The Indonesian Journal of Development Planning
July 2026, Vol 10, No. 2, 282 - 301
https://doi.org/10.36574/jpp.v10i2.787

Research Article

Counter-Information Warfare and Policy Strategy of Digital Propaganda: A Review of Scientific Literature

Nugroho Sulistyo Budi[1], Guntur Eko Saputro[2], Robby M. Taufik [3]

[1, 2, 3] Indonesia Defense University, Jakarta, Indonesia

Abstract

This article presents an in-depth review of the literature on the complex strategy of fighting information wars and digital propaganda in the age of algorithms and artificial intelligence. The information space is currently turning into a cognitive battleground for state and non-state actors who utilize social media, automation, and generative technologies intended to manipulate public viewpoints, lower trust, and deepen social polarization. This study uses a qualitative approach with a systematic and thematic literature review method by utilizing the search and selection of academic sources, policy reports, and institutional documents based on relevance, novelty, and credibility. Data analysis in this study was through thematic categorization and narrative synthesis to identify conceptual patterns, strategic dimensions, and research gaps. The results show that there is no single effective approach to combating disinformation. Cross-sectoral synergy between government, digital platforms, academia, media and society by combining technology-based interventions, media literacy education, psychological immunization, and procedural counternarratives can be successful strategies to combat disinformation. Anti-disinformation policies and regulations need to be ethically crafted with procedural safeguards, restrictions on access to personal data, and independent oversight mechanisms to prevent misuse. There is a need for continuous evaluation and adaptation to assess the effectiveness of strategies, detect adaptations by disinformation actors, and adjust responses in real time. Information warfare and digital propaganda can be countered by: (1) the application of a layered approach to build cognitive resilience in society; (2) strengthening cross-sector collaboration; (3) the formulation of ethical governance and regulations; (4) continuous evaluation and adaptation of strategies; and (5) strengthening individual and collective capacity through media literacy and critical digital education.

Article History:
Received: 10 December, 2025
Accepted: 2 February, 2026
Published: 30 July, 2026

Keywords: information wars; digital propaganda; counter-strategies

Correspondence Email:
nugroho.budi@doktoral.idu.ac.id

1. Introduction

Today's rapidly developing digital technology has given rise to a new invisible environment, namely the information environment. Iacovitti (2022) revealed that the movement of information occurs quickly, massively, and cannot be completely controlled. The dissemination of information that is widely carried out on digital platforms is able to move dynamically so that the direction, scope, and impact of information dissemination are complex to predict. The ease with which each individual can obtain and distribute information without obstacles causes the information space to become increasingly open and vulnerable to manipulation that can affect the social, political, and security life of a country (Weiss & Briemann, 2023).

The increase in the volume and speed of information dissemination in the digital information space indicates the existence of information war. Content that is spread across the digital space is usually deliberately created to achieve

a specific goal that can benefit individuals or groups. Targeted messaging strategies are a form of information manipulation that is often used in digital propaganda (Zhu & Fu, 2024; Nieubuurt, 2021). The narrative built by the individual or group aimed at the target group is intended so that the perception of the target group can be influenced. This form of propaganda is delicately packaged and disseminated through various digital platforms (Bjola & Papadakis, 2020).

Digital propaganda has a great influence on the stability of democracy and the dynamics of the political process. Digital media is used to shape public opinion conducted by individuals or groups with special interests (Al Fatih *et al*., 2024). The dissemination of information is seen in a neutral manner, but there is a message that is actually shown for a specific political purpose. The resulting narrative can change public perception due to the issues that arise, as a result of which pro and con groups are formed from manipulative narratives. Marigliono et al., (2024) suggest that digital discussion spaces will experience debate and polarization of conflicts due to interactions between pro and con groups. In addition, public trust in institutions will be decreasing, thus encouraging social fragmentation. Thus, digital propaganda is a mechanism that can change the direction of the democratic process and broader political stability, not just a space for the dissemination of information.

Provocative propaganda and digital information warfare can create social segregation. The spread of manipulative narratives makes a person more susceptible to prolonged polarization and disorientation of information (Qureshi & Bhatt, 2024; Azzimonti & Fernandes, 2023). The eroding public trust in state institutions and institutions is the main target of digital propaganda. This causes the effectiveness of government policies to be disrupted, public spaces to be unstable, and the potential for social conflict to increase (Timotheou et al., 2023). Information operations in national security can be exploited by external actors, this is done to influence domestic policy, undermine the legitimacy of the government, and create political instability. Contemporary security studies are an important strategic aspect that needs attention, as threats to social cohesion and public trust undermine government legitimacy, and create political instability (Prieto Curiel, 2021).

Algorithms and the emergence of many bot accounts that help speed up the circulation of content greatly affect the dissemination of digital information (Gonzales & De Domenico, 2021). The use of algorithms on digital media platforms is intended to personalize information based on user preferences. Individuals can only be exposed to information that matches or matches the user's view due to the existence of this algorithm (Dahlgren, 2021). The use of automation and bots can trigger messages to be disseminated in a short period of time. Mergers between algorithms and bots can be more efficient in disseminating information but are difficult to detect (Taylor & Choi, 2022). This indicates that the information threat comes from a device or technology that supports the distribution of information. Therefore, the presence of information warfare and digital propaganda emphasizes the need for research to understand information counter-warfare and digital propaganda strategies in academic literature.

This article presents a comprehensive review of the scientific literature on anti-information warfare and digital propaganda strategies by synthesizing academic research, policy reports, and doctrinal analysis with the aim of providing a holistic and structured understanding of the contemporary threat landscape as well as the various responses that have been proposed, implemented, and evaluated. There are five main variables compiled in this article, namely (1) Threat Conceptualization, which defines and categorizes problems; (2) Taxonomy of Counter-Strategies, which maps the spectrum of responses; (3) Effectiveness Evaluation and Metrics, which examines how success is measured, and (4) Governance, Ethics, and Consequences, which addresses normative dilemmas. Based on this systematic approach, this article is expected to be able to become a relevant scientific reference for policymakers, academics, practitioners, and the general public in understanding and developing strategies to overcome one of the most pressing information security challenges in this digital era.

2. Methods

In this journal, a structured and thematic literature review is used to support qualitative research methods. Qualitative approaches in research can act as a strategy for exploring and understanding the meaning of various social and

humanitarian phenomena (Creswell, 2016). Not only does it help to collect data, but the qualitative approach also helps to express the experiences of individuals or groups in a social context using the depiction of subjective meaning.

Through a qualitative perspective, researchers try to understand the complexity of problems by studying various aspects, situations, and social interactions around them. Therefore, the understanding of problem phenomena through qualitative research will feel more complete because it studies values, beliefs, and social dynamics that cannot be described through numerical aspects in quantitative research.

The research process is carried out through several stages. The first stage is to formulate keywords and searches that are relevant conceptually and practically to research topics from various primary and secondary sources such as *academic* databases*,* official publications, institutional reports, and other sources.

The material that has been collected from various sources is then sorted and documented systematically based on predetermined categories, such as topic relevance, novelty, methods used, and accessibility. Furthermore, thematic analysis and manuscript synthesis processes are carried out to determine patterns, *key concepts,* and relationships between themes. Then triangulation and interpretive testing are carried out to ensure the reliability and validity of the research, in this stage the principles of trust are also applied, such as credibility, transferability, reliability, and confirmation. From the several stages that have been carried out, the writing results will follow the principle of transparency with documentation in the form of reporting on search procedures, reasons for exceptions, and methodological limitations.

3. Results, Analysis, and Discussion

3.1 Conceptualization of Threats (Definition and Taxonomy of Problems)

In formulating an effective information warfare and digital propaganda strategy, it is necessary to map the threat landscape by combining various views of the scientific literature and military doctrine so that they can be categorized and analyzed. This aims to create a foundation for understanding ranging from military doctrine to the psychology of information conflict.

The Dimension of Information Warfare

One of the approaches to information warfare is related to the Gerasimov Doctrine introduced by Valery Gerasimov from Russia. The "Gerasimov Doctrine" is a hybrid war concept that integrates hard and soft forces of various domains in an attempt to develop an operational concept for Russia's confrontation with the West as a proponent of the actual doctrine that has guided Russian policy for more than two decades. This doctrine is a non-military war tactic that uses cyberattacks, disinformation and the use of proxies in a region combined with conventional forces (Rumer, 2019).

