COGNITIVE WARFARE IN THE AGE OF AI: A DESK-BASED STUDY THROUGH NARRATIVE REVIEW, STRATEGIC CASE ANALYSIS, AND GLOBAL POLICY EVALUATION.

SH Muansinghe[1] and R Rajapakse[2]

Abstract

This study theorizes cognitive warfare as an intentional and organized form of psychological exploitation, intentionally intensified by artificial intelligence (AI) and used by sovereign states to disrupt democratic solidarity, erode epistemic trust, and change socio-political narratives. We want to show cognitive warfare, not just as information distortion, but as an organized system acting on human cognition through automation, affect, and narrative. This study utilized a desk-based approach and seeks to blend a narrative literature review with a comparative case study and a synthesis of policy documents. It studies three cases focusing on the cognitive or psychological element: Russia's disinformation activities for democratic elections; AI-enabled psychological operations in the Russia–Ukraine war; and China's long-range influence campaign aimed at political culture and democratic resilience in Taiwan. The findings identify a convergent architecture of enhanced cognitive warfare across the three cases which involved (1) technological convergence that consists of synthetic media and algorithmic amplification; (2) psychological manipulation of the audience's emotional and cognitive complex of biases, namely fear (e.g., Jan 6th insurrection), repetition (e.g., slogans, denier hashtags), and identity appeal; and (3) narrative engineering designed to change collective memory, and legitimacy and trust in institutions. These operations are beyond the ambit of the term "cybersecurity", which is usually concerned with technical stability and infrastructure, not cognitive vulnerabilities. This paper calls for security and defense strategies to go back and reposition a new security paradigm around psychological resilience, epistemic integrity, and informational sovereignty. It calls for an interdisciplinary approach to study these intensively AI amplified psychological operations in contemporary hybrid conflict by bringing together cognitive science, algorithmic accountability, and strategic communications.

Keywords: Artificial intelligence, Cognitive security, Disinformation, Hybrid warfare, Psychological manipulation

[1]Department of Statistics & Computer Science, University of Kelaniya, Sri Lanka

Email: hansi.munasinghe12@gmail.com https://orcid.org/0009-0001-7276-6860

[2]School of Computer Science, Faculty of Engineering, University of Sydney.

Email: saumya.rajapakse@sydney.edu.au https://orcid.org/0000-0002-9851-7774

Received date: 27.06.2025 Accepted date: 22.08.2024

Print Publishing Date: 31.10.2025 Web Publishing Date: 31.10.2025

Introduction

In an interconnected and digitally mediated world, warfare has changed drastically. Traditional military matters have become one method of trying to impose control over narratives, perception, and cognition; cognitive warfare is one term to describe the efforts to prepare for warfare considering how warfare may capitalize on people's cognition, e.g., attention, memory, choosing, emotion, etc and ultimately impact behavior and beliefs (NATO Innovation Hub, 2021). Cognitive warfare is a non-kinetic approach that operates within information environments, including social media and AI-mediated technologies, and employs psychological operations. Similarly, elements of cognitive warfare can occur over long periods, unnoticed, and transnationally. Artificial intelligence offers a distinct advantage in making cognitive warfare easier and faster. For example, elements of AI could include deepfakes along with systems such as recommender algorithms, emotion recognition agents, and autonomous influence bots that can be used to influence public discourse, temporality, agency, shared reality, and public trust (Brundage et al., 2018).In a not-too-far-fetched fashion, one could imagine a powerful AI that propagates disinformation; potentially at the level of inundation, but with psychological precision employing an underlying emotional basis and cognitive approach as an example of possible belief preference to persuade(Tversky & Kahneman, 1974).

Literature Review

Cognitive psychology and behavioral science have identified several mechanisms that allow us to be influenced. There is classic research by Tversky and Kahneman (1974) on heuristics and biases showing that human thought often reduces complex decision-making processes to simple mental shortcuts. However, this leads to poor predictions when uncertainty is present. Pennycook and Rand (2019) suggest that the tendency to be cognitively lazy and favor intuitive thinking over analytical thinking enables the spread of misinformation. In the digital media, Lewandowsky, Ecker, and Cook (2017) argue that 'emotionalized lies' are believed and shared more readily than equally verified factual information, especially given the context of information overload and reduced confidence in sources of knowledge. Such vulnerabilities are being intentionally designed to be exploited in global political contexts. For instance, in the 2016 U.S. presidential election, Russia's use of AI-enabled bot networks and micro-targeted misinformation polarized the electorate and undermined democratic institutions (Jamieson, 2018; Mueller, 2019). In East Asia, China has employed strategic influence operations to shape attitudes about Taiwan through coordinated media and messaging. Finally, narratives developed from AI during the Russia-Ukraine war have been designed to undermine institutions, confuse the reality of the situation, and sway international opinion (Marigliano, Ng, & Carley, 2024; Mandic & Klaric, 2023; Bozalka, 2023). Despite increasing awareness of these threats, the international response is still largely reactive and fractured. Organizations such as NATO, the European Commission, and the UN have acknowledged and recognized the role of cognitive and information security in a strategic competition environment (European Commission, 2018; Dusan, 2023), but the responses to these threats are often overly simplistic, lack interoperability, and do not include technology, interdisciplinary frameworks, or long-term thinking.

