EWACS as a Backbone for Wargaming and Decision Support in the Information Environment

Alberto F. Olivieri[1,2], Rosanna E. Guadagno[2], Stefano Solari[1] and Enrico Russo[3]

[1]Leonardo, Fast Prototyping Lab of Leonardo Cyber & Security Division, Via Raffaele Pieragostini, 80, Genova, Liguria, Italy
[2]University of Oulu, Pentti Kaiteran katu 1, 90570 Oulu, Finland
[3]Università di Genova, Via Balbi, 5, 16126 Genova, Italy

Abstract

Western democracies are facing an increasing number of adversarial information operations that exploit well-known psychological mechanisms and the capabilities of digital social platforms to distort public debate and erode trust. This paper proposes a conceptual architecture and methodological blueprint integrating an early warning system with a psychologically informed, agent-based simulation environment and a social media "cyber range". The Early Warning and Control System (EWACS) for the information space would continuously update the narrative and network picture of the information environment and would supply the necessary data to parametrize a red–blue agent-based simulator. Defender Courses of Action (CoAs) are intended to be modeled as parameterized intervention packages and would be compared via repeated runs using diffusion and belief-shift metrics. Cyber Threat Intelligence (CTI) inspired ontologies (DISARM) and a reimagined Pyramid of Pain structure the intervention space, estimating where actions impose cost on narratives, network structures, Tactics, Techniques, and Procedures (TTPs), and strategic objectives. Together, these components would form a decision support and wargaming environment that could allow planners to compare simulated outcomes across alternative CoAs, stress-test escalation pathways, and coordinate disparate tactical measures into a more coherent multi-channel strategy for deterring, disrupting, and mitigating adversarial information operations. This paper does not demonstrate empirical effectiveness; its contribution is to illustrate a conceptual specification intended as a stepping stone for future work on implementation and validation.

Keywords

Early Warning Systems, Information Operations, Persuasion Theory, Decision Support Systems, Wargaming

1. Introduction

The global spread of false information that significantly influences individuals' worldviews and behaviors has become one of the preeminent social and security issues. The Western information environment is increasingly contested, with adversarial powers using competing narratives as tools of influence. This dynamic enables distrust and uncertainty among both the general public and decision-makers. The contemporary geopolitical landscape is marked by a state of confrontation between Western countries, mainly the United States, and major powers such as China, Iran, and Russia [1]. Available public reports suggest an asymmetry between the attackers and the defenders in the Information Environment, driven by uneven national capacities, attribution and coordination frictions, and platform affordances that favor rapid adversarial adaptation [2], [3], [4]. This asymmetry is exacerbated by the reluctance of numerous Western governments to allocate resources to enhance informational resilience and to develop offensive and defensive capabilities and methodologies to counter adversarial Information Operations [5].

The objective of this study is to examine the potential for integrating Early Warning and Control System (EWACS)-type monitoring [6], [7] with a proposed AI-driven simulation environment operating in concert with a *social media cyber range*. This framework would allow decision-makers to simulate various scenarios, estimate plausible trajectories and outcomes across alternative CoAs, and practice responses before taking action; consequently, it could facilitate the design and evaluation of actionable intervention strategies intended to enhance the resilience of selected target populations. In the current

Joint National Conference on Cybersecurity (ITASEC & SERICS 2026), February 09-13, 2026, Cagliari, IT
albertofedericoolivieri@gmail.com (A. F. Olivieri); rosanna.guadagno@oulu.fi (R. E. Guadagno); stefano.solari@leonardo.com (S. Solari); enrico.russo@unige.it (E. Russo)
0000-0001-7223-3522 (A. F. Olivieri); 0000-0001-8247-5154 (R. E. Guadagno); 0000-0002-1077-2771 (E. Russo)

information environment, EWACS-powered tools could help enable defenders to proactively counter hostile narratives.

Briefly, this paper makes the following contributions.

- It links an EWACS-type monitoring layer to a simulation and wargaming environment for decision support.
- It integrates CTI-inspired taxonomies with a reimagined Pyramid of Pain to structure and compare defender CoAs.
- It illustrates the resulting framework through a real information-operations incident, highlighting the operational issues it is designed to address.

Structure of the paper. The rest of the paper is organized as follows. Section 2 reviews early-warning systems and wargaming research. Section 3 summarizes the psychological and conceptual background. Section 4 presents the Bronze Soldier use case. Section 5 details the proposed architecture and workflow. Section 6 discusses how the methodology could be applied to the case and highlights operational issues, limitations, and risks. Section 7 concludes.

2. Related Work

Research on early warning and monitoring in the information environment has mostly evolved in two different design focuses. The first is mainly concerned with tracking the spread of misinformation on social media and its eventual debunking. Hoaxy is a good example of this, it automatically collects, organizes, and displays web articles on specific disinformation narratives and their fact-checking. Hoaxy shows that it is possible to monitor information flows at a large scale and almost in real time [8] (now discontinued, but useful as a reference design). The second design focus is conceptually related to the EWACS (Early Warning and Control System), as it is a tool with early warning capabilities. Yankoski et al. suggested a system that uses AI to identify harmful online disinformation (e.g., memes that dehumanize others, calls to violence) and generate alerts. The system is described as "useful to journalists, peacekeepers, election monitors, and others who need to understand how manipulated content is spreading online during elections and in other contexts" [9]. These examples show how automated systems could be used to monitor, warn, and try to stop new information operations and harmful narratives before they spread. Unfortunately, these systems often focus only on early detection and warning, and they aren't integrated into wargaming or decision-support processes.

