"Pseudo-environment." People rely on stereotypes and media images to make sense of the world.	Two-step flow model. Opinion leaders mediate mass communication.
Researcher on persuasion, group conformity, group decision making.	Cognitive Dissonance Theory. People reduce psychological discomfort.	Attitude Change Theory. Focus on self-attitudes and self-referent beliefs.	System 1 (fast) vs. System 2 (slow) thinking. Biases shape judgment.	Heuristics and biases. People rely on mental shortcuts to evaluate information and risk.

These theories explain why some counter-propaganda works, and why audiences accept, resist, or reject messages.

11. LAW, ETHICS, AND DEMOCRATIC SAFEGUARDS

Democratic Counter-Propaganda Should Be:

- ✓ Transparent about sources and motives
- ✓ Factually accurate and evidence-based
- ✓ Proportionate and necessary
- ✓ Subject to legal oversight
- ✓ Respect freedom of expression and press freedom
- ✓ Enable independent verification
- ✓ Protect citizen autonomy and privacy

Avoid or Resist:

- ✗ Censorship and prior restraint
- ✗ Coercive or deceptive manipulation
- ✗ Targeting lawful dissent
- ✗ Excessive surveillance and profiling
- ✗ Weaponizing disinformation claims

12. MEASURING EFFECTIVENESS: BEYOND VANITY METRICS

Meaningful Indicators

- ① Reach among target audiences
- ② Changes in credibility of sources
- ③ Attitude and belief change
- ④ Behavioral change (offline/online)
- ⑤ Rates of disinformation diffusion
- ⑥ Network propagation patterns
- ⑦ Engagement quality and depth
- ⑧ Resilience and long-term immunity

Why Views/Impressions Fail

High exposure does not equal persuasion. People can view, but not believe. Content, trust, and cognitive processing matter most.

13. CASE STUDIES: SUCCESS AND FAILURE

Successful Examples

- BBC World Service (Cont. War): Credible, multiplatform, long-term investment in trust.
- "I Am an American" (Leifur Eiriksson): Effective in building trust and credibility.
- Smallpox Immunization Campaign (Ethiopia, 1970s): Used local trust, religious endorsement, and community health workers.
- Ukraine Government Strategic Communication (2022-23): Rapid mobilization, consistent messaging, global outreach.

Failed / Unintended Examples

- U.S. "We Will Buy You" Leaflets (Vietnam War): Threatening messages perceived as arrogant, limited effectiveness.
- Soviet Disinfo After Chernobyl (1986): Denial of the truth, damaged credibility.
- Irish RINOs Claims (2003): Misleading claims damaged trust in U.S. and allies.
- Disinformation Against COVID-19 Vaccines: Misled messages and polarization undermined public health efforts.

Success depends on credibility, timing, audience understanding, message quality, channel choice, and long-term consistency.

2. KEY CONCEPTS: BOUNDARIES AND DISTINCTIONS

Concept	Definition	Focus	Typical Actors
Propaganda	Systematic efforts to shape perceptions and influence behavior to advance a cause.	Promote a cause, shape beliefs	States, groups, organizations
Counter-propaganda	Strategic efforts to neutralize or undermine adversary propaganda and protect audiences.	Undermine hostile propaganda, protect	States, orgs., civil society
Counter-messaging	Specific messages designed to rebut or contrast false or harmful claims.	Deliver alternative or corrective messages	Gov't, NGOs, communities
Strategic Communication	Coordinated use of communication to achieve policy objectives.	Inform, persuade, influence audiences	Governments, institutions
Public Diplomacy	Engagement with foreign publics to build understanding and trust.	Build relationships, enhance reputation	Governments, cultural orgs.
PSYOP / MISO (Psychological Ops / Military Information Support Ops)	Military operations to convey selected information and indicators to influence emotions, motives, reasoning.	Influence enemy behavior, protect own forces	Military, defense organizations
Information Operations (IO)	Integrated employment of information-related capabilities.	Affect decision-making of adversaries	Military, intel agencies
Information Warfare	Conflict in the information domain to gain advantage and cause harm.	Degrade, disrupt, or exploit info systems	States, militaries, non-state actors
Influence Operations	Covert or overt efforts to shape attitudes, opinions, and behaviors.	Long-term influence and manipulation	States, groups, networks
Cognitive Warfare	Operations to influence cognition and perception at scale.	Change how people think and decide	States, groups, credentialed actors
Fact-checking	Verification of factual claims against evidence.	Establish factual accuracy	Media, orgs., platforms
Debunking	Expose and explain false claims after they appear.	Reduce belief in false claims	Media, fact-checkers, experts
Pre-bunking (Inoculation)	Prepare audiences in advance by exposing tactics and building resistance.	Prevent belief in false claims	Educators, gov't, platforms
Media Literacy	Skills to access, analyze, evaluate, and create media.	Empower critical consumption	Educators, NGOs, schools
Counter-disinformation	Broad efforts to address false information, including prevention, detection, response.	Reduce harm from disinformation	Multi-stakeholder efforts

5. PSYCHOLOGICAL MECHANISMS THAT INFLUENCE EFFECTIVENESS

Repeated exposure to a claim can increase belief, even if it is repeatedly refuted.	Refutations may strengthen false beliefs among highly committed individuals.	People accept information that aligns with their goals and reject contrary evidence.
Favoring information that confirms existing beliefs.	Rejecting information that threatens group identity or worldview.	Choosing information that aligns with one's beliefs, avoiding contradictory content.
Information environments where dissenting views are absent.	Like-minded groups adopt more extreme positions over time.	Messages from trusted sources are more persuasive.
	Understanding these mechanisms is essential to design counter-propaganda that informs without reinforcing, protects identity, and reaches the right audiences with the right messages through trusted channels.	
People look to others to decide what is true or acceptable.		