Research Problem

While multiple studies address components of cognitive warfare from AI ethics and information systems to media manipulation and behavioral responses, no integrated synthesis systematically brings together insights from psychology, AI, and geopolitics. This fragmentation limits the formulation of effective international policy and hampers public awareness and preparedness.

Research Objectives

This study addresses this critical gap by conducting a desk-based, multidisciplinary investigation of cognitive warfare in the era of artificial intelligence. The research aims to:

1. Synthesize scholarly literature on AI-enabled cognitive manipulation and human psychological vulnerabilities;
2. Analyze key geopolitical case studies that illustrate real-world deployment of cognitive warfare tactics;
3. Evaluate existing international policy frameworks to identify gaps, weaknesses, and areas for coordinated governance.
4.

Research Strategy

The research is exploratory and integrative, with an explicit goal of combining existing, dispersed knowledge related to cognitive psychology, artificial intelligence, information warfare, and strategic studies. The advantage of desk-based research occurs when the research aim is to organize, re-conceptualize, and critically evaluate existing knowledge, rather than to collect new empirical data (Snyder, 2019). (All data utilized is from recognized, reputable, and verifiable secondary sources - peer-reviewed journal articles, institutional white papers, investigative journalism, and official policy literature. This constitutes the best practice for qualitative desk-based

research where the aim is not generalizability through sampling; instead, it is conceptual richness and practical depth (Baumeister & Leary, 1997).

Methodology

The research design, which was of qualitative, desk-based approach, was based on an interpretive analysis. Information was taken from three major areas of primary sources:

1. Academic sources on cognitive warfare, disinformation, AI-driven manipulation, and cyber psychology that have been peer-reviewed.
2. Case reports are publicly available in sources of established institutions and governments.
3. Think tanks and international organizations' strategic policy works (e.g., NATO, EU, RAND).

Inclusion and Exclusion Criteria

Sources were purposively sampled using four criteria:

- Relevance to cognitive warfare or AI-enabled manipulation
- Credibility of the source (e.g., peer-reviewed, government-issued, or institutional publication)
- Influence on discourse or policymaking

The exclusion criteria were limited to materials that were not published in English, duplicate records, papers that are not affiliated with an institution, and/or institutional affiliation and academic rigor, papers that are not transparent to be authored by authorship.

Thematic Analysis and Coding Protocols:

Multi-phase coding method was applied to the analysis of policy documents:

- Open Coding: Terms of interest were identified in the first segments to obtain issues of cognitive warfare governance (e.g., surveillance, trust, psychological influence).
- Axial Coding: Codes were then clustered to determine their relationship, i.e., the incorporation of AI tools and legal uncertainty.
- Selective Coding: Synthesis of core themes was carried out, such as the narrative control and governance vacuums and strategic ambiguity.

They used the six-stage thematic analysis model developed by Braun and Clarke **(**Braun & Clarke, 2006**)** in this process:

1. The introduction of the data
2. Creation of codes at the first level
3. The quest for themes
4. Reviewing themes
5. Theme characterization and labeling
6. Preparation of the report

To reach parity and to triangulate themes in various contexts, the cross-comparison of the case studies and the narrative literature was conducted.

Triangulation and Limitations.

The validity of interpretation was confirmed through triangulation of literature, case data, and policy monitoring that limited the bias of uniquely based sources. The research is, however, restricted by the source of secondary research because it was not field-tested. The results do not have statistical generalizability but are supposed to promote conceptual generalizability (Snyder, 2019; Baumeister & Leary, 1997). The emphasis has been on methodological transparency, which is, however, limited by levels limiting reproducibility because of the availability of different sources in different periods and the interpretive nature of thematic coding.

Findings from the Narrative Literature Review

Cognitive warfare represents a major epistemological shift in the study of conflict that seeks to replace traditional models centered on the use of physical force or the conquest of territory with strategies that aim to penetrate and manipulate cognition. Instead of seeking to manipulate geography and infrastructure, adversaries seek to manipulate attention, as well as distort memories and re-engineer belief systems in the wake of advances in artificial intelligence (AI), behavioral science, and algorithmic content consumption. This narrative review integrates literature across cognitive psychology, computational propaganda, and security studies to outline the systematic vulnerability of human cognitive architecture within digitally saturated contexts.

The seminal work of Tversky and Kahneman (1974) on judgment in uncertain circumstances helped establish that human thinking deviates systematically from normative logic due to cognitive heuristics–for example, availability

bias, anchoring, and representativeness. These heuristics minimize cognitive load but introduce consistent biases into judgments and decisions, especially as the level of uncertainty, emotional (hot) state, or time constraints increase (Kahneman, 2011). These biases are amplified in the digital information environment. There is considerable evidence that cognitive overload, given the sheer frequency and volume of information shared online, impairs the brain's ability to engage in the kind of deliberative thinking presumed to be more reflective (Lewandowsky, Ecker, & Cook, 2017; Swire-Thompson & Lazer, 2020). As cognitive loads increase, individuals tend to rely on fast, heuristic thinking (Type 1 thinking) rather than the slower, analytic - and more methodical, and often normatively correct use of Type 2 thinking: Individuals will accept false information and conspiracy theories as long as it fits their previously held beliefs; as revealed through the social psychological processes that underlie social network echoes (Stanovich & West, 2000; Vosoughi, Roy, & Aral, 2018).