At the same time, there is a long history of using wargames to support analysis, planning, and decision-making in complex or adversarial environments. Perla's book "The Art of Wargaming" provides the classic conceptual foundation, arguing that well-designed wargames can help decision-makers explore the consequences of plans, understand opponent behavior, and identify vulnerabilities. Perla also emphasizes the risks of misusing games as prediction engines or allowing poorly specified objectives to bias outcomes [10]. Longley-Brown's book "Successful Professional Wargames" builds on this work by looking at it from a practitioner's perspective. He explains the best ways to play professional and defense wargames and documents common pitfalls that can make these games fail. The most common are confirmation bias, a lack of rigorous testing, and inadequate integration with real-world data and decision cycles [11]. Both of these texts show how wargaming is a complex and valuable way to support professional decision-making.

Combining these two ideas into one, would create an opportunity for a tool that integrates real-time monitoring with strategic wargaming. Current public early warning systems only detect and alert, while wargaming is usually built by creating either static or ad-hoc scenarios. EWACS could function as the backbone of a system of systems that uses real-time data from the information environment, and turns it into forecasts, wargaming inputs, and decision support. This has the potential of transforming strategic decision-making processes for the information environment by providing dynamic, constantly updated, data-driven insights.

Going beyond academic prototypes, commercial platforms (e.g., Gerulata Sentinel [12]) already provide operational early-warning systems for hostile information activities. However, they don't provide much public information about their architecture or how they work with wargaming and decision support workflows, and this leaves us with a limited understanding of their full capabilities.

3. Background

This section introduces the key concepts needed to understand how modern information operations work and how they influence audiences.

3.1. Propaganda, Disinformation, and the Information Environment

The definitions and understandings of terms like propaganda and disinformation vary significantly across different fields. In this work, the term "propaganda", alternatively referred to as "information operations" (IO), does not inherently imply malevolent intent. The U.S. Department of Defense defines IO broadly, describing capabilities that range from cyber to electronic warfare, which expands the traditional understanding of propaganda to include modern technological dimensions [13]. "Propaganda" is a neutral term [14], and it is defined as "The deliberate, systematic attempt to shape perceptions, manipulate cognitions, and direct behaviour to achieve a response that furthers the desired intent of the propagandist". This definition fits various endeavors, from social awareness campaigns designed to benefit the public, such as reducing alcohol abuse, to adversarial action focused on influencing election results. Propaganda can be evaluated through its transparency of origin and trustworthiness of information. A transparent and trustworthy propaganda is known as White Propaganda, on the other side we have Black Propaganda, and then shades of Gray Propaganda between those two extremes [14], [15]. An example of white propaganda is a government-led campaign promoting public health by reducing alcohol abuse, clearly identifying its source and intent. A Black Propaganda example is the disinformation campaign engineered by the USSR, pinning the origin of the AIDS virus to the US [16], [17], [18]. Lastly, examples of Grey Propaganda could be seen in some of the frontline updates shared by Russian and Ukrainian sources.

One of the main components of Black Propaganda is disinformation, a type of information that was "purposefully created" to be misleading. It is important to note that disinformation is not necessarily false, as it does not need to be, as the main intent is to disrupt the information space. IOs are often based on a central "rational core", which is then misrepresented [19], to push the audience to reach the conclusions of the propagandist or, at least, to make the information space murky, leading to inaction. Disinformation being intentional is what sets it apart from misinformation, as this information type is misleading, but not intentional. Common disinformation examples are deceptive advertising, government propaganda, doctored photographs, forged documents, fake online engagement, and more [20].

While the contemporary information landscape evolves rapidly, the underlying psychological principles, such as cognitive biases and persuasion techniques, remain consistent. That's why principles and methodologies associated with Cold War-era propaganda are still used for modern adversarial IOs [21]. The fundamental logic of influence remains unchanged, however, the infrastructure through which it is disseminated has undergone significant evolution. The information space is full of individuals in positions of media influence who possess the capacity to manipulate public discourse and perceptions, often employing a selective approach to present biased viewpoints that are congruent with their personal or professional agendas. These actors have become increasingly reliant on highly connected digital platforms for their operations. The rise of user-generated content democratizes information dissemination, allowing both accurate information and disinformation to spread more widely and rapidly, impacting public perception [22].

3.2. Psychological mechanisms

Contemporary adversarial IOs can be interpreted through different psychological lenses to understand their effects on attitudes and behavior. Two primary models would be taken into consideration to understand those effects: the dual process model of persuasion, such as the Elaboration Likelihood Model (ELM) [23], and Cialdini's seven principles of influence, focusing on commitment and consistency [24], [25], [26], [27]. According to ELM, an attitude change can occur via two main routes: (i) a more effortful central route when motivation and the ability to elaborate are high, or (ii) it can occur via a more cue-driven peripheral route when these factors are low (e.g., under time pressure, cognitive load, or information overload) [23]. Alongside this, Cialdini's influence principles are defined as: (i) reciprocation (returning favors), (ii) commitment and consistency (aligning with prior acts), (iii) social proof (following others), (iv) liking (responding to affiliation), (v) authority (deferring to expertise), (vi) scarcity (valuing rarity), and (vii) unity (acting as "we") [25], [26].