7. MAJOR STATES AND REGIMES: APPROACHES AND SYSTEMS

United States	United Kingdom	Soviet Union / Russia	China	Nazi Germany	North Korea
• Open society model • Transparency and free press • VOA, cultural diplomacy • Counterterrorism messaging • Challenges partisan polarization, credibility gaps	• BBC World Service • Strong tradition of public broadcasting • Counter-disinformation units (e.g., GC&D, JTBG Disinfo Unit) • Emphasis on credibility	• KGB active measures • Disinformation, deception (maskirovka) • RT, Sputnik • Today, info ops, cyber, social media manipulation • Emphasis on credibility	• Global media expansion (CGTN, Xinhua) • United Front Work • Narrative control and censorship • "Three Warlines": public opinion, psychological, legal	• Joseph Goebbels' Ministry of Propaganda • Total control of media and culture • Extensive domestic and foreign propaganda	• Total information control • Propaganda cult • Total control of media and culture • External disinfo and cyber ops • Very limited info

Evaluation based on historical evidence and scholarly research, not government claims.

9. ANALYTICAL LEVELS AND VARIABLES OF EFFECTIVENESS

Level	Key Variables
Message	Clarity, accuracy, emotional appeal, framing, consistency, specificity.
Source	Credibility, expertise, trustworthiness, transparency, motives.
Channel	Reach, accessibility, platform norms, algorithmic treatment.
Audience	Demographics, beliefs, values, media habits, identity, literacy.
Frame	Interpretive context, values, metaphors, causal attributions.
Narrative	Coherence, characters, conflict, resolution, moral logic.
Network	Information flows, influencers, community structures, beliefs.
Algorithm	Recommendation systems, ranking signals, amplification biases.
Institution	Policies, transparency, accountability, resources, legitimacy.

10. DIGITAL ERA: NEW ACTORS, TOOLS, AND TRANSFORMATIONS

Artificial accounts amplify or distort narratives, create false consensus, harass, and intimidate.	Realistic but fabricated audio, video, and images undermine trust and sow confusion.	Optimize for engagement, which can amplify polarization and disinformation.	Serialized, shareable content spreads rapidly, often detached from context.	Personalized, trustworthy-seeming and niche communities.	Targeted messages exploit psychological vulnerabilities.	Accelerate production of text, images, and videos; increase information volume and complexity.

These tools lower the barrier to propaganda and counter-propaganda alike, but also increase risks of manipulation, confusion, and information overload.

Key Challenge: Speed, scale, and anonymity reduce human verification and traditional institutions.

14. INTEGRATED MODEL: PROPAGANDA PROCESS AND COUNTER-PROPAGANDA DEFENSE



Counter-propaganda (disinformation) is not a neutral act. It is the construction of a resilient, long-term information ecosystem grounded in:

© 2024 The American Newspaper & AmericanTV

Figure 2. Generated infographic-style synthesis of the information environment. Included as an illustrative visual aid; specific claims should be checked against the cited research in the text and bibliography.

11. Democracy, Law, Human Rights, Journalism, and the Risk of State Abuse

11.1 The Democratic Paradox

A democracy must be capable of defending elections, public safety, national security, and citizens against covert foreign manipulation. Yet the tools required—intelligence collection, attribution, strategic communication, platform contact, content labeling, and public warnings—can be repurposed for domestic political advantage. Counter-propaganda therefore requires constitutional and procedural safeguards precisely because some threats are real.

11.2 International Human Rights Framework

Article 19 of the International Covenant on Civil and Political Rights protects the right to hold opinions without interference and the freedom to seek, receive, and impart information and ideas. Restrictions on expression must satisfy legal requirements and, under the Covenant’s structure and Human Rights Committee guidance, should be necessary and proportionate for specified legitimate aims. Article 20 separately requires prohibition by law of propaganda for war and advocacy of national, racial, or religious hatred that constitutes incitement to discrimination, hostility, or violence. These provisions create neither an unlimited “anti-disinformation” power nor an absolute bar on regulation.

European human-rights law similarly protects political expression and information while recognizing narrowly tailored restrictions. The core democratic principle is that falsehood alone is not a sufficient basis for unrestricted state censorship. Laws targeting manipulation should define harms and procedures carefully, preserve judicial review, and avoid criminalizing ordinary error, satire, dissent, investigative journalism, or disagreement with official narratives.

11.3 Domestic Law: General Principles

Domestic rules vary sharply. In the United States, government efforts encounter strong First Amendment constraints, especially when they involve domestic speech, coercion of private platforms, viewpoint discrimination, or retaliation against journalists. Military and foreign-information authorities operate under separate statutes and policies. Other democracies may allow broader regulation of hate speech, election disinformation, foreign interference, or platform duties, but remain constrained by constitutional and human-rights norms. Any comparative analysis must therefore distinguish foreign-targeted state communication, domestic public information, intelligence activity, law enforcement, and private platform moderation.

11.4 Platform Rules

Platforms are not states, but their rules shape public discourse. They commonly regulate coordinated inauthentic behavior, impersonation, manipulated media, harassment, terrorist content, spam, and certain forms of election interference. The legitimacy of platform intervention improves when rules are public,

consistently applied, accompanied by notice and appeal, and supported by transparency reporting. Government requests to platforms should be documented where possible and subject to legal oversight so that informal pressure does not become an unreviewable censorship system.

11.5 Journalistic Ethics

Journalism is not counter-propaganda in the governmental sense, even when it exposes propaganda. Its distinct legitimacy comes from independent verification, source criticism, correction, separation of news from advocacy, protection of confidential sources, and willingness to scrutinize all sides—including the government conducting the counter-propaganda campaign. Journalists should avoid becoming passive conduits for official “debunking,” intelligence claims, leaked attribution, or platform labels. The correct posture is verification, not enlistment.