A significant psychological limitation is the continuous impact effect - or the idea that misinformation continues to distort memory and reasoning, even if it is explicitly rejected (Lewandowsky et al., 2012). The illusory truth effect has also been shown to enhance perceived truth simply by repetition (Fazio et al., 2015). Factors such as this help to explain why disinformation campaigns continue to influence cognition, influenced by AI, once they have become entrenched despite efforts at intervention and fact-checking. Penneycook & Rand (2019, 2020) demonstrate via empirical data that susceptibility to misinformation is more closely related to cognitive reflection than partisan leanings; thus, cognitive warfare is deploying a common mode of psychology not contingent on ideological split. As Roozenbeek & van der Linden (2019) suggest, a potential inoculation approach may still have limited reach and scale as a last-ditch effort to build resistance to persuasion, misinformation, and disinformation.

At the same time, it is recognized that the role of AI and algorithmic systems in turning cognitive warfare into practice is well established. For example, social media applications use AI to speed the delivery of content based on engagement metrics (i.e., clicks, shares, likes), rather than the epistemic truth of the content (Benkler, Faris, & Roberts, 2018; Zuboff, 2019). The result is a feedback loop in the form of algorithmically curated epistemologies, where emotionally engaging content is prioritized for amplification, resulting in biased belief reinforcement and accelerating polarization (Bakir & McStay, 2018). Brundage et al. (2018) point out that AI technologies can have a dual use: features of AI designed to facilitate user personalization and experience can be repurposed to conduct automated psychological operations. For example, state-of-the-art deep learning models are now being used to generate synthetic images, video, and text (e.g., deepfakes; large language models) that allow high-fidelity impersonation and emotional manipulation on a mass scale (Chesney & Citron, 2019; Weidinger et al., 2021). Recently, research has turned its attention to the cognitive and social harms caused by synthetic media. Hancock and Bailenson (2021) claim that although deepfakes create deception, the more concerning outcome is their capacity to devalue the authenticity of real media and/or facts, an “epistemic uncertainty” or ambiguity in knowing what is true, which could lead to a breakdown of public trust, or in democratic discourse. This is especially concerning in the context of crises where misinformation that is emotional and timed is set to overwhelm deliberative processing (Buchanan, 2020; Ferrara, 2020).

Despite budding empirical literature, cognitive warfare exists in a conceptually disjointed sphere, with constructs of the related disciplines appearing across methodological and disciplinary divides. For example, cognitive psychologists and researchers focused on individual-level differences of susceptibility to misinformation (e.g., cognitive reflection and epistemic vigilance) (Pennycook & Rand, 2019), focused on the individual, and did not acknowledge that spaces of exposure are governed by an algorithmic infrastructure. Similarly, scholars writing in the ethics of AI and mutual security explicitly semanticize disinformation as a content governance and counter-threat issue and not at all addressed the psychological persistence of algorithmic persuasion, the detection of characteristic abuses, and the recourse to the system an ethical lens may provide (Brundage et al. and Cinelli et al., 2018). These divergent realms reflect critical disciplinary disjunctions that compound the study of how cognitively potent narratives, and heightened messaging induced by AI systems, travel through social media and ultimately affect public salience or belief formation. We need to connect these realms of study to develop a coherent theory of cognitive manipulation through AI.

In conclusion, the literature evidences that cognitive warfare is not a new concept; rather, it is a temporary merging of psychological understanding and accelerated technology. Here, cognitive rationalities, algorithmic amplification, and adversarial AI concepts come together to create a new information ecology, where the mind now becomes the new battlefield. This transition to cognitive warfare forces a change in security paradigms from cyber threat to cognition, or from controlling content to controlling cognitive control. Dealing with this threat will require a transdisciplinary research agenda that draws on cognitive psychology, AI ethics, media studies, and strategic communications. This research aims to contribute to that agenda through an integrative, desk-based

research synthesis of how vulnerabilities of cognition and AI systems co-evolve and lead to the new field of cognitive warfare.

Findings from Case Study Analysis

This section presents an evidence-based analysis of how state actors are using AI-enabled cognitive warfare in a geopolitical realm. The analysis is demonstrated through illustrative examples from Russia, Ukraine, and China, clearly presenting how operationalized digital platforms and algorithmic tools can be weaponized to manipulate cognition, develop public opinion, as well as disrupt sociopolitical conditions. In each case, we typify the situation as a distinct strategy: rapid disruption, real-time operational distortion, and identity erosion. In this way, we have illustrated that cognitive warfare is not a one-off approach, but a multimodal and evolving strategic doctrine.