In the context of information operations, threat actors deliberately design content to reduce deliberation and increase reliance on peripheral cues, like evoking emotionally vivid examples and extreme cases, using highly stereotyped narratives that match pre-existing themes. Russian information warfare practices often rely on a recognizable "kernel of truth" that is widely accepted at an intuitive level, which is then selectively framed and distorted to steer audiences toward the desired conclusion or, at minimum, to erode trust in alternative sources [19], [28]. This combination of a rational core and biased framing creates an environment in which peripheral cues and heuristics dominate because many people lack the time, resources, or incentive to scrutinize arguments.

Platform design significantly influences communication by shaping how information is presented and consumed, a vulnerability that can be exploited by actors adept at manipulating these features. Short-form platforms such as X (formerly Twitter) limit message size and fragment discussions into branching threads, which makes sustained, complex argumentation difficult to follow. This could push a larger share of users toward peripheral processing, where they rely more heavily on simple cues (likes, retweets, source heuristics, emotional tone) rather than on systematic evaluation of arguments [29], [30]. This creates an opening for adversaries willing to exploit such signals to manipulate and influence the audience.

According to Cialdini the commitment and cognitive consistency principle is described as people striving to maintain coherence between their past actions, and current attitudes [25], [26]. This tendency can be systematically exploited, as discussed by Cialdini in the case of U.S. prisoners of war in the Korean War. Chinese guards obtained repeated cooperation from many U.S. POWs by first requesting small, trivial statements mildly critical of the United States or supportive of communist positions, often in exchange for improved living conditions. Over time, the demands escalated to more public acts of collaboration. The more the POWs complied, the harder it became to disown their previous behaviour. Persuasion is gradual rather than instantaneous, small and sustained commitments, can accumulate, producing in the long term large behavioural shifts. In general, repeated compliance and frequent exposures can lead to the formation of stable attitudes without the need for extensive conscious deliberation. The "becoming a collaborator" trajectory described in the Korean War case is an emblematic example of incremental commitment over time [25], [26]. Online, analogous sequences can be observed when users are first asked to endorse innocuous statements, then to share slightly more radical content, and finally to participate in overtly conspiratorial or extremist narratives. Each subsequent step is both cementing and pushing forward the change in attitude and behavior, as it is necessary for maintaining internal consistency with own prior statements.

Taken together, the mechanisms of influence, such as commitment, consistency, and reliance on heuristics, indicate that threat actors operate on a well-studied psychological terrain. Rather than inventing new forms of manipulation, they systematically combine well-established influence principles and take advantage of predictable features of human cognition. The combination of complex platform dynamics, adversarial adaptation, and cognitively constrained audiences suggests that static best-practice guidelines, or passive debunking approaches alone are insufficient. Instead, defenders require environments in which adversarial and defensive strategies can be tested iteratively against realistic

models of human decision-making. As a gamified demonstration, Roozenbeek and van der Linden [31] show that a purpose-built browser game can inoculate players against common misinformation tactics, improving their ability to recognise and resist them. If even simple games can enhance resistance at the individual level, then more sophisticated simulation environments should be explored, in order to integrate psychological mechanisms as tools for decision support, strategic communication planning, and the wargaming of information-operations scenarios.

3.3. Early Warning and Control System

An EWACS is a proactive tool designed to enhance monitoring in the information environment. This allows for early detection of viral trends and emergent narratives, creating a time buffer for strategic communicators to plan and deploy calibrated responses, before hostile campaigns gain major traction. EWACS is described as a system that can support the flagging of individual disinformation artifacts and suspicious accounts, enable AI-assisted narrative discovery and pattern identification, contribute to tracking coordinated networks, and help monitor shifts in adversary Tactics, Techniques, and Procedures (TTPs). This keeps operators and policymakers informed about evolving capabilities and longer-term objectives of threat actors. As an intelligence-delivery tool, EWACS leverages the information environment's capacity for retention to build an expanding knowledge base of campaigns, actors, and affiliations. The goal is to accelerate attribution and improve the timeliness and coordination of defensive actions [6], [7].

4. Case Study

This section outlines the Bronze Soldier case and extracts the key operational issues it reveals, providing the basis for the later methodological framework.

4.1. The Bronze Soldier Case

A widely cited case of an information operation is undoubtedly the 2007 incident involving the Bronze Soldier statue in Tallinn. In 1947, the Soviet Union placed a bronze statue of a Soviet soldier in the center of the Estonian capital to celebrate the victory over Nazi Germany. Present day protesters used the statue to celebrate the Soviet regime in Estonia, prompting the government to relocate it outside the city center. This led to massive riots by pro-Russian activists that lasted three days, resulting in one death, 153 injuries, and the detention of more than 1,000 people. On the cyber side, Estonia's cyber infrastructure was under attack for twenty-two days. Estonian websites belonging to Estonia's president, prime minister, parliament, government agencies, banks, and news agencies were subjects of repeated DDoS attacks and shut down. Some non-critical websites were hacked and defaced. The following fallout was even worse, with Russian diplomats spreading false information about the Estonian government. Ultimately, the Russian Duma released a statement calling the relocation a glorification of Nazism, painting the whole Estonian government as pro-Nazi [32], [33].

During the tense period from the announcement of the intent to relocate until after the relocation, the Russian government, through the youth organization Nashi, created and incited protests and riots in Tallinn. This organization and others were also responsible for violent protests at the Estonian embassy in Moscow. While the Moscow Police did not intervene to stop the protesters from disrupting embassy activities, the Kremlin delivered water, tents, and food to them [33].