Criterion	Democratic counter-propaganda	Authoritarian / problematic information control
Transparency	Disclose official authorship, funding, evidence basis, and material conflicts when feasible.	Hidden state sponsorship, undisclosed proxies, astroturfing.
Factual accuracy	Use auditable evidence; distinguish facts, estimates, and judgments.	Knowingly false counter-messaging or fabricated evidence.
Source disclosure	Explain provenance and confidence while protecting legitimate intelligence methods.	Anonymous official assertion presented as certainty without basis.
Independent verification	Enable journalists, courts, researchers, and civil society to check key claims.	State monopoly on classification of “truth.”
Proportionality	Match intervention to demonstrated harm and reach.	Mass suppression for low-level or speculative threats.
Legal authority	Use clear statutory or constitutional authority; maintain review and appeal.	Informal coercion outside legal process.
Press freedom	Allow reporting critical of the counter-propaganda campaign and the government.	Punishing or discrediting independent media merely for dissent.
Citizen autonomy	Provide information and tools that improve choice.	Manipulation, coercion, forced ideological conformity.
Contestability	Permit correction of official errors and public disagreement.	Official narratives insulated from challenge.

Operational test: If a “counter-disinformation” program cannot tolerate independent journalists checking its evidence, courts reviewing its authority, citizens criticizing its conclusions, or researchers documenting its errors, it is drifting from democratic resilience toward information control.

12. Evaluating Effectiveness

Views, impressions, followers, reposts, and media mentions measure exposure, not persuasion. A campaign can go viral because audiences mock it; a low-reach correction can still be strategically important if it reaches journalists, election administrators, commanders, or a vulnerable community. Evaluation should therefore separate output, exposure, cognition, attitude, behavior, diffusion, and resilience.

Indicator	What to measure	Interpretive question
Reach / exposure	Unique target-audience exposure, frequency, channel penetration.	Did the intended audience actually encounter the response?
Credibility change	Trust in the hostile source, trusted source, or institution before/after.	Did source evaluation change?

Indicator	What to measure	Interpretive question
Factual belief	Accuracy or confidence about specific claims.	Did people update what they believe happened?
Frame acceptance	Changes in causal attribution, problem definition, or moral interpretation.	Did the underlying interpretive structure change?
Attitude change	Support/opposition, favorability, perceived legitimacy.	Did evaluation of actors or policies change?
Behavior change	Voting intention, compliance, reporting, sharing, recruitment, defection, donation, protest.	Did communication alter action?
Diffusion rate	Reproduction number of a narrative, cascade depth, speed, cross-platform spread.	Did the false narrative spread less after intervention?
Network change	Removal of coordinated nodes, reduced bridge activity, fragmentation of amplification network.	Was the distribution system disrupted?
Source resilience	Ability of trusted institutions to retain credibility under repeated attack.	Is trust stable through future incidents?
Cognitive resilience	Ability to recognize manipulation techniques and resist novel variants.	Did learning transfer beyond the single claim?
Institutional resilience	Speed, transparency, coordination, correction capacity, legal oversight.	Can the system respond repeatedly without eroding legitimacy?

12.1 Experimental and Observational Designs

- Randomized experiments: compare exposure to corrections, pre-bunks, alternative narratives, or source labels; measure factual belief, trust, and behavior.
- Panel studies: follow the same participants over time to assess durability and decay.
- A/B platform tests: estimate changes in sharing, click-through, or exposure while monitoring unintended distribution effects.
- Network analysis: compare cascade speed, centrality, community penetration, and cross-platform propagation before and after intervention.
- Interrupted time series: evaluate whether major warnings, sanctions, takedowns, or evidence releases change diffusion patterns.
- Qualitative audience research: interviews, focus groups, ethnography, and community observation reveal meaning and trust mechanisms not visible in clicks.
- Operational measures in conflict: surrender rates, defection, intercepted communications, morale indicators, prisoner interviews—use cautiously because causal attribution is difficult.

12.2 Why Simple Exposure Metrics Mislead

Exposure is necessary for persuasion but not sufficient. High view counts may reflect adversarial engagement. Repetition can increase awareness without changing belief. Messages can change beliefs without changing behavior because material incentives dominate. A platform intervention can reduce public posts while pushing communities into encrypted channels. Long-term resilience may produce no spectacular spike in engagement at all. The correct metric depends on the theory of change.

13. Comparative Case Studies: Successes, Failures, and Mixed Outcomes

The cases below use “success” and “failure” in limited, evidence-sensitive terms. In information operations, causal attribution is rarely clean: military outcomes, policy credibility, censorship, economic conditions, and social identity interact with communication.

Case	Objective	Audience	Message logic	Channel	Assessment
British PWE / black radio against Nazi Germany, WWII	Undermine morale and cohesion; exploit regime contradictions.	German audiences; soldiers and civilians.	Imitated insider style, corruption stories, selective truth/deception.	Radio and covert media.	Tactically innovative; effects difficult to isolate; ethically problematic by contemporary transparency standards.
RFE/RL during Cold War	Provide uncensored information and weaken monopoly of communist media.	Eastern Europe and USSR.	News, local reporting, history, culture, dissident voices.	Shortwave radio; later multiple media.	Long-run credibility and access were important; early covert CIA funding complicates transparency narrative.
VOA credibility model	Build attention and respect through reliable news while explaining U.S. policy.	Foreign publics.	News plus clearly identified policy content.	Radio, TV, digital.	Institutional separation and credibility often increased effectiveness; government pressure remained a recurring tension.
U.S. Active Measures Working Group, 1980s	Expose Soviet forgeries and disinformation.	Foreign governments, media, elites, public.	Documented attribution and public exposure.	Reports, diplomacy, media engagement.	Frequently cited as effective because it raised reputational costs and used evidence; still operated within broader Cold War political contest.
U.S. information effort in Iraq	Counter insurgent and adversarial narratives; support legitimacy.	Iraqi and regional audiences.	Official messages, PSYOP, media initiatives.	Broadcast, print, paid media, military channels.	Often judged weak because credibility and policy realities undercut messaging; some covert/paid practices damaged trust.
Counter-ISIS messaging	Reduce recruitment and legitimacy.	Potential recruits, sympathizers, communities.	Atrocity exposure, defectors, religious arguments, alternative identities.	Social media, community networks, official campaigns.	Mixed: targeted counter-narratives can shift attitudes, but government-branded messages often lack authenticity; network disruption and territorial defeat mattered greatly.
Ukraine / open-source verification since 2014, especially 2022–	Rapidly rebut battlefield falsehoods, attribute manipulation, sustain domestic and allied trust.	Ukrainian and international audiences.	Geolocated evidence, satellite imagery, captured documents, rapid official updates, OSINT.	Social media, news, government channels, investigator networks.	Strong in speed and evidentiary culture, but wartime claims still require independent verification; propaganda exists on all sides of armed conflict.
EU / platform pre-bunking campaigns	Increase recognition of manipulation techniques before elections and crises.	Broad digital audiences.	Short inoculation videos about tactics such as scapegoating or false dichotomies.	YouTube and other digital platforms.	Experimental evidence supports reduced susceptibility; durability and transfer remain bounded and require reinforcement.