Case A: Russia – Hybrid Information Warfare and Strategic Cognitive Disruption

Russia serves as a prominent example of hybrid cognitive warfare that uses disinformation and psychological operations, along with algorithmically based media manipulation, to undermine public trust and impact the adversary's decision-making environments. For example, during the 2014 annexation of Crimea, the Kremlin used a carefully structured operation that included the messaging of state-sponsored media, brandished manufactured threats that delegitimized the Ukrainian government, and combined these with military engagement (Pomerantsev & Weiss, 2014; NATO STRATCOM COE, 2016). The partnering of a fully realized Internet Research Agency (IRA) following the 2016 election in the U.S., increased the access to and potential success of this model, which coordinated influence operations and was based on social media platforms, troll farms, bots, and fabricated identities based on social and political divides - among class, race and religion, and immigrants (Mueller, 2019; Jamieson, 2018). These influence operations had the advantage of being supported by algorithms built into the platform, and the content strategy was driven by methodologies of social media content strategy framed on evoking and optimizing emotional engagement to become a destabilizing agent for epistemic erosion. If Russia and its use of hybrid cognitive warfare has not proven it yet, the patchwork of weaponized attention, trust, or belief has very much upset the trajectory of kinetic possibility, and toward non-kinetic possibilities for cognitively disruptive tactics in the next generation of strategic information warfare.

Using AI-powered bots and automated content agents, these operations increasingly tried to create some level of engagement and enter digital communities without being detected (Polyakova & Boyer, 2018). The broader intent was not just to spread misinformation, but to build cognitive ecosystems that created additional polarization, diminished trust in democratic institutions, and eroded collective rationality. Many of the disinformation narratives were shaped through algorithms to appeal to morality and identity-based predispositions; in this way, the disinformation could maximize psychological appeal by being salient emotionally and culturally. Social science has shown that the campaigns contributed to affective polarization, increased contention towards institutions, and further radicalized political discourse (Benkler, Faris, & Roberts, 2018; Tucker et al., 2018).

Case B: Ukraine Conflict – Deepfakes and Real-Time AI-Powered Propaganda

The war in Ukraine and Russia is a shining example of how AI (artificial intelligence) technologies are reshaping the cognitive environment of modern warfare. After the invasion in 2022, Russia's influence campaigns utilized Western automated systems and bot networks, as well as emotionally loaded messaging, to deliberately alter public perception of ideologies and disrupt the internal coherence of adversaries. Marigliano, Ng, and Carley (2024) used computational analysis to show that bot campaigns played an important role in circulating strategic narratives via digital media platforms, and these bots engaged users with rhetorical strategies that sought to polarize, destabilize, and allow for a constant reinforcement of pro-Russian messaging. Narrative posts were quickly shared and scripted to tap into and demonstrate sentiments of legitimation, sovereignty, and existential threat that would maximize engagement from both domestic and foreign audiences. Mandic and Klaric (2023) even documented that these operations engaged beyond information to embrace a significant psycho social targeting to activate or leverage against the emotional or cognitive vulnerabilities of the audience. Combined with moral and financial appeals, identity appeals, and repetition, the targeting of Russian disinformation campaigns was designed to avoid analytical resistance and achieve psychological disruption. Rather than persuading an audience through reasoning, these narratives made use of cognitive heuristics to incite activity, emotional arousal, and action that would threaten the perceived trustworthiness of democratic institutions and inhibit public resilience.

In addition, Bozalka (2023) locates these efforts in a broader context of AI-enabled information warfare, which he describes as information operations shifting from persuasion to cognitive overload, disorientation, and epistemic destabilization. The speed of Intel technologies allows them to congest information ecologies, disrupt attention economies, and diminish the scrutiny and critical reasoning of the public. Throughout the conflict in Ukraine, these operations allowed Russian-aligned media ecologies and armies of automated agents to employ a

besiegement of conflicting messages, affect-laden imagery, and violated narrative couplings of soundness, intention, and evidence to leave local populations and the global audience overwhelmed with cognitive bandwidth. Together, these sources point to Russia's information operations in Ukraine as employing a form of hybrid model of traditional propaganda with algorithmic scalability and cognitive (dis)engineering. This case study illustrates that the use of AI will not only change the sites of conflict in the digital age but will also change the psychological terrain of conflict.

Case C: China and Taiwan – Identity-Based Disinformation and Strategic Cultural Saturation

China's influence campaigns against Taiwan consist of a protracted period of cognitive warfare. These types of operations are markedly different from the shorter-cycle disruption operations conducted by Russia. Rather, the Chinese Communist Party (CCP) employs a network of digital efforts to influence national identity, cultural memory, or trust in democracy in Taiwan. The CCP takes action in conjunction with proxies such as media organizations and influencers, which include Key Opinion Leaders (KOLs) and news outlets, to avoid detection. This ecosystem stems from stoking narratives that emphasize unification, marginalized governing democracies (or democratic governance), or portray the people of Taiwan as throngs of divisive actors (Lee, 2021). Finally, the campaigns leverage the process of information laundering, where state-directed messages are encapsulated by trusted non-state intermediaries, such as KOLs and media influencers, to obscure the origins of salient narratives. This method increases perceived credibility and allows elite content to appear organic while reinforcing China's strategic themes of unification and democratic instability (Bush, 2021). Instead of encouraging a quick behavior change, China's approach is to cultivate long-term change in attitudes and use the illusory truth effect and narrative repetition to make slow changes to baseline beliefs. This is not about creating confusion, but fostering a strategic sense of inevitability and political fatigue, not just on what Taiwanese citizens believe to be true, but what they believe to be possible. In essence, in this model, cognitive warfare is a normative recalibration tool, and it can gradually realign public opinion with the strategic goals of the CCP.