The official anti-Estonian narrative was pushed by various state-sponsored media channels in both Russia and Estonia, using the same talking points that were used by the protesters and the government. The entire apparatus was able to work in synchrony. The media added fake elements to the story of the relocation, such as the idea that the statue was cut into pieces and defaced prior to the relocation. They also exaggerated the police response, painting it as extremely brutal and repressive [33].

The narratives Russia pushed forward in this information operation were varied and usually contained at least a kernel of truth. However, the facts were spun to push the Estonian public in the direction

of Russia's will: keeping a memento of Soviet overlordship in the center of Tallinn. This operation aligns with Russia's broader strategy of reestablishing itself as a regional hegemon that can influence smaller nations near its borders. Ultimately, the statue was relocated, however, the three days of riots, one fatality, and subsequent diplomatic crisis underscore how a seemingly local memory dispute can be weaponized into a full-spectrum information operation. Moscow's response included external messaging that framed Estonia's decision as "Nazi glorification" and internal mobilization of Russian and Russian-speaking audiences. The same audiences were prompted on the Russian internet to join the DDoS attacks, carefully synchronized with physical attacks, and many answered the call [32]. This mobilization tapped into the longstanding narrative that any opponent of Russian policy is a Fascist or a Nazi.

4.2. Operational Issues highlighted by the Bronze Soldier Case

The Bronze Soldier episode illustrates an instance of a Russian INFO OPS fusing symbolic politics, street mobilization, and dual-audience propaganda, easily spilling over from the information sphere into physical violence as part of a long-term strategic pressure campaign. This episode exposes four operational issues that are recurrent in contemporary information operations:

- **Late detection**. The Bronze Soldier statue incident occurred in a rapidly changing information environment, with only a narrow warning window to identify the threat, and culminated in three days of riots and an extended cyber campaign. By the time the operation was discovered, the campaign was nearly at its peak, with mobilization, messaging, and cyberattacks already underway.
- **Cross-domain synchronization (online/offline/cyber)**. The incident developed across various domains, fusing street mobilization, cyberattacks, and coordinated messaging into a single pressure campaign. Physical riots occurred alongside sustained DDoS attacks targeting core government, banking, and media sites. Participation was encouraged online, synchronized with offline actions.
- **State coordination**. State-sponsored media pushed an anti-Estonian narrative using consistent talking points aligned with protest messaging, such as claims of desecration and police brutality, which were often fabricated or exaggerated. The Russian Duma's framing of the relocation of Soviet-era monuments as "Nazi glorification" was similar, further reinforcing the claim of a coordinated, top-down operation.
- **Deniability**. The campaign relied on intermediaries such as the youth organization Nashi and associated groups to organize and incite protests. Meanwhile, the state provided logistical support to the protesters, indicating both its alignment with their cause and maintaining plausible deniability for its involvement. Similarly, the online community responsible for orchestrating the cyberattacks seemingly operated without interference or consequences.

5. Methodology

In this section we describe how an EWACS-enabled early-warning capability is combined with CTI-inspired taxonomies and an agent-based simulation of the information environment to support wargaming and CoA selection. We first motivate the need to move from reactive debunking to proactive, EWACS-enabled approaches (Section 5.1). We then introduce the CTI taxonomies used to structure the space of adversarial and defensive behaviours (Section 5.2). Finally, we describe the agent-based model and simulation loop that turn EWACS data into decision-support outputs (Section 5.3 and Section 5.5).