Information warfare is slowly evolving into cognitive warfare which is a war strategy in the identification of the human mind as the main battlefield by NATO and Experts (Marjanović & Smiljaniç, 2025). The focus in cognitive warfare is reasoning in decision-making without physical combat to achieve a specific strategic goal. Russia adopted the concept of "reflection control" in its cognitive warfare doctrine as the basis for decision-making so that the opponent unknowingly had made a beneficial decision (Institute for the Study of War, 2025). Psychological manipulation, synchronization of narratives with geopolitical events, and exploitation of cognitive kinship are strategies in manipulating opponent/target thought processes (Paziuk *et al*., 2025). Cognitive warfare focuses on the use of modern technology in thinking, decision-making and maintaining social solidarity (Marjanović & Smiljaniç, 2025).

The information war that has evolved into cognitive warfare has undergone a strategic change so that many state actors recognize that the cognitive battlefield needs attention, not only about physical and digital. Cognitive warfare aims to deny access to information and also disrupt the target's reasoning process. In traditional information warfare, information and system control are the focus so that information can be controlled (Courter, 2024). According to Tashev *et al*. (2019) Cognitive warfare focuses on targeting the minds of individuals who send, receive, and process information so that the reasoning of opponents/targets is illogical. This explains that in dealing with the symptoms of information war, the strategy that can be done is to check facts or delete content that is less effective. A comparison of some conceptual different forms of operation in the information domain is presented in the table below.

Table 1. Comparative Framework of the Information Warfare Concept

Concept	Main objectives	Main Targets	Typical Methods
Hybrid Warfare (Gerasimov)	Achieving political goals through non-military means	Political will and stability of the target country	Information warfare, cyberattacks, economic pressure, proxies
Information Operations (IO)	Influencing, interfering, or usurping the opponent's decision-making	Information systems and decision-making processes	PSYOP, military fraud, electronic warfare, cyber operations
Cognitive Warfare	Influence the reasoning and behavior of opponents	Human cognition, beliefs, social cohesion	Psychological manipulation, narrative synchronization, exploitation of cognitive bias with AI

***Source**: Compiled by the author based on a synthesis of literature from Gerasimov (2013), NATO Strategic Command (2021), and Smith (2022).*

Psychological operations (PsyOps) are used with the aim of influencing opinions and behaviors through the delivery of persuasive information to opponents or targets by influencing emotions, motives, and objective reasoning to reinforce foreign attitudes and behaviors that are favorable to the perpetrator (Yovana *et al*., 2022; Bachtiar *et al.*, 2023). Whereas, information operations (IO) are actions to protect one's own system by influencing the opponent's information system so that this change makes the boundary between PsyOps and IO fade (Courter, 2024). In 2010, the U.S. Department of Defense replaced the term PsyOps with Military Information Support Operations (MISO) adding to the ambiguity in the meaning that the change was made to avoid "propaganda" and "brainwashing." This adds to the confusion because the change in the term "information support" can have another meaning.

The doctrine of the United States that replaces the concept of IO with a broader framework is Operations in the Information Environment (OIE). The emergence of OIE blurs the line between information as a support and influence as an operational goal (Courter, 2024). Information is a means and influence is a desirable impact on human thought and behavior, yet some literature states that equating "information" and "influence" is a doctrinal error (Courter 2024). In this case, the explanation of the role of cyberspace as the main medium in modern information warfare is based on the relationship between information and its influence. Information warfare and cyber operations are not separated because they are strategically intertwined. Cyber warfare is the use of cyberattacks that are delivered to disrupt, damage, or seize control of an adversary's digital and physical infrastructure (Yang & Zhu, 2025).

Information operations that use cyberspace by exploiting social media to shape public narratives and perceptions will disrupt democratic freedom, so there needs to be resistance with cyber forces (Nayakama, 2022). The cyber dimension is a space that combines information and cognitive aspects through digital propaganda mechanisms. Traditional cyber operations are used defensively to deny an adversary the ability to conduct its own information operations by disabling communication servers (Nayakama, 2022). Cyberattacks can have a maximum or comprehensive impact by targeting operational technology systems (Yang & Zhu, 2025).

Digital Propaganda Mechanism

The involvement of information production and distribution through systematic engineering of human interaction in cyberspace is the integration of information and cognitive aspects in the modern digital era by utilizing technology

that is increasingly sophisticated today. By exploiting, a multidimensional approach through digital platforms that actually exploits human psychological vulnerabilities, manipulates algorithms to prioritize and disseminate information accessed repeatedly that is later directed to the audience group that is considered the most emotionally vulnerable.

Increasingly sophisticated technology in the modern digital era is currently encouraging the integration of information aspects. This involves human interaction as the production and distribution of information in cyberspace. The multidimensional approach is designed to exploit digital platforms and exploit human psychological vulnerabilities. The exploitation was carried out using the manipulation of social media algorithms that prioritized and disseminated information that could be accessed twice. The massive dissemination of certain information to the audience group with emotional susceptibility.

Digital propaganda in this context can be used as a tool to disseminate one-sided information in order to create a coordinated public perspective. Arrangements in the dissemination of information and distribution times are made to maximize psychological and social effects. Digital propaganda has a category of three main three-dimensional mechanisms, namely content type, dissemination techniques, and persuasion methods. Content type is a dimension that relates to the form and substance of the information disseminated, including the accuracy, intent, and degree of manipulation involved. The dissemination technique is a dimension that focuses on the distribution of information through several digital channels such as social media, fake news sites, instant messaging applications, or controlled influencer campaigns. Meanwhile, the last dimension, the persuasion method, is a dimension that focuses on psychological strategies used to shape the audience's opinions, emotions, and behaviors.

Digital propaganda has the main forms of framework used such as Disinformation, Misinformation, and Malinformation (DMM) (Dinfos Pavilion Team, 2025). In this case, the framework works as the basis of the content in distinguishing the level of truth of the information motives behind its dissemination. False information that is deliberately disseminated to deceive, manipulate, and harm a party is defined as disinformation (Princeton Public Library, 2025). Disinformation is a lie that is deliberately carried out to achieve a certain goal. While the dissemination of false information without malicious intent is called misinformation, this can happen because a person believes that information is widely spread, so that it does not directly cause dissemination (Dinfos Pavilion Team, 2025). Malinformation is defined by information that is widely disseminated based on facts, but is shared with the intention of harming one party. The main framework of this digital propaganda is related to a person's personal and privacy data which often experiences leaks and breaches that create negative perceptions (Dinfos Pavilion Team, 2025).

Disinformation, Misinformation, and Malinformation (DMM) are differentiated into harmful subtypes of content. This content includes fabricated content, original content that is imitated and manipulated, misleading information, and satirical or comedic content Information that is presented as fact (Dinfos Pavilion Team, 2025). Information generated from completely fabricated content is the most extreme form, as there is no basis in facts made up so it is deliberately done for fraudulent purposes. Other information such as manipulated content is content that is created to support a narrative. In addition, fraudulent content is created by imitating credible styles and sources, in this case official media or government agencies. False context content with the presentation of correct information but the context is misinterpreted as misleading. The latter content is satirical or parody that has a humorous or critical nature, but the fact is often mistaken for correct information by some.

The implementation of digital propaganda depends on the persuasion techniques and the substance of the message used, the techniques involve social media, fake news sites, automated bots, instant messaging apps, and influencer accounts as digital channels. This digital propaganda is divided into several dissemination techniques such as Troll, Bot, and Astroturfing. The Troll Farming Technique employs hundreds of human operators in an organized manner to create and manage fake social media accounts funded by countries such as the Russian Internet Research Agency (IRA) (NATO, 2020). In the process, the flow of messages is done in rotation all the time to ensure a stable and consistent reach right to the audience.

The automatic deployment technique is called the botnet technique where the network is programmed to perform certain tasks simultaneously, such as liking, retweeting, or commenting on certain posts (Feldman & Nahmias, 2025).

But the messages generated by these bots are easily recognized as spam because they are of low quality. Nonetheless, bots have a primary function as content created by human trolls to be amplified. Fraud of platform algorithms by using humans to think that narratives are more popular to be widely accepted than they actually are (NATO, 2020). Third, the creation of illusions that seem real but in fact are engineered by certain parties is digital astroturfing. The practice is a top-down activity for a campaign that is made to appear to come from an individual initiative (Kovic et al., 2018). Digital astrosurfing aims to form an idea, a policy for political figures to provide broad support so as to create social pressure and false legitimacy.

Troll, bot, and astroturfing propaganda techniques work synergistically within a single ecosystem. Content generated by human trolls is difficult to detect automatically. Reach is amplified by the volume and speed of the bot's network's reach. Social media platforms provide algorithms to interpret popular content when it reaches a certain limit and has the advantage of spreading to more social media users. Then the content that generates thousands of likes and variety is considered credible news (fake social proof) and spreads more widely and massively.