Cross-Case Synthesis: Cognitive Warfare as a Strategic System

While the cases occurred at differing times and targets, each exhibited an underlying architecture of AI-enabled cognitive warfare:

1) Technological Convergence: State actors capitalized on AI applications expanding across deep fakes to NLP (natural language processing) bots to recommender systems to scale influence operations and utilize automation to persuade.

2) Psychological Exploitation: These campaigns manipulate emotional triggers (fear, anger, nostalgia) and cognitive biases (confirmation, repetition, and illusory truth) to operationalize these factors to short-circuit deliberative reasoning.

3) Narrative Engineering: The purpose of these campaigns was not simply to disseminate misinformation, but to recode and reposition interpretive frameworks to change what populations trust, remember, and emotionally connect to. Most importantly, these cases illustrate that traditional approaches to information security, upon which informatics, national, and cyber narratives are built, are grounded in relatively passive defensive thinking (e.g., malware, unauthorized access) and do not distinguish cognitive intrusions that impact perceptions or beliefs. Cognitive warfare is not a matter of devices or devices and traditional infrastructure; it concerns human cognition interfacing with and using machines. This underscores the need for new security paradigms to grow out of cognitive warfare: psychological resilience, epistemic integrity, and informational sovereignty.

Table 01: Comparative Analysis of AI-Enabled Cognitive Warfare Across Russia, China, and the United States

Element	Russia	China	United States
Objective	Destabilize democratic systems; erode public trust	Reinforce CCP narratives; suppress dissent	Preserve information integrity; promote cyber resilience
Modality of AI Use	Deep-fakes, bot farms, and micro-targeted content via troll networks	Mass surveillance, algorithmic censorship, narrative alignment	Public-private partnerships for detection and transparency mechanisms

Target Groups	Foreign publics (e.g., Ukraine, U.S., EU)	Domestic citizens and diaspora	Internal and foreign audiences are exposed to influence operations
Narrative Techniques	Conspiracy seeding, denialism, and emotional dissonance	Harmony, nationalism, and anti-West framing	Pro-democracy, transparency, and anti-disinformation messaging
Policy Mechanisms	Operate outside formal regulation (e.g., state-embedded troll farms)	Legal reinforcement via Cybersecurity Law, Internet Sovereignty	DHS, CISA, and AI oversight via civil society and tech regulation
Documented Example	Ukraine's hybrid war; U.S. election interference	COVID-19 propaganda; Xinjiang surveillance	2020 U.S. disinfo task force; collaboration with Meta/Twitter/X

Conceptual Framework: The C3F Model for AI-Enabled Cognitive Warfare

This research paper presents a theoretical framework, known as the C3F Framework, shortened as Cognition, Channels, Convergence, and Feedback, to promote the theoretical aspects of the AI-enabled cognitive warfare and the practical countering measures. The model is multi-layered and says that adversarial actors use a symbiotic loop between technological, narrative, and policy processes to influence thinking at scale by strategically exploiting the capabilities of artificial intelligence (AI) technologies.

In essence, the framework has Cognition, the main object of such operations, which includes human memory, emotions, prejudices, and trust systems. The next layer, above it, is called the Channels layer and contains AI technologies like large language models (LLMs), deep fakes, and bots as vehicles of delivery of persuasive or deceptive messages. These vectors are driven by the Convergence layer, a tactical level where stories are built, joined, and permeated by the synchronized digital involvement of influence activities.

More importantly, the Feedback layer, which includes policy mechanisms, governing structures, and normative frameworks, is crucial in accentuating or breaking the cognitive warfare loops. The meeting of AI tools and narrative operations has the potential to chronically diminish cognitive resilience in democratic societies, without the strong feedback loops through which human beings improve their collective understanding.

This conceptualization excavates lessons of scrupulous case studies (Russia, China, United States) and the latest academic literature in the field of psychological operations, disinformation campaigns, and hybrid warfare. The C3F Framework is therefore an interdisciplinary and scalable framework for comprehending and opposing the emerging danger of cognitive manipulation via AI.

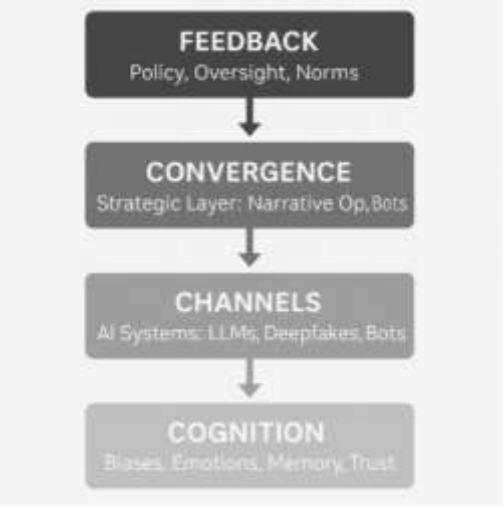

Figure 09: The C3F Framework: Understanding and Countering AI-Enabled Cognitive Warfare.