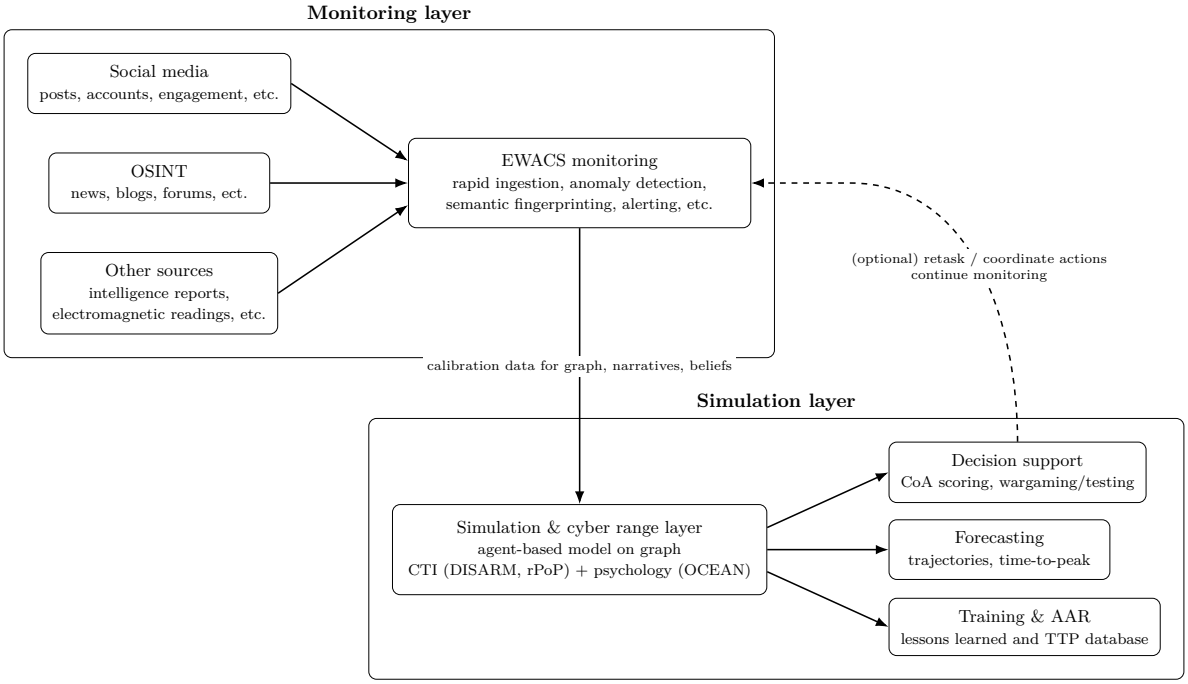

Figure 1: High-level architecture of our methodology.

5.1. From reactive debunking to proactive EWACS-enabled defense

Debunking, a prevalent method in counter-disinformation campaigns, often involves fact-checking and dissemination of accurate information, but its effectiveness is limited when deployed reactively after a disinformation campaign has gone viral. Moreover, even after its deployment, not everyone would be convinced, or even reached, by the counter-narrative campaign [34]. It should also be considered how the modern information environment is structured to favor quick information delivery, and its correctness is not considered as important. Knowing this, attackers use the sheer content volume to overwhelm defenses based on debunking, that are unable to keep pace. This means that speed is essential in any proposed strategy to counter adversarial IOs, and the proposed countermeasures are more and more focused on proactive, preemptive, and deterring approaches [35], [36].

Rapid data ingestion and analysis are essential for tools monitoring the information environment, as exemplified by a proposed EWACS (Early Warning and Control System) for the information space [6]. A tool with EWACS-like capabilities should be integrated as the backbone of a system that aims at improving decision making through simulating various scenarios, estimating plausible (in-simulation) outcomes, and practicing responses before taking action.

The simulation system should model complex interactions through an agentic AI-driven simulation environment, operating in concert with a social media cyber range able to inject narratives and simulate their effectiveness and spread over time. These two systems are going to positively interact with each other: the data is collected from various social networks, then it is analyzed and shared to the simulator to calibrate agent behavior, narrative trajectories, and network structure. Here the simulation can be used to forecast possible outcomes with, or without interventions, and it should help decision makers wargame scenarios from which they could extract the best Course of Action (CoA). At this stage, with the knowledge generated by the simulation, the decision maker would be able to go back to the EWACS system and apply countermeasures live, or just continue monitoring the developing information space situation.

Figure 1 summarizes the EWACS-enabled pipeline, from continuous monitoring of the information environment to the simulation-based evaluation of defender CoA.

5.2. CTI-informed taxonomies and CoA scoring (rPoP and DISARM)

To provide a structured way of describing adversarial behaviours and defender responses, the simulator adopts CTI taxonomies adapted to the cognitive domain.

Drawing inspiration from established concepts and frameworks from the field of cybersecurity and threat intelligence, and adapting them to meet the specific requirements of the cognitive domain, is common practice. This simulator should use revised Cyber Threat Intelligence (CTI) frameworks, like the Reimagined Pyramid of Pain [7] (rPoP) and the Disinformation and Risk Management (DISARM) framework [37], to interpret the efficacy of narratives and counter narratives, and to help classify various defender CoAs and their impact on the adversary.

Generally, taxonomies have been demonstrated to be beneficial tools for the identification and allocation of resources. Behavioral taxonomies, in particular, can enhance understanding by categorizing actions and responses, thereby improving strategic decision-making. The MITRE ATT&CK framework [38] is widely regarded as one of the best taxonomies in CTI, as it provides a structured workflow for integrating disparate pieces of information and artifacts into standardized TTPs. The Adversarial Misinformation and Influence Tactics and Techniques (AMITT) project, renamed DISARM in 2021, employs a similar approach for disinformation by systematically cataloging attackers and responder behaviors. The objective of these frameworks is to explore the underlying strategies and structures present in adversarial campaigns, which can reveal patterns and inform more effective countermeasures [37]. By highlighting emerging and recurrent procedures (e.g., seeding, laundering, amplification), these frameworks enable defenders to strategically allocate efforts, like deploying tailored responses to specific attack patterns. This should be an integral part of the wargaming and decision support side of the EWACS-powered tool.

Bianco's Pyramid of Pain [39] was adapted to shift from cyber-technical indicators to information-operations artifacts, enabling comparative analysis of alternative CoAs. The rPoP retains its six-level structure, ranging from individual messages to strategic objectives. The levels are delineated as follows: Level 1 (individual messages), Level 2 (accounts), Level 3 (content themes and narratives), Level 4 (network structures), Level 5 (tactics, techniques, and procedures - TTP), and Level 6 (strategic objectives). In the simulation, Levels 1–2 are shown as aggregate flows, such as the volume of posts and active accounts. Defender actions focus on Levels 3–5, targeting narratives, network structures, and TTPs.The rPoP serves as a cost-of-operation lens, suggesting that interventions at higher levels are more disruptive to adversaries but also pose greater demands and risks to defenders. This classification should help rank and prioritize the proposed CoAs, where the higher their impact is on the rPoP, the more effective and painful they become for the attacker, while also evaluating the defender price, for a more in-depth cost-benefit analysis of the actions.