13.1 Why the Active Measures Working Group Is Often Treated as a Relative Success

The U.S. interagency effort of the 1980s is frequently cited because it shifted from rebutting Soviet accusations in the abstract to documenting specific forgeries, fronts, planted stories, and dissemination chains. The method targeted source credibility and operational tradecraft rather than only the content of the claim. The broader lesson is “show your work”: attribution becomes more persuasive when external audiences can inspect documents and patterns themselves.

13.2 Why Iraq Is Often Treated as a Counter-Propaganda Failure

In Iraq, official messaging frequently confronted a credibility deficit created by violence, political instability, detainee abuse scandals, contested intelligence claims, and the gap between strategic promises and lived experience. Some paid or unattributed media practices also blurred the line between journalism and influence. The failure was therefore not simply insufficient messaging technique. It illustrates a fundamental law of strategic communication: communication cannot indefinitely compensate for visible policy contradictions.

13.3 Counter-ISIS: The Messenger Problem

ISIS propaganda offered identity, belonging, grievance, heroic purpose, religious justification, and visual spectacle. A sterile factual rebuttal did not address those functions. More promising interventions used defectors, survivors, credible religious voices, and community networks, or undermined ISIS’s claim to competence by documenting corruption, brutality, and battlefield losses. Even then, communications effects were intertwined with military defeat, loss of territory, platform enforcement, and changes in recruitment opportunity.

13.4 Ukraine: Speed, Evidence, and the Risk of Wartime Overclaim

The Ukraine case demonstrates the value of open-source evidence ecosystems: journalists, satellite firms, researchers, citizens, and official institutions can rapidly authenticate or challenge battlefield claims. This plural verification can strengthen counter-propaganda because no single institution needs to monopolize truth. At the same time, war creates intense incentives for all belligerents to frame information strategically. The safest analytical rule is to distinguish verified evidence from official wartime assertion even when one side has a documented record of large-scale disinformation.

14. Integrated Propaganda–Counter-Propaganda Model

14.1 Propaganda Process

- 1. Propaganda input: claim, symbol, image, rumor, narrative, fabricated evidence, selective fact, or emotional cue.
- 2. Cognitive and emotional stimulation: attention, fear, anger, grievance, pride, disgust, hope, humiliation, status, uncertainty.
- 3. Frame formation: definition of the problem, causal attribution, moral evaluation, and implied solution.

- 4. Integration with beliefs and identity: the new frame attaches to prior ideology, social identity, memory, and trusted sources.
- 5. Behavior or diffusion: sharing, voting, compliance, recruitment, protest, defection, harassment, withdrawal, or institutional distrust.

14.2 Counter-Propaganda Defense Mechanism

- 1. Detection: identify anomalous narratives, coordinated activity, manipulated media, or high-risk rumors before they mature.
- 2. Source verification: establish provenance, sponsorship, authenticity, ownership, and confidence level.
- 3. Fact-checking: test discrete factual propositions against transparent evidence.
- 4. Frame dismantling: reveal omissions, false causal links, manipulative binaries, scapegoating, and emotionally loaded premises.
- 5. Contextualization: place accurate facts in chronology, scale, comparison, uncertainty, and institutional context.
- 6. Alternative explanation: provide a coherent causal account that answers the audience's real question and preserves explanatory closure.
- 7. Pre-bunking / inoculation: teach manipulation techniques and likely narrative patterns before the next exposure.
- 8. Strengthening social resilience: build trustworthy institutions, independent media, civic networks, transparency, and durable media literacy.

Stage 1		Stage 2		Stage 3		Stage 4		Stage 5
Propaganda	→	Cognitive / emotional stimulation	→	Frame formation	→	Belief & identity integration	→	Behavior / diffusion
Counter-defense	→	Detection + source verification	→	Fact-checking + frame dismantling	→	Context + alternative explanation	→	Pre-bunking + resilience

The model is recursive rather than linear. Propaganda actors adapt after exposure; audiences feed back through engagement; platforms change ranking; states alter policy; and counter-propaganda can itself become propaganda if it abandons evidence and transparency. Resilience therefore requires learning institutions that continuously update their assumptions.



1 Definition and Conceptual Origins

對抗宣傳 (Duìkàng xuānchuán):
 對抗 (duìkàng) means "resistance," "counteraction," or "opposition."
 宣傳 (xuānchuán) means "propaganda," "publicity," or "public communication."
 Together, 對抗宣傳 refers to the systematic effort to resist, neutralize, or counteract hostile propaganda and harmful influence.

English term counter-propaganda:
 emerged during World War I, when states sought to counter enemy wartime propaganda. It developed through the interwar period, expanded enormously in World War II and the Cold War, and has evolved in the digital and AI era.