Findings from Policy Document Analysis
The analysis of policy documents within the scope of this study demonstrates a deepening strategic asymmetry: the means and methods associated with cognitive warfare are increasingly advanced, mainly due to the impact of artificial intelligence (AI) and the exploitation of behavioral data, while international and national policy responses are still disparate, abstract, and poorly enforced. Agencies and organizations, such as NATO, the EU, and extensive U.N. structures, have recognized the theoretical and practical importance of the potential instability created through psychological and informational manipulation, and they attempt to address the issue; however, there are approximately no legal, technical, or doctrinal frameworks capable of solving the problem comprehensively.

NATO: Conceptual Recognition Without Operational Doctrine.
The policy document analysis carried out in this study uncovers a deepening and growing strategic asymmetry: the tactics and technologies underpinning cognitive warfare are becoming infinitely more sophisticated primarily through artificial intelligence (AI) in combination with the exploitation of behavioral data while the international and national policies to respond to cognitive warfare are still fragmented, high-level, and unenforceable. Organizations such as NATO, the European Union, and the United Nations have identified the destabilizing nature of psychological and information manipulation; however, they lack legal, technical, and doctrinal mechanisms to comprehensively deal with this new threat.

United Nations: Diplomatic Recognition, Legal Ambiguity
At the global level, the United Nations has addressed cyber conflict using tools such as the Group of Governmental Experts (GGE) and the Open-ended Working Group (OEWG). These institutions acknowledge the destabilizing impact of information and communication technologies (ICTs) on international peace and security. However, their paradigms are centered on cyber infrastructure rather than meddling with human cognition. The United Nations Institute for Disarmament Research (UNIDIR, 2021) sees cognitive influence as a factor of modern hybrid threats but does not set legal limits to psychological manipulation.
From the point of view of international law, Tallinn Manual 2.0 regards information operations as such only insofar as they disturb key infrastructure or national sovereignty in material ways. The non-kinetic, emotional, and perceptual dimensions of AI-driven influence operations remain outside its scope. As Greene and Kulesza (2022) explain, no treaty or customary principle in international law at present specifies cognitive warfare as a violation of sovereignty, nor does it regulate AI-enabled psychological interference as a violation of international humanitarian law. Such a gap in the law permits state and non-state actors to conduct psychological operations unattributed, legally held to account, or diplomatically sanctioned.

National Responses: Doctrinal Gaps and Strategic Asymmetry
Policy responses remain varied and disjointed across the nation. Foreign disinformation and influence operations feature in strategic documents such as the National Defense Strategy 2022 and various cybersecurity frameworks in the United States, but there is no unified national doctrine for cognitive defense. Existing AI governing policy, that is, Executive Order 13960, comprises responsible and ethical uses of AI for public innovation and service, without consideration of hostile or opposed uses of AI in influence campaigns, perception operations, or manipulation.

By contrast, China has made a real doctrine for information control. The reconciling campaigns tend by "Three War-fares" strategy, psychological warfare, media warfare, and legal warfare into the action of shaping public opinion, discrediting the enemy, and normalizing strategic interests (Kania, 2019). These principles are not theoretical but rather put into active use in influence operations against Taiwan and in the South China Sea, where psychological pressure, narrative framing, and AI-enabled amplification all have central roles to play. Russia continues to be busy in the industry of strategic ambiguity, which is characterized by non-attributable influence operations with proxies and information laundering that are oftentimes amplified by bots and emotionally manipulative content. Unlike the liberal democracies, these authoritarian states have centralized command, legal impunity, and asymmetric tolerance regarding offensive postures.

European Union: Regulatory Action Without Psychological Depth
The most visible policy response to disinformation from the EU is the Action Plan Against Disinformation (European Commission, 2018). The plan lays out a 4-pillar plan: (1) Increase detection of disinformation; (2) Increase responsibility of the platform; (3) Support for fact-checkers and researchers; (4) Build resilience in citizens through media literacy. It resulted in significant progress in urging platforms, such as Facebook, Twitter, and Google, to increase transparency in political ads and eliminate content that is proven false. The EU's strategies nevertheless conceptualize disinformation as a form of content regulation and do not engage with the psychological mechanisms at play that motivate users to share or endorse such content. For example, among other

neglected areas of the Action Plan, engage with how AI systems personalizing experience introduce cognitive biases in content performance or the increasing impact of synthetic media (e.g., deepfakes, text generation via large language models). Although newly enacted laws, such as the Digital Services Act, create a requirement for systemic risk assessments for very large online platforms in their use of algorithms, the primary focus is still algorithmic transparency, and not on protecting against cognitive attacks powered by artificial intelligence, whether that affects democracy, election reliability, or trust in others.

Cross-Cutting Policy Gaps and Strategic Risks

In the comparison of institutional frameworks, three critical, persistent global policy gaps have become evident to see. The first is a disconnect between concept and operation; cognitive warfare is widely perceived as a threat, but actionable strategies for detection, defense, and deterrence are underdeveloped or nonexistent. Second, there is legal ambiguity concerning defining and treating cognitive attacks. Non-kinetic, AI-driven manipulations of cognition find no resonance in the current laws of war, international humanitarian law, and cybersecurity treaties. Thus, psychological aggression lives in a grey zone-not clearly lawful nor explicitly prohibited. Finally, most extant policies deal with platform content regulation and algorithmic accountability or user data protections, keeping the human-level vulnerabilities-cognitive overload, emotional reactivity, motivated reason-spoiled by cognitive warfare unattended.