5.3. Agent-based representation of narratives and communities

We then instantiate an agent-based model that captures how narratives propagate across communities and how individual differences modulate susceptibility to influence.

As we previously discussed, modern information operations utilize well-known influence mechanisms, such as Cialdini's principles of commitment, social proof, authority, and unity. They also consider the difference between effortful, argument-based processing and fast, cue-driven processing as described by Dual Process theories. In the simulation model, these mechanisms are simplified by focusing on key elements rather than simulating full cognitive processes. Instead of simulating full cognition, agents use personality effects, specifically the Big 5 personality factors: Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism or OCEAN [40], values should be based on empirical findings. These effects are combined with value alignments and the structural properties of narratives and communities. This approach approximates the likelihood of agents accepting, sharing, or resisting a given narrative, based on their personality and community influences.

5.4. Information Space Representation and Data Flow

The information environment (social media platforms) is represented as a directed graph, where nodes symbolize individual accounts or groups, and edges illustrate the connections such as follows, friendships, or subscriptions and define how narratives can spread. Each node is assigned a side (blue, red, green, orange, or gray) and may be flagged as a persistent spreader, such as bots, paid influencers, or institutional channels. Nodes are associated with attributes like social media activity, OCEAN personality traits, Political/Economic/Social (PES) values, memberships in socio-political communities, a "stubbornness" parameter indicating resistance to change, and narrative-specific states. These states indicate whether the node is ignorant, exposed, a believer, an active diffuser, or no longer interested in each narrative.

Narratives are modeled without regard to their specific content, focusing instead on their structural and relational attributes. Each narrative is assigned a typology, such as "disinformation" or "debunking." It is also given a side, topical tags, a novelty score, PES alignment values, and a position in a correlation matrix. This matrix encodes the mutual exclusivity, conflict, irrelevance, or mutual reinforcement of pairs of narratives. For each individual agent or group of agents, and for each narrative, a Belief Strength value in the range [1, 1] is provided. This value serves to summarize how strongly the narrative is accepted or rejected. The value is determined by the following factors: (i) correlations with narratives the node already believes or disbelieves, (ii) the fit between the narrative's PES profile and the node's own values, and (iii) the fit with the values of the communities the node belongs to. When a node encounters a new narrative, its Belief Strength for that narrative is updated, and the other entries are adjusted once to reflect how this new "frame" shifts the relative salience of existing beliefs. As time progresses, state-transition rules cause nodes to shift between the following states: ignorant, exposed, believer, diffuser, and disengaged. These transitions are influenced by exposure to spreaders, activity levels, narrative novelty, personality weights, and stubbornness. Collectively, these parameters approximate susceptibility and persistence. They modulate how agents respond to salient cues (e.g., social proof, in-group unity, and emotional tone) as well as the duration that narratives remain active in various sections of the graph.

5.5. Simulation loop and decision-support outputs

Finally, we describe how EWACS data are used to initialize the model and how simulation outputs are turned into decision-support indicators for planners and wargamers.

EWACS provides the live and historical data necessary to initialize and parameterize this model. This data includes several components: the graph structure (agents, groupings, communities, edges), distributions of activity and inferred traits, observed narratives and correlations, and approximate Belief Strength profiles for key segments. Following the data initialization, the model simulates campaigns where agents of the Red and Blue teams are pitted against one another on this graph, with the defender's CoAs varying. These CoAs may include alternative timings and targets for counter-narratives, varying levels of amplification or throttling, or more aggressive account or network interventions. For each CoA, the simulator should provide trajectories. These are derived from standard diffusion metrics (reach, time-to-peak, persistence) and higher-level indicators (such as changes in beliefs in protected or contested segments, rPoP levels primarily affected, TTPs involved, and resilience of adversarial structures) are computed across multiple runs.

In practice, this simulation layer functions as the core for a decision-support tool, a wargaming environment, and a CoA exploration engine within strategic planning. As a decision-support tool, it allows planners to compare CoAs using a parameterized model, providing a more structured basis than intuition or static checklists. As a wargaming environment, it enables red–blue exercises for strategic communication (stratcom), information operations (InfoOps), psychological operations (PsyOps), or military deception (MILDEC), stress-testing coordination and revealing weaknesses in the decision chain. As a CoA exploration engine, it allows stakeholders at tactical, operational, and strategic levels to explore "what-if" scenarios. For example, they can consider earlier pre-bunking, different protection

strategies for green audiences, or higher-level interventions, and observe how these choices affect the distribution of "pain" across rPoP levels for the adversary. By integrating EWACS data with a psychologically informed agent model, the tool is designed to enable faster, better-informed, and more resilient decision-making within the information environment, helping achieve Cognitive Superiority [41].

6. Discussion

This section examines how the proposed architecture would operate in a real scenario, using the Bronze Soldier incident (see Section 4) as a guiding example. We outline how early detection, simulation, and CoA exploration would unfold in such a case, highlight the operational issues it exposes, and discuss the broader limitations and risks of employing EWACS-enabled decision support in the information environment.

6.1. Bronze Soldier statue Incident

We use the 2007 incident involving the Bronze Soldier statue in Tallinn as an illustrative exercise to understand how an EWACS-like system, linked to a decision support system, could materially improve response timing and decision quality in such a crisis.