Why "merely refuting" is inadequate

✗ It is reactive, not proactive.

- It ignores frames, emotions, identity, and underlying narratives.
- It may inadvertently reinforce false claims (familiarity effect).
- It overlooks source credibility, audience values, and channels.
- It fails to build long-term resilience and alternative narratives.
- It can become part of the adversary's information strategy.

3 Principal Objectives of Counter-Propaganda

🔍	Undermine credibility of hostile propaganda
🗨️	Expose false information and deception
🛡️	Dismantle frames and manipulative narratives
👤	Verify credibility of sources and messengers
🧠	Build cognitive immunity and resilience
🗣️	Provide alternative explanations and narratives
👥	Protect social trust and institutional legitimacy

4 Major Forms of Counter-Propaganda

Form	Advantages	Limitations / Risks	Best Applied When	Potential Unintended Effects
✓ Direct Refutation	Clear, straightforward rebuke of false claims	May trigger familiarity effect; perceived as confrontational	Claims are new, audience is reachable and open	Backfire among hostile audiences
🔍 Indirect Refutation	Avoids repeating false claims, less reactive	May be perceived as evasive	When false claim is widely-known	May be ignored as not addressing claim
🔍 Fact-Checking	Builds credibility, provides evidence	Limited reach; slow to counter viral claims	For factual claims with verifiable evidence	Seen as biased or "establishment"
👤 Source Exposure	Undermines trust in propaganda sources	Requires strong evidence	When source is dishonest or corrupt	Legal risks; may be dismissed as smear
⚖️ Logical Contradictions	Appeals to reason and consistency	Audience may be unmoved by logic	When audience values consistency	Can be too complex for mass audiences
🗨️ Reframing	Changes perspective; weakens enemy frame	Hard to shift deep-seated frames	When current frame is manipulative	Reframe may be seen as spin
📖 Alternative Narratives	Fills vacuum; inspires positive identity	Takes time to build credibility	When audiences seek meaning and hope	If weak, may be seen as propaganda
🛡️ Pre-bunking / Inoculation	Builds resistance before exposure	Requires audience engagement	Before likely disinformation campaigns	May not reach those most at risk
😄 Humor / Satire	Engaging; shareable; lowers defenses	May trivialize issues; misfire	When issue allows light-hearted approach	Can offend or reduce seriousness
📺 Platform-Level Responses	Reduces spread at scale	Raises free speech concerns	When content violates rules or is harmful	Over-moderation; trust in platforms declines
📖 Education / Media Literacy	Long-term empowerment and resilience	Long-term; needs resources	For sustainable societal resilience	Results not immediately measurable

6 Historical Development of Counter-Propaganda

WWI	WWII	Cold War	Korea / Vietnam	Post-Cold War	Post-Korea	Russia War	Age of Social Media
1914-1918	1939-1945	1947-1991	1950s-1970s	1990s	2001+	2014-2019	2020s+
Birth of modern propaganda; leaflets, radio, government-ordered information bursts.	Mass propaganda; film, and censorship.	US-Soviet propaganda competition; Radio Free Europe/Liberty, VOA, BBC; leaflets, cultural diplomacy.	Leaflets, radio broadcasts, and heart-and-minds campaigns.	Everyday citizens; Total information control; and build national control.	Counterterrorism communication, CVE, and global narratives.	Hybrid warfare, disinformation, cyber operations, social media.	Sophisticated online content; mass online memes, and global recruitment.

5 Level of Effectiveness

Level	Reacts	Conducts propaganda de Operations
News	Credibility	Is message clear, credible, and resonant?
Source	Context	Is the source trustworthy and relevant?
Starts	Channel	Does the channel fit the audience's format?
Audience	Personas	Who are they and what motivates them?
Frame	Interpretive	What frame dominates their thinking?
Narrative	Stories	What story are they in—and why?
Network	Intergroup	How does information spread socially?
Resilience	Institution	Does it also align those groups?

15 Conclusion

Effective counter-propaganda is not simply the creation of stronger propaganda. It is the construction of a resilient, long-term information ecosystem based on accurate information, trustworthy institutions, transparent sources, independent journalism, media literacy, critical thinking, social trust, and institutional resilience.



2 Conceptual Boundaries: How Key Terms Differ

Term	Focus	Primary Purpose	Scope	Typical Actors
Propaganda	Promote a perspective; shape perceptions and behavior	Influence for political, ideological, or strategic ends	Messages, narratives, emotions, cognition	States, groups, movements, individuals
Counter-propaganda	Systematically counter hostile propaganda	Undermine, expose, neutralize, and replace	Messages, frames, sources, narratives, audiences	States, institutions, civil society, platforms
Counter-messaging	Specific messages that oppose or reject extremist narratives	Change attitudes and prevent mobilization	Targeted messages	Governments, NGOs, communities
Strategic Communication	Integrate words and deeds to achieve goals	Synchronize communication with policy and action	All communication activities	Governments, militaries, institutions
Public Diplomacy	Engage foreign publics to build understanding	Build relationships and favorable perceptions	International audiences, culture, values	Governments, cultural institutions
PSYOP / MISOP	Military information activities to influence emotions, motives	Affect adversary behavior in specific operations	Targeted groups in conflict zones	Military, defense forces
Information Operations (IO)	Coordinated employment of information-related capabilities	Gain information advantage over adversary	Systems, networks, people, processes	Military, intelligence, agencies
Information Warfare	Use of information to achieve military or political objectives	Disable, disrupt, deceive, degrade, destroy	Broad spectrum of information domain	States, militaries, non-state actors
Influence Operations	Covert or overt attempts to influence decisions	Shape choices and behaviors	Policies, elections, opinions	States, groups, individuals
Cognitive Warfare	Target cognition and decision-making processes	Manipulate perceptions, beliefs, and reasoning	Mind, identity, morality, culture	Advanced actors, state and non-state
Fact-checking	Verify factual claims	Determine truthfulness	Specific claims	Media, NGOs, platforms
Debunking	Expose false claims after they spread	Reduce belief in falsehoods	False or misleading claims	Fact-checkers, experts, institutions
Pre-bunking (Inoculation)	Prepare audiences before exposure	Build resistance to misinformation	Potential future narratives	Educators, communicators, platforms
Media Literacy	Teach skills to access, analyze, evaluate media	Empower critical thinking	Individuals, communities	Schools, NGOs, institutions
Counter-Disinformation	Broad effort to counter false information	Detect, expose, disrupt, and build resilience	Ecosystem of information	Governments, platforms, civil society