Indeed, these gaps reflect drainages by which democratic societies become particularly vulnerable to AI-enhanced influence operations. Missing binding legal norms, integrated cognitive defense infrastructures, and cross-border psychological security cooperation leave adversarial actors with the ability to manipulate public perception and trust in institutions while undermining the legitimacy of democracy at relatively low cost and risk.

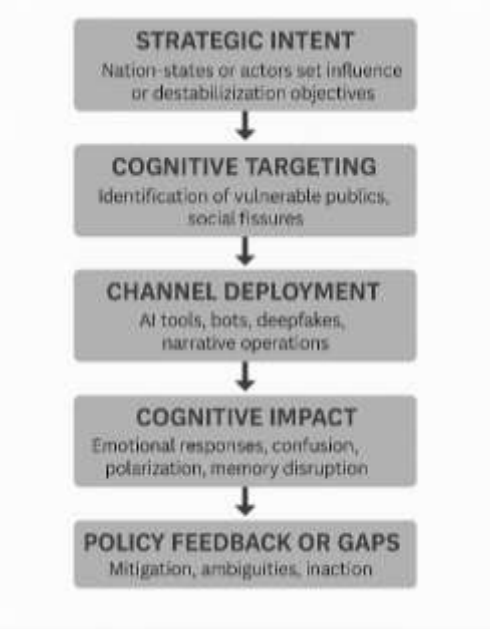

Figure 010: Cognitive Warfare Policy Process Model

This is a diagram of the step-by-step policy-relevant development of cognitive warfare using an AI-enabled cognitive warfare solution from a state-level strategic intent measure to a psychological effect on target populations and then to the closing feedback loop (or gap) of regulatory and policy action. It also at the same time outlines the important parts of targeting, putting into practice technologies, and Hewitt disruption, showing where the failure or uncertainties by the policy chosen to enable these activities.

Conclusions and Recommendations

The study finds that the most contested battlefield of the 21st century has turned out to be the human mind. By synthesizing a broad range of literature, actual political operations, and fragmented policy responses, the article shows how cognitive warfare, supported by artificial intelligence, is transforming warfare. Cognitive warfare is targeted at the architecture of human thought-in contrast to traditional warfare that targets land, infrastructure, or cyberspace. This means affecting how people think-abusing AI powers to deliver messages, to shape perception, to amplify influence, and to diminish the mental sovereignty of persons and societies in a realm of speed, scale, and deniability. These are not dangers on the horizon; they are, rather, present-day operational realities underlying global strategic conduct.

The implications are immediate and deep. All known security mechanisms and doctrines have been structurally designed to be unfit to deal with this issue. They treat information as content to be moderated while ignoring its weaponized potential for psychological disruption. This denies any protection for democratic societies against adversarial structures that do nothing less than affect what people think and how people think. In this new environment, cognitive freedom is no longer guaranteed by law; it must be defended through strategic foresight, interdisciplinary collaborations, and technological protection.

Though this study amalgamates policy papers and secondary literature in order to trace trends, it does not carry direct empirical testing (for example, experiments or interviews). Therefore, some results are interpretative and exploratory in type. Future studies could potentially profit from the integration of the first data to verify or broaden this knowledge.

In this paper, the authors will be suggesting a paradigm shift in defining, defending, and governing cognition in the digital age. It insists that cognitive security is not simply a niche concern; it is the core linchpin for societal resilience, democratic function, and future peace. The defense of cognition should now constitute the fourth pillar of global security strategy, alongside the defense of borders, data, and infrastructure. Failure to acknowledge this fact equals losing not just the information space but the very autonomy of human thought. If the 20th century has taught us that war could be total, the 21st century teaches us that it could be invisible, personal, and psychological; the destruction capability is no less.

References

Bakir, V., & McStay, A. (2018). Fake news and the economy of emotions. *Digital Journalism, 6*(2), 154–175. https://doi.org/10.1080/21670811.2017.1345645

Baumeister, R. F., & Leary, M. R. (1997). Writing narrative literature reviews. *Review of General Psychology, 1*(3), 311–320. https://doi.org/10.1037/1089-2680.1.3.311

Benkler, Y., Faris, R., & Roberts, H. (2018). *Network propaganda: Manipulation, disinformation, and radicalization in American politics*. Oxford University Press.

Bozalka, D. (2023). *Information warfare in the age of artificial intelligence* (Strategic Brief No. 62). Institut de Recherche Stratégique de l'École Militaire (IRSEM). https://www.irsem.fr/strategic-brief-no-62-2023-information-warfare-in-the-age-of-artificial-intelligence.html

Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., ... & Amodei, D. (2018). The malicious use of artificial intelligence: Forecasting, prevention, and mitigation. *arXiv preprint*. https://arxiv.org/abs/1802.07228

Buchanan, B. (2020). *The hacker and the state: Cyber attacks and the new normal of geopolitics*. Harvard University Press.

Bush, R. C. (2021). *How China is remapping Taiwan's information environment*. Brookings Institution. https://www.brookings.edu/articles/how-china-is-remapping-taiwans-information-environment/

Clarke, V., & Braun, V. (2013). Teaching thematic analysis: Overcoming challenges and developing strategies for effective learning. *The Psychologist*, *26*(2), 120–123.