Propaganda that is successfully disseminated will then be ensured that the message can influence the audience. Success indicators can be achieved by using digital platforms and playing human psychologists to reinforce the persuasive effect. Micro-targeting and the creation of information bubbles (echo chambers and filter bubbles) are two methods of persuasion. The focus of micro-targeting is the use of personal data collected massively through online activities, interactions in social media, and other digital footprints. To convey the message, it will be tailored to the psychological character, and demographics of individuals or small groups. In convincing the message received, the perpetrator uses propagandist content that can understand the individual's vulnerability, preferences, and cognitive biases (Tucker, 2020).

Information environments that are closed or known as echo chambers and filter bubbles can form digital propaganda. The concept is in the form of a phenomenon that describes that internet users will be exposed to information that strengthens the view in user preferences. The main source is the difference between these two phenomena. Human choices and behaviors form echo chambers. Aligned beliefs (selective viewing) with like-minded users (homophilia) encourage to keep searching, sharing, and interacting (Moller, 2021). Meanwhile, filter bubbles are generated from the algorithmic mechanism of social media such as Facebook, YouTube, Twitter, etc. which has an automatic mechanism to adjust the content feed according to the preferences of online behavior that was done previously. Therefore, unknowingly being directed to information that is "liked" and considered "relevant", algorithms systematically reduce information that is contrary to the user's preferences (Möller, 2021). The fragmentation of information caused by echo chambers and filter bubbles depletes rational public spaces, digital propaganda can develop effectively and sustainably under ideal conditions (Arguedes et al., 2022). Echo chambers and filter bubbles need to be understood in order to examine the actors and processes that occur in digital propaganda attribution.

Actors and Attribution

In digital propaganda, there are actors that can be divided into state actors and non-state actors. State actors (state-sponsored) are actors who in achieving geopolitical goals make information operations a tool of foreign policy and national security. State actors that carry out active and sophisticated information operations are Russia, China, Iran, and North Korea (Institute for the Study of War, 2025). The operations of such countries are carried out by intelligence agencies or special military units with the integration of broader military strategies (Tashev, *et al.*., 2019). Non-state actors are usually filled by "patriotic" hacker groups, PR firms, or even criminal organizations as proxies to carry out operations (Hill, 2019). The use of proxies will provide strategic advantages so that the sponsoring country for engagement and avoid international accountability (Malyarenko & Kormych, 2022). This provides an unclear boundary between cybercrime, espionage, cyber warfare, and political activism (Raba, 2025).

It is the process of reliably identifying the source of an attack or campaign, attribution in the context of cyber operations can also pose significant challenges because it is multidimensional, involving technical, legal, and political aspects. It can take the form and use of stealth techniques such as routing attacks through servers in different countries, using virtual private networks (VPNs), and exploiting uninvolved third-party computer infrastructure. Hill (2019)

states that it is things like this that cause the attribution process to be very complex and time-consuming because it complicates the process of digital forensic investigations to establish a definitive relationship between the attack and the source.

From a legal point of view, international law in relation to actions between non-state actors and a state is set within a very high threshold. The sponsoring country that exercises "effective control" regarding a particular operation using proxies must be proven by the victim country (Hill, 2019). In addition, international law does not yet have specific rules regarding relevant cyber attribution (Bank, 2021). The complexity of determining a proportionate response to the law in cyberspace is complicated by the lack of a mutually agreed definition of "armed attack" or "use of force" (Hill, 2019).

From a political point of view, public attribution takes into account diplomatic and national security factors, not just technical analysis. Strong technical and intelligence evidence in a particular country is not published with diplomatic considerations, the risk of escalating conflicts, or the protection of intelligence sources and methods (Hill, 2019).

The challenge of the "attribution issue" is a challenge to the strategy deliberately exploited by state actors, not just a technical challenge. Cyber operations are designed by the state to remain below legal and technical thresholds with clear attribution objectives, using proxies, and technical concealment (Finlay & Payne, 2019). The goal of the evil actor in crippled the capabilities of the target country can be achieved and still defend itself under applicable international norms. In this situation, it proves that an effective response does not depend only on perfect technical attribution, but includes the development of a new diplomatic and legal framework for achieving accountability based on behavior, indirect evidence, and collective effort. The relationship between response achievement and attribution can be used as a basis for examining the results of information war targets and digital propaganda.

Impact (Targeted Outcomes)

The effects of information warfare and digital propaganda have been designed according to specific objectives. Much of the literature proves that the main effect of information warfare and digital propaganda is the erosion of public trust in institutions, authorities, and truths that fail to be conveyed to the public. Government institutions, the media, and the judiciary are among the targets in disinformation attacks to undermine public trust (Paziuk et al., 2025). When people fail to understand who can be trusted, the ability of people to actively participate in meaningful public interests is also significantly reduced. In addition, disinformation campaigns also target the fields of science and public health as easy targets. During the Covid-19 pandemic, the propaganda that fueled the disinformation campaign was successfully coordinated to the maximum, causing anxiety and doubt about the scientific consensus, so that vaccine hesitancy spread among the public in the form of rejection of public health measures so as to hinder the government in overcoming the worse impact (Alaminos-Fernández & Alaminos-Fernández, 2025). The same thing is also used in undermining public trust in scientific consensus on crucial issues such as climate change (Druckman, 2022). It can be understood that the existence of digital campaigns and propaganda has the main purpose of controlling public trust in institutions, science, and the ability of individuals to judge the truth by utilizing digital media to spread false information. This decline in trust can lead to a weakening of social cohesion and political stability in many countries. The magnitude of these impacts can inform alternative strategies in responding to information wars and digital propaganda.

3.2 Taxonomy of Countervailing Strategies (Response & Intervention)

Information warfare is becoming an increasingly complex threat, along with this, there is a need for countermeasures that can be applied to various levels of society. The Taxonomy of Countermeasures section presents the results of the analysis which have been categorized into four main levels of analysis, namely: macro (regulations and policy-based strategies), meso (technology/platform-based strategies), micro (individual and social-based strategies), and organizational (strategies based on collaborative approaches).

Policy-Based Regulations and Strategies (Macro Level)

At the highest level, international institutions and governments seek to form information systems through laws and regulations and diplomacy. This step aims to engage users, countries, and corporations in the digital world. National laws and regulations are one of the approaches taken in many countries. The circulation of fake news is suppressed by providing restrictions on its use through applicable regulations. In Indonesia, this step is realized through the Electronic Information and Transaction Law (ITE). The existence of these laws and regulations provides a legal framework in the gray area of pre-existing digital law enforcement (Harefa & Laila, 2024). However, the implementation did not always run smoothly. Several studies show that the definition of "hoax" in the ITE Law is still ambiguous and can give rise to different interpretations, as law enforcement is often inconsistent (Harefa & Laila, 2024). Technical obstacles in law enforcement that are often encountered are when perpetrators use anonymous identities or fake accounts, exacerbated by limited resources in digital tracking (Harefa & Laila, 2024). On the other hand, various criticisms have arisen regarding the rubber provisions in the ITE Law considered to provide an opportunity for excessive actions, which are used to suppress public criticism and differences of opinion in politics (Ufen, 2024).

In addition to domestic regulations, various countries are also focused on the very significant role of digital platforms as a medium for the dissemination of information. Digital platforms are the core where information can be accessed, the number of digital platforms that are developing raises debates about how to manage digital platforms. Initially, the government handed over the responsibility for regulation to the development company through a self-regulatory mechanism, this was followed up by the company by drafting a code of conduct and voluntarily filtering out disinformation. This approach has drawn criticism for its lack of binding sanctions and transparency. Many researchers equate data access from platforms to "episodic and arbitrary" activities, which make surveillance not carried out consistently (Shattock, 2021).

The presence of regulatory challenges does not stop at the national level and the platform; information exchange moves across national borders, and responses that rely on domestic policies are considered inadequate. At the international level, the concept of cyber diplomacy is slowly emerging, this activity is in the form of the use of diplomatic instruments to manage conflicts and problems in cyberspace (Christou, 2024). This diplomacy is aimed at creating a common understanding of acceptable state behavior in cyberspace, preventing the development of conflicts, and stability (Cyber Diplomacy Toolbox, 2025). Countries introduced two main agendas, namely affirming that international law, including the UN charter, remains applicable in cyberspace, and developing voluntary norms regarding responsible state behavior, this is conveyed through multilateral forums (Christou, 2024). The commitment conveyed to implement norms so as not to attack civilian infrastructure is important, and to cooperate in the investigation of cross-border cyber cases (Christou, 2024). Efforts made at the international level that information warfare require coordination across jurisdictions. Whether this is effective or not depends entirely on the ability of actors to translate these norms and commitments into operational mechanisms that are carried out at the meso and platform levels.