5 Psychological Mechanisms That Influence Effectiveness



8 Theories of Propaganda, Persuasion & Cognition

Jacques Ellul Propaganda as systematic manipulation of symbols in modern society.	Carl Hovland Experimental research on persuasion; credibility, source, and message.
Harold Lasswell Who says what, in which channel, to whom, with what effect.	Leon Festinger Cognitive dissonance; people adjust beliefs to reduce discomfort.
Edward Bernays Engineer of consent; public relations and psychological persuasion.	William McGuire Inoculation theory; refutations can strengthen beliefs (in some cases).
Walter Lippmann Stereotypes in our heads; the "pictures" in our heads shape reality.	Daniel Kahneman System 1 (fast) and System 2 (slow) thinking shape judgments.
Paul Lazarsfeld Two-step flow; opinion leaders influence mass attitudes.	Amos Tversky Heuristics and biases; framing effects in decision-making.

11 Law, Ethics: Democracies & Faultlines

Permissible Responses	Problematic Responses
✓ Transparency about fact-checking & corrections ✓ Source disclosure ✓ Eigen Denial ✓ Eigen Denial	✗ Transparency about fact-checking & corrections ✗ Source disclosure ✗ Independent verification ✗ Proportionality ✗ Legal oversight ✗ Press freedom ✗ Citizen autonomy & choice
✓ Source greater doubt? ✓ Source trusted over next? ✓ Patterns ✓ Extension ✓ Clearer disputes	✗ Censorship of speech ✗ Disinformation by the state ✗ Lack of transparency ✗ Excessive surveillance ✗ Punishing dissent ✗ Manipulation of media ✗ Undermining journalism ✗ Abuse of state bailiiffs.

Guiding Principles: Consistent with International Law, Human Rights, and Democratic Values.

10 Digital & AI Transformations

Bots & Trojans Automated disinformation and harassment.	Deepfakes & Synthetic Media Undermine trust in seeing and hearing.
Algorithmic Systems Drive attention, polarization, and virality.	Memes & Influencers Drive attention, polarization, and virality.
Anonymous Accounts Harder accountability and manipulation.	Microtargeting Tailored messages for vulnerability.
Online Communities Build identity and coordinate action.	AI-Generated Content Scale and speed of content creation.



14 Integrated Model: Propaganda vs. Counter-Propaganda



- Selected Key References (Representative)**
- Jowett, G. S., & O'Donnell, V. *Propaganda & Persuasion* (7th ed.). Sage, 2019.
 - Young, R. J. C. *Contentious Information Warfare: Towards an Assessable Framework*. Journal of Information Warfare, 2007.
 - Fallis, D. *Disinformation: A Primer in the Politics of Falsehood*. Oxford University Press, 2015.
 - Hamelers, M., & Van der Meer, T. (Eds.). *The Routledge Handbook of Disinformation*. Routledge, 2020.
 - NATO StratCom Centre of Excellence. *Countering Disinformation: Factual and Effective Communication*, 2020.
 - EU East StratCom Task Force. *Disinformation Review Reports* (2015–present).
 - U.S. Department of State. *Global Engagement Center Reports* (2016–2023).

Note: This document synthesizes extensive scholarly research and historical evidence. Interpretations remain contested in some areas of propaganda history and effectiveness.

Figure 3. Generated conceptual poster illustrating counter-propaganda as a multi-level system. Visual details are illustrative; the integrated model in Sections 14–15 supplies the report’s formal synthesis.

15. Design Principles for a Resilient Information Ecosystem

Diagnose before rebutting: Identify the campaign’s objective, target audience, narrative, network, and emotional function. Do not assume the visible claim is the real target.

Lead with verified reality: Make the accurate account more memorable than the false one. Provide evidence, context, and an alternative explanation.

Treat credibility as infrastructure: Build trust before crises through transparency, correction, competence, and independence. Credibility cannot be improvised at the moment of attack.

Use the right messenger: The most authoritative institution is not always the most persuasive source for a specific community.

Separate exposure from amplification: Respond proportionately. Do not give fringe narratives a national audience merely to demonstrate vigilance.

Target networks and incentives: When manipulation is coordinated, disrupt deceptive behavior, monetization, infrastructure, or amplification rather than only arguing with content.

Use pre-bunking as maintenance: Teach recurring techniques and reinforce periodically. One inoculation is not permanent immunity.

Preserve independent verification: Journalists, academics, courts, civil society, and competing political actors must remain able to challenge official claims.

Publish uncertainty: Use confidence levels, disclose what is unknown, and update publicly. Overconfidence creates future propaganda vulnerabilities.

Protect rights and autonomy: Counter-propaganda should expand citizens’ capacity to decide, not replace one manipulation system with another.

Align words with policy: When lived reality contradicts official communication, propaganda advantage shifts to the adversary. Strategic behavior is part of strategic communication.

Measure resilience, not applause: Track belief accuracy, trust calibration, diffusion, behavioral outcomes, and long-run robustness—not only impressions and engagement.