EWACS detection. In the first period a thorough surveillance of internal and adversarial information space could plausibly have triggered alarms, as the traffic and the narratives were starting to take hold and spread. This could have been achieved by various means, from monitoring of well known threat actor spaces, to noise filters that could catch unusual activity spikes, or from simple keyword match. Once such first alerts were triggered, the second phase could have focused on monitoring trends, and widening the monitoring through semantic fingerprinting of the narratives, thus discovering the real extent of the operation. Sources monitored that are known to be linked to the Kremlin should also function as canary sources that trigger additional alerts of a concerted operation starting to take place. In a mature deployment, this could occur within days, and at that point in time, with enough historical data regarding the internal and adversarial information space collected, and the new narrative detected early through the means previously (non exhaustively) listed, the information could be handed to the simulation engine.

Simulation and CoA. The simulation of the spread of the narrative could generate additional alerts/flags. Decision-makers could at this point be involved by the analysts, and the simulator engine could be dedicated to explore various scenarios and possible responses. This process aims to maximize the pain for the attacker, minimize the damage to the defender, and prevent the expenditure of too much political capital and other assets on the defender side, while, at the same time, increasing the adversarial's expenditures. After choosing and deploying the most effective CoA early on in the campaign, the adversarial attempt could plausibly have been attenuated enough to reduce momentum and coordination. At this point, the Estonian government could have proceeded with their plan with less disruption, and with ample margin of maneuver to manage the protesters. This campaign and the tactics deployed by the Russian side could also be added to a database and, regardless of the final outcome, both analysts and decision-makers could use the campaign simulation to wargame: testing to find new avenues, unpacking eventual errors or finding crucial decision points, as to understand better the adversarial attack, and create lesson learned for tackling similar situation in the future.

6.2. Addressing the Operational Issues

Building on the four operational issues introduced in Section 4, we now outline how an EWACS-enabled wargaming and decision-support architecture could mitigate each of them.

Late detection. In our architecture, late detection should be addressed by the EWACS monitoring layer (anomaly detection + narrative tracking) feeding indicators into the simulation layer to forecast possible critical situations, in order to create alerts before the campaign reaches critical mass.
Cross-domain synchronization (online/offline/cyber). From a methodological perspective, this approach emphasizes the need for multi-domain capabilities: EWACS could facilitate early detection, subsequently these indicators could be paired with other data inputs from other domains, for a cross-domain analysis and simulation that takes into account cyber/physical realities when forecasting or testing CoA.
State coordination. In our methodology, coordinated amplification could be detected through semantic fingerprinting of narratives and 'canary' sources linked to known actors, which serve as data inputs to both alerting operators, forecasting, and scenario initialization.
Deniability. Monitoring and cataloging the *modus operandi* of threat actors is crucial for reducing attribution uncertainty. In our architecture EWACS is thought as tracker for recurring TTP patterns, and the wargaming layer helps formulating the best CoAs under condition of attribution uncertainty.

6.3. Broader Picture

In short, the function of an EWACS-type system is to identify patterns and emergent trends within the information environment earlier on, opening new opportunities for decision-makers and operators. They would be able to react promptly, thus having more time to plan and craft adequate responses to the threat. They could pre-seed counternarratives tailored to resonate strongly in the demographics targeted by the disinformation artifacts, meanwhile the effort of threat actors is still growing and not yet viral.
A broader toolbox of countermeasures should include advanced network analysis, platform policy enforcement, AI-based detection (e.g., of deepfakes), media literacy initiatives, and longer-term measures, such as regulatory reform, diplomacy, and institutional strengthening. Each of these acts on different layers of the problem, and each has well-known limitations. Network analysis is technically demanding and data-constrained. Uneven policy enforcement risks legitimacy and pushes activity into harder-to-monitor spaces. AI detection fuels a costly AI arms race. Media literacy and legal reforms are slow to deliver measurable effects. In this context, an EWACS-type system coupled with a decision support and wargaming layer is not a replacement, but rather a coordinator. By detecting shifts in narratives and networks early and stress-testing alternative CoAs in simulation, it gives policymakers the time and structure needed to sequence and combine these levers. Rather than reacting piecemeal to individual artifacts that can easily be replaced, decision-makers can use EWACS-informed forecasts to transform disparate tactical measures into a coherent, multi-channel strategy that more effectively deters, disrupts, and mitigates adversarial information operations.

6.4. Limitations, caveats, and risks

Developing more assertive CoAs in the information environment using EWACS-type systems and simulation tools poses several risks that must be recognized. First, attribution in information operations is often contested. Acting on imperfect or ambiguous attribution can result in punishing the wrong actors, which can lead to a loss of credibility and trigger cycles of reciprocal escalation. Second, more coercive CoAs in a gray-zone context carry escalation risks. Signaling intended as limited "deterrence by punishment" may be perceived by other parties as disproportionate or hostile. This can result in horizontal spillover into the cyber or diplomatic realms or vertical increases in intensity. Third, liberal democracies operate under constitutional, statutory, and treaty constraints, such as protections of free expression, proportionality, and due process. If counter-disinformation measures are perceived as indistinguishable from the tactics they aim to counter, domestic trust and normative legitimacy will suffer.
There are also risks specific to the use of simulation for decision support. Any model of the information environment is necessarily simplified, and mis-specification in the simulation phase can result in

plausible but misleading narrative trajectories. This can lead to a false sense of security or steer decision-makers toward suboptimal CoAs. A related danger is automation bias or the temptation to overvalue quantitative simulator outputs relative to qualitative intelligence, political judgment, or ethical considerations, which are more difficult to encode. Finally, once sophisticated adversaries understand that EWACS-style monitoring and simulation inform policy, they may attempt to poison the system's data or exploit blind spots in its behavioral assumptions. For these reasons, this tool should help human judgment rather than serve as a substitute for it. Interventions in the information space must be designed with a clear commitment to legal standards, freedom of expression, and a rejection of any role as an "arbiter of truth".