The control of disinformation in democratic countries is important to be developed in an appropriate, measurable structure that relies on strengthening a reliable information ecosystem. From a national security perspective, the main focus emphasized is to ensure that interventions are concentrated on the protection of digital spaces over content that is intentionally or intended to deceive the public and result in public order and safety and political stability being disrupted. To address this, a multi-layered response system is needed as an instrument that clearly integrates regulations, manages digital *platforms* responsibly, and actively engages the public.

Democratic countries need to build an appropriate, measurable, and oriented framework to strengthen the information ecosystem to overcome the information that occurs. From a national security perspective, interventions should be focused on protecting digital spaces from content that deliberately misleads the public that poses a risk to political stability, public order, and public safety needs to be made a top priority. For this reason, it is felt that there needs to

be an architecture in the form of a layered response that integrates clear regulatory instruments, structured digital platform governance, and active public participation.

The government needs to develop regulations on disinformation that have a basis for evidence and risks to ensure that policies are on target without curbing freedom of opinion. In its implementation, law enforcement is obliged to uphold transparency and procedural justice, starting from the clarity of the legal basis to the provision of appeal space for affected users. In line with that, digital platforms must have a moral responsibility to mitigate the spread of harmful information through preventive measures such as context labeling, scheduled and periodic algorithmic audits, and reporting of misinformation dissemination data in an open and easily accessible manner by the public.

Technology-Based Strategy (Meso/Platform Level)

The tools and policies that digital technology companies use at the platform level are used to detect, manage, and provide context for malicious content at scale. The intervention formed is the main line of defense in the information ecosystem. The main approach used is artificial intelligence-based automatic detection. Manual moderation is not possible against the millions of pieces of content that appear every second. Machine learning algorithms and deep learning are used to recognize speech patterns, news structures, and metadata. Studies with models such as BERT, CNN, and LSTM are able to demonstrate the classification of fake news with a very high level of accuracy after being trained using datasets containing real and fake news (Alshuwaier, 2025). Detection of Deepfakes also uses a similar approach, with algorithms trained to be able to read small inconsistencies in manipulated images, videos, or audio that are often unnoticed by humans (Rana, 2022). The promising results do not negate limitations, automatic detection is still limited in capturing context or sarcasm, the risk remains in false positives and false negatives.

In addition to automatic detection, content moderation policies for community rules are also implemented by digital platforms. This policy is in the form of restricting the reach of content (ranking down) so that content is not promoted by algorithms until it is in the form of account deletion (de-platforming). However, research criticizes that the policies made by the platform are often unclear and inconsistent in their implementation (Deng et al., 2024). The effectiveness of account deletion itself cannot be said to be linear. The study found that blocked users became five times more active when moving using alternative platforms such as Gettr (Deng et al., 2024). Survey-based studies show the opposite that strict measures such as freezing and suspending accounts are said to be effective in limiting the spread of misinformation because it instills fear of social isolation (Almufarrij & Toia, 2023). These findings prove that the moderation enforced is necessary to consider the psychological dynamics of users, not just the technical aspects.

Labeling and fact-checking are increasingly being used as alternative strategies. Some digital platforms provide additional warning labels on posts of potentially misleading content rather than removing content. Experimental studies show the credibility of false information will be lost and prevent users from sharing it (Liu & Wang, 2025). The platform also collaborates with independent fact-checking professional organizations in most cases to professionally certify. The tendency of users to trust labels from fact-checkers or professional media rather than automated labels although trust levels can still be politically polarized (Jia & Lee, 2024).

Scalability and trust in technology-based strategies still have fundamental tensions. Automation in the form of scalable AI detection that can process millions of pieces of content in a short time is still not trusted by users and vulnerable to contextual errors. Human-based solutions in the form of fact-checking carried out by professionals are far more credible and accuracy is on target, the disadvantage is that the slow and labor-intensive process cannot keep up with the volume of incoming content. This is a dilemma for platforms because they have to choose scalable interventions that may not be effective due to lack of trust or trusted interventions that cannot be scaled. The most likely solution lies in a hybrid "human-in-the-loop" system, where AI is leveraged to traise and early identify high-risk content and then reviewed by professional content analysis experts. Approaches that leverage technology can be combined with other strategies that focus on social and human aspects.

Human/Social Based Strategies (Micro Level)

In this strategy the focus is on the individual, the goal is to increase the cognitive and social resilience of the community to propaganda and misinformation. The approach used is the use of the best defense premise created by a critical

audience. Digital literacy and media literacy programs make up the most commonly used approaches. The public feels helped by this program because it helps develop practical skills such as checking the credibility of sources, understanding biases, recognizing persuasive techniques, and distinguishing types of misinformation (American Public University, 2025). Several studies have shown that short-term interventions can improve an individual's ability to distinguish between fake and real news (Guess et al, 2020). Based on a systematic review, programs that have a high success rate have direct elements and are aimed at young age groups, especially students (Eyal & Te'eni-Harari, 2024).

However, the ability to recognize false information does not necessarily stop a person from spreading it. MIT Sloan's (2022) research found a difference between the ability to distinguish between true and false information and the decision to share it. Sometimes users constantly share information even though they know it is wrong, citing social acceptance in group identity (Brown, 2022). These findings reinforce that psychological approaches go hand in hand with digital literacy.

This psychological approach is known as psychological inoculation or pre-bunking. With the strategy of giving "little exposure" to manipulative techniques (e.g., propaganda), then showing ways to refute them (Feng, 2023). In this way, users can build "cognitive antibodies" that make them more alert in the face of misinformation and real information. The effect given is an increase in resistance to misinformation and a decrease in the perception of the truth of the content (Han et al., 2025). Many experiments have shown that such an approach is considered effective in terms of health, politics, and climate change. An example of application is the educational game Bad News, which is considered to have succeeded in training millions of users in recognizing misinformation patterns such as polarization and emotional manipulation (Roozenbeen et al., 2020). Some studies point to the fact that its protective effects are able to last more than six weeks (Han et al, 2025).

Reactive strategies use debunking and counter-narrative approaches, which are used when misinformation has spread and is believed. Debunking aims to refute false claims by revealing true evidence. The results of the study show that this method is effectively used when the explanation given is complete, not only conveying that the information is wrong (Anneberg Public Policy Center, 2017). The obstacle regarding debunking is the influence of the continuous effect, where incorrect information will continue to affect the individual even after knowing (Ecker et al, 2024). The alternative is a counter-narrative approach that has an attempt to replace false information with a new narrative that is structured and reasonable (Sangalang et al., 2019). This method works with emotional and mental models so that it can deconstruct a person's beliefs as a result of misinformation (Sangalang et al., 2019). The consideration is the characteristics of each strategy, not choosing the best but using it according to the time and situation. This characteristic can be formed in an organizational approach with the involvement of cross-actor coordination.

Organizational Strategy (Whole-of-Society Approach)

Information warfare cannot be handled by a single actor, but also encompasses the political, social, technological, and psychological spheres of society. Efforts to combine various sectors from society to government are an approach that needs to be emphasized, so that all can work in a coordinated manner (European Commission, 2025). This approach shows that each stakeholder has a different role but can complement each other. For example, governments have a role in setting policies, providing funding, and coordinating national responses, including cyberpolation (European Commission, 2025). Digital platforms are responsible for implementing transparent content moderation, adjusting algorithms to curb the spread of harmful content, and ideally sharing data with researchers so that independent research can be conducted (European Commission, 2025)

Just like the government, independent media and fact-checking organizations also have a very important role in verifying the claims that are circulating, as well as providing true and accurate information to the wider public (Gori, 2024). Meanwhile, according to Gori (2024), academics and researchers play the role of evaluators of the effectiveness of interventions and develop theories and supporting tools. Viewed from another perspective, civil society organizations are involved in advocating for digital matters, education, media literacy, and monitoring government and platform transparency (European Commission, 2025). Similarly to the private sector, investors who use ESG

considerations encourage platforms to take climate disinformation more seriously (Gori, 2024). This approach is also adopted by NATO as a framework for handling information threats in collaboration with allied governments, international organizations, the private sector, and academia to ensure a coordinated response (NATO, 2025). In academia, this concept has been realized in collaborative work models such as penta-heli (government, business, academia, media, and NGOs) or hexa-helix by adding the community as the main stakeholder (Kowalski, 2023).