Chesney, R., & Citron, D. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *California Law Review, 107*(6), 1753–1819. https://doi.org/10.2139/ssrn.3213954

Cinelli, M., Morales, G. D. F., Galeazzi, A., Quattrociocchi, W., & Starnini, M. (2021). The echo chamber effect on social media. *Proceedings of the National Academy of Sciences, 118*(9), e2023301118. https://doi.org/10.1073/pnas.2023301118

European Commission. (2018). *Tackling online disinformation: A European approach*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52018DC0236

Fazio, L. K., Brashier, N. M., Payne, B. K., & Marsh, E. J. (2015). Knowledge does not protect against illusory truth. *Journal of Experimental Psychology: General, 144*(5), 993–1002. https://doi.org/10.1037/xge0000098

Ferrara, E. (2020). What types of COVID-19 conspiracies are populated by Twitter bots? *First Monday, 25*(6). https://doi.org/10.5210/fm.v25i6.10633

Greene, D., & Kulesza, J. (2022). Legal frameworks for AI-enabled cognitive warfare: The limits of international law. *Journal of National Security Law & Policy, 13*(1), 45–74.

Hancock, J. T., & Bailenson, J. N. (2021). The social impact of deepfakes. *Cyberpsychology, Behavior, and Social Networking, 24*(3), 149–152. https://doi.org/10.1089/cyber.2021.29208.jth

Jamieson, K. H. (2018). *Cyberwar: How Russian hackers and trolls helped elect a president*.

Oxford University Press.
Kania, E. B. (2019). *The weaponization of AI and the future of conflict*. Center for a New American Security. https://www.cnas.org/publications/reports/the-weaponization-of-ai-and-the-future-of-conflict
Lee, C. (2021). Disinformation and psychological warfare in the Taiwan Strait. *Asian Security, 17*(1), 32–52. https://doi.org/10.1080/14799855.2020.1762855
Lewandowsky, S., Ecker, U. K. H., & Cook, J. (2017). Beyond misinformation: Understanding and coping with the "post-truth" era. *Journal of Applied Research in Memory and Cognition, 6*(4), 353–369. https://doi.org/10.1016/j.jarmac.2017.07.008
Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest, 13*(3), 106–131. https://doi.org/10.1177/1529100612451018
Mandic, J., & Klaric, D. (2023). Case study of the Russian disinformation campaign during the war in Ukraine – Propaganda narratives, goals and impacts. *National Security and the Future, 24*(2), Article 5. https://doi.org/10.37458/nstf.24.2.5
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*, 170. https://doi.org/10.1007/s13278-024-01322-w
Mueller, R. S. (2019). *Report on the investigation into Russian interference in the 2016 presidential election*. U.S. Department of Justice. https://www.justice.gov/storage/report.pdf
NATO Innovation Hub. (2021). *Cognitive warfare*. NATO ACT. https://www.innovationhub-act.org/sites/default/files/2021-04/CognitiveWarfare.pdf
Pennycook, G., & Rand, D. G. (2019). The implied truth effect: Attaching warnings to a subset of fake news headlines increases perceived accuracy of headlines without warnings. *Management Science, 66*(11), 4944–4957. https://doi.org/10.1287/mnsc.2019.3478
Pennycook, G., & Rand, D. G. (2020). Who falls for fake news? The roles of analytic thinking, motivated reasoning, political ideology, and bullshit receptivity. *Journal of Personality, 88*(2), 185–200. https://doi.org/10.1111/jopy.12476
Polyakova, A., & Boyer, D. (2018). *The future of political warfare: Russia, the West, and the coming age of global digital competition*. Brookings Institution. https://www.brookings.edu/articles/the-future-of-political-warfare
Roozenbeek, J., & van der Linden, S. (2019). The fake news game: Actively inoculating against the risk of misinformation. *Journal of Risk Research, 22*(5), 570–580. https://doi.org/10.1080/13669877.2018.1443491
Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research, 104*, 333–339. https://doi.org/10.1016/j.jbusres.2019.07.039
Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences, 23*(5), 645–665. https://doi.org/10.1017/S0140525X00003435
Swire-Thompson, B., & Lazer, D. (2020). Public health and online misinformation: Challenges and recommendations. *Annual Review of Public Health, 41*, 433–451. https://doi.org/10.1146/annurev-publhealth-040119-094127
Tucker, J. A., Guess, A., Barberá, P., Vaccari, C., Siegel, A., Sanovich, S., Stukal, D., & Nyhan, B. (2018). *Social media, political polarization, and political disinformation: A review of the scientific literature*. Hewlett Foundation. https://hewlett.org/wp-content/uploads/2018/03/Social-Media-Political-Polarization-and-Political-Disinformation-Literature-Review.pdf
Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124
Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science, 359*(6380), 1146–1151. https://doi.org/10.1126/science.aap9559
Weidinger, L., Mellor, J., Rauh, M., Griffin, C., & Uesato, J. (2021). Ethical and social risks of harm from language models. *arXiv preprint*. https://arxiv.org/abs/2112.04359
Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. PublicAffairs.