Conclusion

Counter-propaganda is a necessary concept because hostile influence does not operate only through false facts. It works through source credibility, framing, identity, emotion, social proof, repetition, selective exposure, network coordination, algorithmic amplification, censorship, and institutional weakness. The response must therefore be equally multi-level: verification, source exposure, correction, reframing, alternative narratives, pre-bunking, network analysis, platform governance, independent journalism, and institutional reform.

The historical record also warns against simplistic moral categories. Democracies have used overt truth-based broadcasting, independent journalism, covert deception, psychological warfare, and propaganda; authoritarian regimes have sometimes distributed factually accurate material for strategic ends. The analytical question is not whether a message is labeled “counter-propaganda,” but how it is produced,

whether it is accurate, whether its sponsorship is disclosed, whether citizens can contest it, whether the press remains free, and whether the state is constrained by law.

The strongest long-term defense is a resilient information ecosystem. Accurate information must be easier to verify; institutional claims must be auditable; official errors must be correctable; independent media must have access and protection; citizens must be able to compare sources; platforms must disclose meaningful governance processes; and education must develop critical habits without teaching cynical disbelief in everything. Resilience is not universal skepticism. It is calibrated trust: confidence that rises with evidence and falls when evidence fails.

Effective counter-propaganda, then, is not the creation of stronger propaganda. It is the construction of social and institutional conditions under which manipulation becomes harder to scale, harder to conceal, easier to investigate, and less likely to convert uncertainty into durable distrust. The ultimate strategic asset is not message dominance but a public capable of independent judgment within a plural, transparent, and accountable information order.

Representative Bibliography and Source Guide

The sources below combine classic works, peer-reviewed research, archival collections, government reports, and international-organization guidance. Official sources are included as evidence of institutional history, doctrine, or stated policy; they are not treated as independent validation of contested claims.

1. Lasswell, Harold D. *Propaganda Technique in the World War*. 1927.
2. Bernays, Edward L. *Propaganda*. 1928.
3. Lippmann, Walter. *Public Opinion*. 1922.
4. Lazarsfeld, Paul F., Bernard Berelson, and Hazel Gaudet. *The People's Choice*. 1944.
5. Hovland, Carl I., Irving L. Janis, and Harold H. Kelley. *Communication and Persuasion*. Yale University Press, 1953.
6. Festinger, Leon. *A Theory of Cognitive Dissonance*. Stanford University Press, 1957.
7. Ellul, Jacques. *Propaganda: The Formation of Men's Attitudes*. English edition, 1965.
8. McGuire, William J. Foundational studies on inoculation and resistance to persuasion, early 1960s.
9. Tversky, Amos, and Daniel Kahneman. "The Framing of Decisions and the Psychology of Choice." *Science* 211 (1981): 453–458.
10. Kahneman, Daniel. *Thinking, Fast and Slow*. Farrar, Straus and Giroux, 2011.
11. Lewandowsky, Stephan, Ullrich K. H. Ecker, Colleen M. Seifert, Norbert Schwarz, and John Cook. "Misinformation and Its Correction: Continued Influence and Successful Debiasing." *Psychological Science in the Public Interest* 13 (2012): 106–131.
12. Wood, Thomas, and Ethan Porter. "The Elusive Backfire Effect: Mass Attitudes' Steadfast Factual Adherence." *Political Behavior* 41 (2019): 135–163.
13. Guess, Andrew, and Alexander Coppock. "Does Counter-Attitudinal Information Cause Backlash? Results from Three Large Survey Experiments." *British Journal of Political Science* 50 (2020).
14. Ecker, Ullrich K. H., Stephan Lewandowsky, John Cook, et al. "The Psychological Drivers of Misinformation Belief and Its Resistance to Correction." *Nature Reviews Psychology* 1 (2022): 13–29. <https://www.nature.com/articles/s44159-021-00006-y>
15. van der Linden, Sander. "Misinformation: Susceptibility, Spread, and Interventions to Immunize the Public." *Nature Medicine* 28 (2022). <https://www.nature.com/articles/s41591-022-01713-6>
16. Vosoughi, Soroush, Deb Roy, and Sinan Aral. "The Spread of True and False News Online." *Science* 359 (2018): 1146–1151. <https://www.science.org/doi/10.1126/science.aap9559>
17. Lazer, David M. J., et al. "The Science of Fake News." *Science* 359 (2018): 1094–1096.
18. Lewandowsky, Stephan, John Cook, Ullrich Ecker, et al. *The Debunking Handbook 2020*. <https://sks.to/db2020>
19. Maertens, Rakoén, et al. Research on long-term effectiveness and booster interventions for misinformation inoculation, including *Nature Communications* (2025).