This approach can work if special units or task forces are formed as many countries and organizations realize. An example is the Global Engangement Center (GEC) in the United States, which coordinates interagency efforts to counter foreign propaganda (U.S. Department of State, 2025). NATO's Rapid Response Group (NRRG), which is capable of providing early warning and response to information threats (NATO, 2025). At the national level, several countries have established anti-hoax task forces and strategic communications units to monitor the development of narratives and responses to their spread. However, this approach is not as simple as the concept, the practice opens up the possibility that cross-actor collaboration is often hampered by conflicts of interest and trust. Governments that have political incentives in using anti-disinformation rules as a tool to silence criticism are the result of a breakdown of public trust (Mahapatra et al., 2024). Digital platforms are reluctant to open up private data and are afraid that their internal practices will be exposed (Pasquetto et al., 2020). Meanwhile, civil society organizations and academics are limited in resources in holding government actors and platforms accountable (Gori, 2024).

This results in the potential for mutual blame as the problem continues. To avoid this potential, governance is needed that is able to align profits between actors. Concrete steps have been suggested by some literature including the obligation to share data access securely as implemented by the EU DSA, the establishment of independent watchdogs, and co-funding for the development of media literacy research (European Commission, 2020). With this method, information is made a shared responsibility to conduct relevant evaluations of the effectiveness of the strategies used.

3.3 Evaluation and Effectiveness Metrics

In measuring strategies, it is necessary to conduct a systematic evaluation of the impact produced. Effectiveness evaluation plays an important role in understanding the extent to which the goals of counter-disinformation strategies can be achieved. The following subchapter will discuss several studies that assess the success of counter-disinformation strategies through the identification of indicators of success and obstacles in assessment.

Success Indicators or Key Performance Indicators (KPIs)

The fight against disinformation can be measured in many ways, from platform-level metrics to individual cognitive changes. At the platform level, the main indicator is reduced reach and engagement. This assessment indicator measures the number of views, likes, shares, and other forms of social media interaction with content that contains misinformation after interventions such as downgrading or labeling of content. However, the limitation of this indicator is impression data, which is generally exclusive and can only be accessed by platforms, making it difficult for independent researchers to access it (Pasquetto et al., 2020).

At the individual level, indicators of success focus on increasing resistance to false information and being able to distinguish between false and true information. The most commonly used metric is accuracy wisdom, which is an individual's ability to accurately distinguish between true and fake news. Several studies have shown that interventions such as fact-checking, inoculation strategies, and media literacy training can significantly improve individual scores in this area (Guess et al., 2020). In addition, there is an indicator of sharing wisdom that is judged by the extent to which an individual is able to apply this accuracy assessment in their information sharing behavior. However, a number of studies have shown that an increase in accuracy wisdom is not necessarily followed by a change in sharing behavior, which shows a gap between belief and action (Brown, 2022). Reduced confidence in false information is also another indicator, measured using a likert scale that assesses how far participants deduce a claim to be true or false. Research results show that there are positive effects that are considered consistent in this aspect, especially after interventions in the form of labeling or debunking (Porter & Wood, 2021). At the individual level, indicators of success focus on increasing resistance to false information and being able to distinguish between false and true information. The most commonly used metric is accuracy wisdom, which is an individual's ability to accurately distinguish between

true and fake news. Several studies have shown that interventions such as fact-checking, inoculation strategies, and media literacy training are able to significantly increase individual scores in this area (Guess et al., 2020). In addition, there is an indicator of sharing wisdom that is judged by the extent to which an individual is able to apply this accuracy assessment in their information sharing behavior. However, a number of studies have shown that an increase in accuracy wisdom is not necessarily followed by a change in sharing behavior, which shows a gap between belief and action (Brown, 2022). Reduced confidence in false information is also another indicator, measured using a likert scale that assesses how far participants deduce a claim to be true or false. Research results show that there are positive effects that are considered consistent in this aspect, especially after interventions in the form of labeling or debunking (Porter & Wood, 2021).

Table 2. *Types of Interventions, Psychological Mechanisms, and Effectiveness Against Disinformation*

Types of Interventions	Major Psychological Mechanisms	Effectiveness on Confidence (Average)	The Effectiveness of Sharing Intent	Key Challenges/Context
Fact Labeling	Giving a hint of external credibility	Quite effective in reducing trust in false claims (Porter & Wood, 2021)	Varied effects; can reduce the intention to share (Liu & Wang, 2025)	Trust in labeling sources (e.g., fact-checkers vs. algorithms) is critical (Jia & Lee, 2024); Partisan bias
Dismantling	Correcting incorrect mental models	Effective in reducing trust, but residual effects often remain (ongoing influence) (Annenberg Center for Public Policy, 2017; Ecker et al., 2024).	Less studied, but likely to be	Risk of boomerang effect (rare) (Swire-Thompson et al., 2022); details and explanations are better than simple rejections (Annenberg Public Policy Center, 2017).
Psychological Inoculation	Building resistance through pre-bunking	It is highly effective in reducing susceptibility to persuasion in the future (Han et al., 2025; Johansson et al., 2023)	Effective in reducing the intention to share (Johannson et al., 2023)	Requires active involvement; Scalability through games and videos shows promise (Roozenbeek et al., 2020)
Media Literacy	Improve critical thinking skills	Effective in improving the ability to distinguish accuracy (Guess et al., 2020)	Ineffective in improving the ability to differentiate sharing (Brown, 2022)	Requires long-term educational investment; the effect may be stronger on younger audiences (Eyal & Te'eni-Harari, 2024)

Sources: *Porter & Wood (2021); Liu & Wang (2025); Jia & Lee (2024); Annenberg Center for Public Policy (2017); Ecker et al. (2024); Swire-Thompson et al. (2022); Han et al. (2025); Johansson et al. (2023); Roozenbeek et al. (2020); Guess et al. (2020); Brown (2022); Eyal & Dean-Harari (2024); Author analysis, processed*

It should be understood that every change in attitude and behavior also has indicators. These KPIs concern aspects related to the level of trust in institutions, the level of social division, and actual behavior such as voting decisions or vaccine acceptance. A number of studies have given positive results at this level, the impact is generally small, and it is unlikely to be compared to the perception of accuracy (Bowles et al., 2025).

Research methodology and its social impact still have significant gaps. The most causally rigorous experimental methods, such as laboratory experiments and controlled surveys, tend to have small social outcomes. The most significant social impacts are quite the opposite, such as reduced public trust, political polarization, anarchic actions, difficult to measure causally and can usually only be analyzed using correlation or qualitative approaches. The limited ability of researchers to test the direct relationship between extreme actions resulting from misinformation meant that this study assessed how accurate participants were in judging news headlines. Methodological disparities suggest that the presence of strong evidence that certain interventions can affect confidence in a controlled environment, the existence of evidence related to the impact of preventing massive and widespread social consequences remains limited.

Table 2 shows information on the various interventions, psychological mechanisms, and their effectiveness against disinformation. These differences highlight the types of strategies that work through certain psychological mechanisms and the application of contexts that affect the results. Understanding how each intervention and mechanism has limitations is essential before proceeding with the discussion of the challenges and obstacles in their implementation.

Implementation Challenges and Barriers

Many strategies have been found to be effective in experimental studies, but their implementation still faces a number of challenges. Obstacles and challenges regarding implementation are divided into scalability, adversarial adaptation, boomerang effects, and unexpected side effects.

- **Scalability:** The dissemination of information in a huge digital platform based on speed and volume. Human involvement in its interventions, such as in-depth fact-checking or intensive media literacy programs, illustrates that it is very difficult to balance and scale up (Roozenbeek et al., 2023).
- **Boomerang Effect:** What practitioners worry about is that correcting misinformation can backfire, reinforcing false beliefs, especially between highly fanatical individuals. However, the emerging scientific consensus on the effect of a boomerang can increase misconceptions is rare (Swire Thompson et al., 2022). Many studies show that corrections generally result in reduced public trust in misinformation (van der Meer et al., 2023). The first study concluded that the boomerang effect was likely influenced by the less reliable measurements used, which gave false results due to the reckle to the mean (Swire-Thompson et al., 2022). However, the echo of affective beliefs persists even though individuals have updated their factual understanding (Sangalang et al., 2019).
- **Enemy Adaptation:** An adaptive and dynamic enemy depicts the act of disinformation. They evolved with a variety of techniques, tactics, and procedures (TTPs) with the goal of avoiding detection and counter-intervention (Roozenbeek et al., 2023). As the platform developed to be better at detecting bots, the perpetrators switched to using troll farms operated directly by humans. Audiences sensitive to vulgar "fake news," perpetrators are turning to more subtle forms of malinformation, by distorting accurate information out of context.
- **Unwanted Side Effects:** Interventions of course also have negative consequences. Several studies have concluded that audience trust can be eroded by excessive pressure on disinformation threats, including credible and high-quality sources. General cynicism can grow and trust in the media decreases (Kozyreva et al., 2023).

The application of counter-strategies in digital information warfare and propaganda remains challenging, not having a single perfect solution. To create an effective strategy, one must consider scalability, enemy adaptation, boomerang effects, and the potential for unexpected side effects. The multidimensional approach is considered the most optimal by combining technology, interventions, policies to address disinformation patterns, and evaluation with continuous

adjustments to remain relevant and effective. At this stage, it is necessary to understand before discussing governance, ethics, and the consequences of implementing counter-disinformation.

3.4 Governance, Ethics, and Consequences

Politically or ethically, interference in the media is never neutral. Any attempt to confront and control propaganda can have an unexpected impact. The following chapters will discuss polemics in counter-disinformation strategies that focus on the tension between freedom and security of expression, privacy issues, and the potential misuse of existing anti-disinformation tools.

Privacy and Surveillance Issues

Individual privacy can be threatened by the methods used to combat propaganda. Efforts to eradicate propaganda can be done through the collection and analysis of personal data using micro-targets. Data on behavior are used to identify targets and vulnerable populations with media interventions. However, this practice is considered incompatible with privacy and ethics, as the data collected does not have the consent of the relevant individual (Tucker, 2020). The data collected is not controlled by companies and governments capable of creating "surveillance capitalism" (Thomson Reuters Foundation & Tow Center for Digital Journalism, 2023). Disinformation is diverted with the aim of domestic surveillance of citizens. The design of tools to track bots and foreign trolls can be used to monitor political activists or journalists. This creates a slippery slope towards a "surveillance state", where individual freedoms are eroded under the pretext of national security (Thomson Reuters Foundation & Tow Center for Digital Journalism, 2023). Recommendations for such solutions are the implementation of strong personal data protection laws that can restrict governments from accessing citizens' personal data, which usually requires intervention from judicial oversight (Thomson Reuters Foundation & Tow Center for Digital Journalism, 2023). It is not enough just to protect data, because laws and regulations regarding anti-disinformation can be misused for the political advantage of certain groups.

"Anti-Disinformation" weaponry

Instead of functioning as a mechanism to protect the public interest in order to protect the truth, today the fight against disinformation that should be carried out by the relevant stakeholders is used to achieve oppressive political goals by making it a legitimacy of power. Mahapatra et al (2024) state that in countries that adhere to authoritarian and liberal regimes, they respond to this by passing "anti-fake news" laws. They believe that this law can be used to intimidate and imprison journalists, activists, and political opposition positions. In the same source, it is said that the response to the phenomenon has fostered the term "Anti-Fake News Lawfare" (AFNL) which refers to the granting of the authority to prosecute all criticism of the government by "fake news" or "disinformation". Therefore, fact-checking organizations have a big role in this. However, as Bowles et al (2025) say, this organization may have a great potential for habituation by accident or intention. For this reason, fact-checking organizations, according to Jia & Lee (2024), must focus on transparency and objectivity.

Easy to twist and used to undermine democracy, laws, content moderation, and fact-checking are used to suppress dissent and erode freedom of speech. Harefa & Laila (2024) argue that "fake news" that seems to be supported by law is defined as disinformation that is actually a threat to a country. Furthermore, the same thing was also expressed by Mahapatra et al (2024) and responded to the existence of disinformation supported by the law that can later be used by the ruling government to target political criticism or investigative journalism. Therefore, to prevent misuse of information, any anti-disinformation intervention must be designed with strong procedural safeguards.

4. Conclusions and Recommendations

Multidimensional threats in the cognitive realm are a consequence of the development of information warfare and disinformation. This form of multidimensional threat targets public perception and belief through the role of social media, including content manipulation, digital algorithms, and integrated actor operations. In overcoming this, a response that has an adaptive, collaborative, and integrated strategy with ethical policies and governance according to the principles of proportionality, transparency, and accountability is needed. Responses to multidimensional threats

can be pursued by increasing cognitive resilience through media literacy, fact-checking, and psychological immunization. However, there are still limitations in the form of scalability, methodological gaps, and adaptive dynamics of disinformation faced by perpetrators.

This research has a gap that emphasizes the need for new strategies to solve the complexity of threats in the digital age such as the phenomenon of artificial intelligence and social algorithms. The development of technologies such as AI, VR, and AR can trigger the presence of new threats through the excessive delivery of disinformation and personalized to the public amid empirical validity and replication that are increasingly hampered by limited access to digital *platform* data.

In a methodological context, longitudinal and cross-contextual experimental research is becoming increasingly important because too much of the dominance of *cross-sectional studies* has not been able to reveal more about the ongoing effects of disinformation and cognitive phenomena gradually. Theoretically, while theoretically, the relationship between human cognitive bias, social network dynamics, and recommendation algorithms is no longer sufficient to be explained through classical mass communication models.

The focus in strategic handling efforts to deal with the information war and digital infrastructure includes: (1) building public cognitive resilience through the application of layered approaches that can harmonize technology, social intervention, and education; (2) strengthening cross-sectoral cooperation, namely between the government, digital platforms, academics, media, and civil society in using secure and ethically guided data; (3) formulate ethical anti-disinformation governance and regulations through procedural safeguards, restrictions on access to personal data, and independent oversight procedures; (4) assessing the effectiveness of interventions and adjusting strategies through the development of a sustainable evaluation and adaptation system; and (5) improve the public's ability to recognize, assess, and reject disinformation by strengthening individual and collective capabilities through media literacy, psychological immunization, and critical digital education.

References

Al Fatih, I. Z., Putera, R. A., & Umar, Z. H. (2024). Peran algoritma media sosial dalam penyebaran propaganda politik digital menjelang pemilu. Jurnal Kajian Stratejik Ketahanan Nasional, 7(1), 6.

Alaminos-Fernández, A. F. & Alaminos-Fernández, P. (2025). Democratic Resilience Under Strain: Threats and Public Trust in Greece. Athens Journal of Social Sciences, 12 (4): 259–282. https://doi.org/10.30958/ajss.12-4-3.

Almufarrij, I., & Tolia, N. (2023). You've been fact-checked! Examining the effectiveness of social media fact-checking against the spread of misinformation. Available from: https://www.researchgate.net/publication/372995024_You've_Been_Fact-Checked_Examining_The_Effectiveness_of_Social_Media_Fact-checking_Against_the_Spread_of_Misinformation [Accessed 3 November 2025].

Alshuwaier, F. A. & Alsulaiman, F. A. (2025). Fake news detection using machine learning and deep learning algorithms: A comprehensive review and future perspectives. Computers, 14 (9): 394. https://doi.org/10.3390/computers14090394

American Public University. (2025). Why is digital media literacy important to us and society? Available from: https://www.apu.apus.edu/area-of-study/business-and-management/resources/why-is-digital-media-literacy-important/ [Accessed 3 November 2025].

Annenberg Public Policy Center. (2017). Debunking study suggests ways to counter misinformation and correct 'fake news'. Annenberg School for Communication, University of Pennsylvania. Available from https://www.asc.upenn.edu/news-events/news/debunking-study-suggests-ways-counter-misinformation-and-correct-fake-news [Accessed 3 November 2025]

Arguedas, A. R., Robertson, C. T., Fletcher, R. & Nielsen, R. K. (2022). Echo Chambers, Filter Bubbles, and Polarisation: A Literature Review. London: The Royal Society. Available from https://royalsociety.org/-/media/policy/projects/online-information-environment/oie-echo-chambers.pdf [Accessed 3 November 2025].

Azzimonti, M., & Fernandes, M. (2023). Social media networks, fake news, and polarization. European Journal of Political Economy, 76, 102256. https://doi.org/https://doi.org/10.1016/j.ejpoleco.2022.102256.

Bachtiar, A., Dahlian Persadha, P., & Supriyadi, E. (2023). Propaganda Intelijen Melalui Media Sosial Dalam Mendukung Perpindahan Ibu Kota Negara. Innovative: Journal Of Social Science Research, 3(5), 8696–8709. Retrieved from https://j-innovative.org/index.php/Innovative/article/view/5221.

Banks, W. (2021). Cyber Attribution and State Responsibility. International Law Studies, 97 (1): 1040-1068. Available from https://digital-commons.usnwc.edu/cgi/viewcontent.cgi?article=2980&context=ils [Accessed 3 November 2025].

Bjola, C., & Papadakis, K. (2020). Digital propaganda, counterpublics and the disruption of the public sphere: the Finnish approach to building digital resilience. Cambridge Review of International Affairs, 33(5), 638–666. https://doi.org/10.1080/09557571.2019.1704221

Bowles, J., Croke, K., Larreguy, H., Liu, S. & Marshall, J. (2025). Sustaining exposure to fact-checks: misinformation discernment, media consumption, and its political implications. American Political Science Review. https://doi.org/10.1017/S0003055424001394.