20. Biddlestone, Mikey, et al. "Video Inoculation against Election Misinformation across 12 EU Nations." *Nature Human Behaviour / Communications Psychology*, 2026. <https://www.nature.com/articles/s44271-025-00379-3>
21. National Institute of Standards and Technology. Reducing Risks Posed by Synthetic Content: An Overview of Technical Approaches to Digital Content Transparency. NIST AI 100-4, 2024. <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-4.pdf>
22. National Institute of Standards and Technology. Generative AI Profile, NIST AI 600-1, 2024.
23. UNESCO. Guidelines for the Governance of Digital Platforms. 2023. <https://www.unesco.org/en/internet-trust/guidelines>
24. Ireton, Cheryl, and Julie Posetti, eds. Journalism, "Fake News" & Disinformation: Handbook for Journalism Education and Training. UNESCO, 2018.
25. United Nations. International Covenant on Civil and Political Rights, Articles 19 and 20. <https://www.ohchr.org/en/instruments-mechanisms/instruments/international-covenant-civil-and-political-rights>
26. OHCHR. Rabat Plan of Action and materials on freedom of expression and incitement. <https://www.ohchr.org/en/freedom-of-expression>
27. European Court of Human Rights. Case-law guides concerning Articles 10 and 17, updated through 2026.
28. United States Committee on Public Information / George Creel. Complete Report of the Chairman of the Committee on Public Information. 1920. Library of Congress: <https://www.loc.gov/item/20026826/>
29. Creel, George. How We Advertised America. 1920. Library of Congress: <https://www.loc.gov/item/20010648/>
30. U.S. National Archives. Records of the Office of War Information (Record Group 208). <https://www.archives.gov/research/foreign-policy/related-records/rg-208>
31. U.S. National Archives. "A Reel Story of World War II" and United News records on overseas counter-propaganda.
32. The National Archives (UK). Research Guide: Propaganda; records of the Political Warfare Executive (FO 898). <https://www.nationalarchives.gov.uk/help-with-your-research/research-guides/propaganda/>
33. Radio Free Europe/Radio Liberty. Institutional history. <https://about.rferl.org/our-history/> (official institutional source; historical claims should be cross-checked with archival scholarship).
34. Voice of America. Historical materials and VOA Charter. <https://www.voanews.com/> (official institutional source).
35. U.S. Holocaust Memorial Museum. "Nazi Propaganda," "Nazi Propaganda and Censorship," and "Ministry of Propaganda and Public Enlightenment." <https://encyclopedia.ushmm.org/>
36. U.S. intelligence / declassified records on Soviet active measures and disinformation, CIA Reading Room. Examples include "Soviet Active Measures and Disinformation."
37. U.S. Department of State, Global Engagement Center. Counter-Disinformation Literature Review (2022–2024 archive) and reports on foreign information manipulation. (Official source; GEC authorization ended December 23, 2024.)
38. European External Action Service. 3rd Report on Foreign Information Manipulation and Interference Threats. 2025. https://www.eeas.europa.eu/eeas/3rd-eeas-report-foreign-information-manipulation-and-interference-threats-0_en
39. European External Action Service. 4th Annual Report on FIMI Threats. 2026. <https://www.eeas.europa.eu/>
40. NATO. "NATO's Approach to Counter Information Threats." Updated 2025. <https://www.nato.int/en/what-we-do/wider-activities/natos-approach-to-counter-information-threats>
41. NATO Strategic Communications Centre of Excellence / partners. Attributing Russian Information Influence Operations, 2026.
42. Eisenman, Joshua. China's Media Propaganda in Africa: A Strategic Assessment. U.S. Institute of Peace, 2023.
43. DiResta, Renée, et al. Telling China's Story: The Chinese Communist Party's Campaign to Shape Global Narratives. Stanford Internet Observatory, 2020.
44. Xu, Jian, and Qian Gong. "Telling China's Story Well as Propaganda Campaign Slogan." *Media, Culture & Society*, 2024.
45. Bélanger, Jocelyn J., et al. "Do Counter-Narratives Reduce Support for ISIS? Yes, but Not for Their Target Audience." *Frontiers / experimental research*, 2020. <https://pmc.ncbi.nlm.nih.gov/articles/PMC7325943/>
46. Fernandez, Alberto M. Here to Stay and Growing: Combating ISIS Propaganda Networks. Brookings Institution, 2015.
47. RAND Corporation. Research and commentary on counter-ISIS messaging and audience segmentation, 2016.
48. Lieberman, Ariel Victoria. "Terrorism, the Internet, and Propaganda: A Deadly Combination." *Journal of National Security Law & Policy*, 2017.
49. European Union / EEAS. FIMI Response Framework and annual activity reports, 2024–2026.
50. Clark, S., et al. Research on continued influence of AI-generated deepfake videos, 2026. <https://www.nature.com/articles/s44271-025-00381-9>
51. Groh, Matthew, et al. "Human Detection of Political Speech Deepfakes across Transcripts, Audio, and Video." *Nature Communications*, 2024.

52. Salvi, Francesco, et al. “On the Conversational Persuasiveness of Large Language Models: A Randomized Controlled Trial.” 2024 preprint; subsequent research should be checked for final publication status.
53. Science (2025). “The Levers of Political Persuasion with Conversational Artificial Intelligence.”
<https://www.science.org/doi/10.1126/science.aea3884>
54. Science Advances (2026). “Biased AI Writing Assistants Shift Users’ Attitudes on Societal Issues.”
<https://www.science.org/doi/10.1126/sciadv.adw5578>
55. OHCHR. Commission of Inquiry on Human Rights in the Democratic People’s Republic of Korea, 2014, and follow-up reporting.
<https://www.ohchr.org/en/hr-bodies/hrc/co-idprk/commission-inquiryon-h-rin-dprk>

Selected Web and Archival Entry Points

Library of Congress – Committee on Public Information: <https://www.loc.gov/item/20026826/>

U.S. National Archives – Office of War Information: <https://www.archives.gov/research/foreign-policy/related-records/rg-208>

UK National Archives – Propaganda research guide: <https://www.nationalarchives.gov.uk/help-with-your-research/research-guides/propaganda/>

RFE/RL history: <https://about.rferl.org/our-history/>

USHMM – Nazi propaganda: <https://encyclopedia.ushmm.org/content/en/article/nazi-propaganda>

OHCHR – ICCPR: <https://www.ohchr.org/en/instruments-mechanisms/instruments/international-covenant-civil-and-political-rights>

UNESCO – Digital Platform Governance Guidelines: <https://www.unesco.org/en/internet-trust/guidelines>

EEAS – FIMI reports: https://www.eeas.europa.eu/eeas/information-integrity-and-counterintelligence-foreign-information-manipulation-interference-fimi_en

NATO – Counter Information Threats: <https://www.nato.int/en/what-we-do/wider-activities/natos-approach-to-counter-information-threats>

NIST – Synthetic Content Risk Reduction: <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-4.pdf>

Editorial note: This report is a synthesis for analytical and educational use. It is not a manual for covert manipulation. Historical descriptions of deception, black propaganda, or psychological operations are included to understand the field and its ethical limits.