Brown, S. (2022). Study: Digital literacy doesn't stop the spread of misinformation. MIT Sloan School of Management: Ideas Made to Matter. Available from https://mitsloan.mit.edu/ideas-made-to-matter/study-digital-literacy-doesn't-stop-spread-misinformation [Accessed 3 November 2025].

Campbell, J. (2024). Free speech and the regulation of social media content. Available from https://www.ebsco.com/research-starters/communication-and-mass-media/free-speech-and-regulation-social-media-content [Accessed 3 November 2025].

Christou, G. (2024). Cyber diplomacy: From concept to practice. Tallinn Papers. NATO Cooperative Cyber Defence Centre of Excellence (CCDCOE). Available from: https://ccdcoe.org/uploads/2024/06/Tallinn_Papers_Cyber_Diplomacy_From_Concept_to_Practice_Christou.pdf [Accessed 3 November 2025].

Courter, I. J. (2024). US Military Doctrine Treats Information and Influence as the Same Thing—and That's a Problem. Modern War Institute. Available from https://mwi.westpoint.edu/us-military-doctrine-treats-information-and-influence-as-the-same-thing-and-thats-a-problem/ [Accessed 3 November 2025].

Creswell JW, Poth CN. (2016). Qualitative inquiry and research design: Choosing among five approaches. Sage publications.

Cyber Diplomacy Toolbox. (2025). What is cyber diplomacy? Available from https://www.cyber-diplomacy-toolbox.com/Cyber_Diplomacy.html [Accessed 3 November 2025].

Deng, E., Cho, T., Shah, D., & Lomborg, S. (2024). "Community guidelines make this the best party on the internet": An in-depth study of online platforms' content moderation policies. arXiv preprint arXiv:2405.05225. Available from: https://arxiv.org/html/2405.05225v1 [Accessed 3 November 2025].

Dinfos Pavilion Team. (2025). Types of Information Disorder. Available from: https://pavilion.dinfos.edu/Article/Article/3582943/types-of-information-disorder/ [Accessed 3 November 2025].

Druckman, J. N. (2022). Threats to Science: Politicization, Misinformation, and Inequalities. The Annals of the American Academy of Political and Social Science, 700 (1): 8–24. doi:10.1177/00027162221095431.

Ecker, U. K. H., Prike, T., Paver, A. B., Scott, R. J., & Swire-Thompson, B. (2024). Don't believe them! Reducing misinformation influence through source discreditation. Cognitive Research: Principles and Implications 9(1): 52. https://doi.org/10.1186/s41235-024-00581-7.

European Commission. (2025). Countering information manipulation. Available from https://commission.europa.eu/topics/countering-information-manipulation_en [Accessed 3 November 2025].

Eyal, K. & Te'eni-Harari, T. (2024). Systematic review: Characteristics and outcomes of in-school digital media literacy interventions, 2010-2021. Journal of Children and Media, 18 (1): 8-28. https://doi.org/10.1080/17482798.2023.2265510

Feldman, D. K. D., & Nahmias, Y. (2025). From bots to ballots: Democratic integrity in the era of digital manipulation. Minn. JL Sci. & Tech., 26 (1): 229-292. https://scholarship.law.umn.edu/cgi/viewcontent.cgi?article=1573&context=mjlst.

Feng, G. C. (2025). Inoculation theory revisited: reinterpretations, extensions, and new directions. Frontiers in Communication, 10: 1576772. https://doi.org/10.3389/fcomm.2025.1576772

Finlay, L. & Payne, C. (2019). The attribution problem and cyber armed attacks. American Journal of International Law, 113: 202-206. https://doi.org/10.1017/aju.2019.35.

González-Bailón, S., & De Domenico, M. (2021). Bots are less central than verified accounts during contentious political events. Proceedings of the National Academy of Sciences, 118(11), e2013443118. https://doi.org/10.1073/pnas.2013443118

Foundation for Individual Rights and Expression. (2025). Free speech and social media. Available from https://www.thefire.org/research-learn/free-speech-and-social-media [Accessed 3 November 2025]

Guess, A. M., Lerner, M., Lyons, B., Montgomery, J. M., Nyhan, B., Reifler, J., & Sircar, N. (2020). A digital media literacy intervention increases discernment between mainstream and false news in the United States and India. Proceedings of the National Academy of Sciences of the United States of America, 117 (27), 15536-15545. https://doi.org/10.1073/pnas.1920498117

Han, X., Liao, K., Chen, Y., Jin, X., & Han, W. (2025). Mechanism and long-term effects of psychological inoculation on individuals' attitudes towards commercial misinformation: Experimental research. Information Technology & People 39(4): —. https://doi.org/10.1108/ITP-10-2024-1340.

Harefa, A., & Laila, F. (2024). Criminal Law Implication on the Spread of Hoax News on Social Media in the Perspective of ITE Law. Journal of Strafvordering 1(5): 1-10. https://doi.org/10.62872/10z57x15

Hill, A. G. (2019). The Ultimate Challenge: Attribution for Cyber Operations. Wright Flyer Paper No. 70. Maxwell AFB, AL: Air University Press.

Iacovitti, G. (2022). How technology influences information gathering and information spreading. "Church, Communication and Culture," 7(1), 76–90. https://doi.org/10.1080/23753234.2022.2032781

Institute for the Study of War. (2025). A primer on Russian cognitive warfare. Understanding War. Available from https://understandingwar.org/research/cognitive-warfare/a-primer-on-russian-cognitive-warfare [Accessed 3 November 2025].

Jacobs, L. (2023). Fighting disinformation online: The Digital Services Act in the European Union. Justice Studies Faculty Publications, Montclair State University. Available from: https://digitalcommons.montclair.edu/cgi/viewcontent.cgi?article=1248&context=justice-studies-facpubs [Accessed 3 November 2025].

Jia, C. & Lee, T. (2024). Journalistic interventions matter: Understanding how Americans perceive fact-checking labels. HKS Misinformation Review, 5 (2): 1-17. https://doi.org/10.37016/mr-2020-138.

Johansson, P., Enock, F., Hale, S., Vidgen, B., Bereskin, C., Margetts, H., & Bright, J. (2023). How can we combat online misinformation? A systematic overview of current interventions and their efficacy. The Alan Turing Institute. Available from https://www.turing.ac.uk/sites/default/files/2023-11/tackling_online_misinformation_report_2023_v3.pdf [Accessed 3 November 2025

Kovic, M., Rauchfleisch, A., Sele, M. & Caspar, C. (2018). Digital astroturfing in politics: Definition, typology, and countermeasures. Studies in Communication Sciences, 18 (1): 69–85. DOI: 10.24434/j.scoms.2018.01.005.

Kowalski, M. (2023). Responses to digital disinformation as part of hybrid threats: A conceptual review. Security and Defence Quarterly, 43(3): 26–39. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC10446021/ [Accessed 3 November 2025].

Kozyreva, A., Herzog, S. M., Lewandowsky, S., Hertwig, R., Lorenz-Spreen, P., Leiser, M., & Reifler, J. (2023). Resolving content moderation dilemmas between free speech and harmful misinformation. Proceedings of the National Academy of Sciences of the United States of America 120(7): 1-12. https://doi.org/10.1073/pnas.2210666120

Kozyreva, A., Herzog, S.M., Lewandowsky, S., Hertwig, R., Lorenz-Spreen, P., Leiser, M. & Reifler, J. (2023). Resolving content moderation dilemmas between free speech and harmful misinformation. Proceedings of the National Academy of Sciences of the United States of America, 120 (7): 1–10. https://doi.org/10.1073/pnas.2210666120.

Lahmann, H., Custers, B., & Scott, B. I. (2025). The fundamental rights risks of countering cognitive warfare with artificial intelligence. Ethics and Information Technology 27(4): 49. https://doi.org/10.1007/s10676-025-09868-9

Liu, Y. & Wang, S. 2025. The nudging effect of fact - checking labels: an inquiry based on the platform context. Social Science. 1-16. https://doi.org/10.20944/preprints202503.1737.v1

Malyarenko, T., & Kormych, B. (2022). The barbarism of hybrid warfare. Kennan Institute/Wilson Center. Available from https://www.wilsoncenter.org/blog-post/barbarism-hybrid-warfare [Accessed 3 November 2025].

Marigliano, R., Ng, L. H. X., & Carley, K. M. (2024). Analyzing digital propaganda and conflict rhetoric: a study on Russia's bot-driven campaigns and counter-narratives during the Ukraine crisis. Social Network Analysis and Mining, 14(1), 170.

Marjanović, A., & Smiljaniç, D. (2025). Cognitive warfare – The human mind as the new battlefield. In Proceedings of the Defence and Security Conference, 1 (1) pp. 84-114). Zagreb, Croatia. Available from https://www.researchgate.net/publication/391704397_Cognitive_warfare_-_the_human_mind_as_the_new_battlefield [Accessed 3 November 2025].

Möller, J. (2021). Filter bubbles and digital echo chambers. In H. Tumber, & S. Waisbord (Eds.), The routledge companion to media disinformation and populism. Routledge. 92-100. https://doi.org/10.4324/9781003004431-10.

Nakayama, B. (2022). Democracies and the future of offensive (cyber-enabled) information operations. The Cyber Defence Review, 7(3): 49-65. Available from: https://cyberdefensereview.army.mil/Portals/6/Documents/2022_summer_cdr/04_Nakayama_CDR_V7N3_Summer_2022.pdf [Accessed 3 November 2025].

NATO. (2005). DEEP Portal 2: Troll Factories [Infographic]. Available from: https://www.nato.int/nato_static_fl2014/assets/pdf/2020/5/pdf/2005-deepportal2-troll-factories.pdf [Accessed 7 November 2025].

NATO. (2020). DEEP ADL Portal: Troll factories. Available from: https://www.nato.int/nato_static_fl2014/assets/pdf/2020/5/pdf/2005-deepportal2-troll-factories.pdf [Accessed 3 November 2025].

NATO. (2025). NATO's approach to counter information threats. Available from: https://www.nato.int/cps/en/natohq/topics_219728.htm [Accessed 3 November 2025].

Nieubuurt, J. T. (2021). Internet memes: Leaflet propaganda of the digital age. Frontiers in Communication, 5, 547065.

Paziuk, A., Lande, D., Shnurko-Tabakova, E., & Kingston, P. (2025). Decoding manipulative narratives in cognitive warfare: A case study of the Russia–Ukraine conflict. Frontiers in Artificial Intelligence, 8: 1566022. https://doi.org/10.3389/frai.2025.1566022.

Petrov, I., & Kumar, R. (2025). Disinformation and legal responsibility: Regulating digital speech without curtailing dissent. Interdisciplinary Studies in Society, Law, and Politics, 4 (2): 303-314. https://doi.org/10.61838/kman.isslp.4.2.26

Porter, E. & Wood, T. J. (2021). The global effectiveness of fact-checking: Evidence from simultaneous experiments in Argentina, Nigeria, South Africa, and the United Kingdom. Proceedings of the National Academy of Sciences of the United States of America, 118 (37). https://doi.org/10.1073/pnas.2104235118

Prieto Curiel, R. (2021). Opinion Dynamics and the Inevitability of a Polarised and Homophilic Society (R. López-Ruiz (ed.)). IntechOpen. https://doi.org/10.5772/intechopen.96989

Princeton Public Library. (2025). Misinformation, Disinformation & Malinformation: A Guide. Available from: https://princetonlibrary.org/guides/misinformation-disinformation-malinformation-a-guide/ [Accessed 3 November 2025].

Qureshi, I., & Bhatt, B. (2024). Social media-induced polarisation. Information Systems Journal, 34(4).

Raba, I. (2025). Challenges in the attribution and regulation of potential state cyberattacks. Global Affairs, University of Navarra. Available from: https://en.unav.edu/web/global-affairs/desafios-en-la-atribucion-y-regulacion-de-posibles-ciberataques-estatales [Accessed 3 November 2025]

Rana, M. S., Nobi, M. N., Murali, B. & Sung, A. H. (2022). Deepfake detection: A systematic literature review. IEEE Access, 10: 25495 25513. https://doi.org/10.1109/ACCESS.2022.3154404

Roozenbeek, J., Culloty, E., & Suiter, J. (2023). Countering misinformation: Evidence, knowledge gaps, and implications of current interventions. European Psychologist, 28 (3): 189-205. https://doi.org/10.1027/1016-9040/a000492

Roozenbeek, J., van der Linden, S., & Nygren, T. (2020). Prebunking interventions based on "inoculation" theory can reduce susceptibility to misinformation across cultures. The Harvard Kennedy School (HKS) Misinformation Review, 1(2): 1-23. https://doi.org/10.37016//mr-2020-008

Rumer, E. (2019). The Primakov (Not Gerasimov) Doctrine in Action Carnegie Endowment for International Peace. Available from https://carnegieendowment.org/research/2019/06/the-primakov-not-gerasimov-doctrine-in-action?lang=en [Accessed 3 November 2025].

Sangalang, A., Ophir, Y., & Cappella, J. N. (2019). The potential for narrative correctives to combat misinformation. Journal of Communication 69(3): 298-319. https://doi.org/10.1093/joc/jqz014

Shattock, E. (2021). Self-regulation 2:0? A critical reflection of the European fight against disinformation. HKS Misinformation Review, 2(3). https://doi.org/10.37016/mr-2020-73

Swire-Thompson, B., Miklaucic, N., Wihbey, J. P., Lazer, D., & DeGutis, J. (2022). The backfire effect after correcting misinformation is strongly associated with reliability. Journal of Experimental Psychology: General 151(7): 1655-1665. https://doi.org/10.1037/xge0001131

Tashev, B., Purcell, M., & McLaughlin, B. (2019). Russia's Information Warfare: Exploring the Cognitive Dimension. MCU Journal, 10(2): 129-147. https://doi.org/10.21140/mcuj.2019100208.

Taylor, S. H., & Choi, M. (2022). An initial conceptualization of algorithm responsiveness: Comparing perceptions of algorithms across social media platforms. Social Media+ Society, 8(4), https://doi.org/20563051221144320.

Thomson Reuters Foundation & Tow Center for Digital Journalism. (2023). Weaponising the law: Attacks on media freedom. Available from https://www.medialegalattacks.com/ [Accessed 3 November 2025]

Timotheou, S., Miliou, O., Dimitriadis, Y., Sobrino, S. V., Giannoutsou, N., Cachia, R., Monés, A. M., & Ioannou, A. (2023). Impacts of digital technologies on education and factors influencing schools' digital capacity and transformation: A literature review. Education and Information Technologies, 28(6), 6695–6726.

Tucker, P. (2020). Why data privacy is crucial to fighting disinformation. Defense One. Available from https://www.defenseone.com/technology/2020/12/why-data-privacy-crucial-fighting-disinformation/170445/ [Accessed 3 November 2025].

U.S. Department of State. (2025). Report to Congress on efforts to combat disinformation of foreign adversaries and authoritarian regimes. Available from: https://www.state.gov/wp-content/uploads/2025/08/Report-Efforts-to-Combat-Disinformation-of-Foreign-Adversaries-006070-Accessible-07.15.202.pdf [Accessed 3 November 2025].

Ufen, A. (2024). The rise of digital repression in Indonesia under Joko Widodo. GIGA Focus Asia 1/2024. GIGA – Hamburg. Diakses 3 November 2025 dari https://www.giga-hamburg.de/en/publications/giga-focus/the-rise-of-digital-repression-in-indonesia-under-joko-widodo

Van der Meer, T. G. L. A., Hameleers, M., & Ohme, J. (2023). Can fighting misinformation have a negative spillover effect? How warnings for the threat of misinformation can decrease general news credibility. Journalism Studies 24(6): 803-823. https://doi.org/10.1080/1461670X.2023.2187652

Weiss, M., & Biermann, F. (2023). Cyberspace and the protection of critical national infrastructure. Journal of Economic Policy Reform, 26(3), 250–267. https://doi.org/10.1080/17487870.2021.1905530

Wikipedia. (2024). Psychological operations (United States). Wikipedia. Available from https://en.wikipedia.org/wiki/Psychological_operations_%28United_States%29 [Accessed 3 November 2025].

Yang, Y.-T. & Zhu, Q. (2025). Toward a Multi-Echelon Cyber Warfare Theory: A Meta-Game-Theoretic Paradigm for Defense and Dominance. arXiv preprint arXiv:2509.08976v1. Available from: https://arxiv.org/abs/2509.08976.

Yovana, K. & Gatra, A.R.P.P. (2022). Strategi Kampanye Donal Trump pada Pemilihan Umum Amerika Serikat. Moestopo Journal International Relations (MJIR), 2(1), 1-16.

Zhu, Y., & Fu, K. (2024). How Propaganda Works in the Digital Era: Soft News as a Gateway. Digital Journalism, 12(6), 753–772. https://doi.org/10.1080/21670811.2022.2156366.