7. Conclusion

This paper argues that contemporary information operations can be better understood, monitored, and countered by combining three approaches typically treated separately: early warning systems for hostile narratives, behavioral and CTI taxonomies for structuring observations and interventions, and professional wargaming practices. Drawing on well-established psychological mechanisms (e.g., Cialdini's principles and dual-process theory), we discussed how threat actors exploit incremental commitment, heuristic processing, and platform affordances to influence attitudes and create behavior change. We then propose an operationalization of these mechanisms into a model of the information environment, using belief strength, OCEAN traits, value alignment, and community structure to approximate how narratives diffuse across large-scale social graphs.

The core contribution of this work is the integration of EWACS-type early warning systems, a reimagined Pyramid of Pain integrated with CTI ontologies (ATT&CK and DISARM), and an LLM-based multi-agent simulation layer into a single environment that supports decision-making and training in the information space. EWACS provides continuous, multi-source monitoring of internal and adversarial narratives and network structures. The rPoP and CTI taxonomies offer a cost-of-operation lens for classifying and "scoring" defender courses of action. As the conceptual extension of a "cyber range", the simulator enables decision-makers to explore alternative courses of action, estimate potential trajectories, and practice responses prior to acting. The Tallinn Bronze Soldier case illustrates how such an integrated system could detect emerging narratives earlier, support more targeted counter-messaging, and provide a basis for wargaming different pathways and managing political risk.

However, EWACS-powered tools are only one part of a broader defense posture. They are best suited for crisis management, decision support, and institutional learning. They do not eliminate the need for long-term investments in media literacy, critical thinking, institutional resilience, and legal and regulatory frameworks that address the structural drivers of vulnerability to disinformation. Therefore, the deployment of these tools must be governed by clear ethical and legal constraints, and there must be a strong commitment to protecting freedom of expression and avoiding any role as "arbiter of truth". Future work should focus on prototyping the simulator using restricted datasets, validating its simulated trajectories with historical information campaigns, and using it in controlled red–blue wargames with practitioners to refine modeling assumptions and organizational standard operating procedures.

Acknowledgments

The authors thank the Fast Prototyping Lab of Leonardo Cyber & Security Division for invaluable discussions and feedback. This work also benefited from opportunities provided by the EUCINF project under the European Defence Fund (EDF) and was partially funded by the NextGenerationEU project "Security and Rights in CyberSpace" (SERICS).

Declaration on Generative AI

During the preparation of this work, the author(s) used ChatGPT in order to: Grammar and spelling check. After using these tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] NATO, STRATEGIC CONCEPT (2022).

[2] NATO, NATO's approach to counter information threats, 2024. URL: https://www.nato.int/en/about-us/official-texts-and-resources/official-texts/2024/10/18/natos-approach-to-counter-information-threats.

[3] C. Bjola, Algorithmic invasions: How information warfare threatens NATO's eastern flank, 2025. URL: https://archives.nato.int/nato-review-algorithmic-invasions-how-information-warfare-threatens-natos-eastern-flank.

[4] Insikt Group, Threats to the 2025 NATO Summit: Cyber, Influence, and Hybrid Risks, 2025. URL: https://www.recordedfuture.com/research/threats-2025-nato-summit.

[5] D. Psaledakis, US State Department closing office aimed at countering foreign disinformation | Reuters, 2025. URL: https://www.reuters.com/business/media-telecom/us-state-department-closing-office-aimed-countering-foreign-disinformation-2025-04-16/.

[6] A. F. Olivieri, R. Guadagno, Strategies for Combating Adversarial Information Operations: Theory and Practical Applications, in: European Conference on Cyber Warfare and Security, volume 23, 2024, pp. 341–347. Issue: 1.

[7] A. F. Olivieri, R. E. Guadagno, Integrating EWACS and the Reimagined Pyramid of Pain: Proactive Strategies against Adversarial Information Operations., Journal of Information Warfare 24 (2025).

[8] C. Shao, G. L. Ciampaglia, A. Flammini, F. Menczer, Hoaxy: A platform for tracking online misinformation, 2016, pp. 745–750.

[9] M. Yankoski, T. Weninger, W. Scheirer, An AI early warning system to monitor online disinformation, stop violence, and protect elections, Bulletin of the Atomic Scientists 76 (2020) 85–90. URL: https://www.tandfonline.com/doi/abs/10.1080/00963402.2020.1728976. doi:10.1080/00963402.2020.1728976, publisher: Routledge.

[10] P. Perla, The art of wargaming: A guide for professionals and hobbyists, United States naval institute, 1990.

[11] G. Longley-Brown, Successful Professional Wargames: A Practitioner's Handbook, History of Wargaming Project, 2019.

[12] G. Technologies, Gerulata Sentinel, 2020. URL: https://www.gerulata.com/products/gerulata-sentinel.

[13] Joint Chiefs of Staff, JP 3-13: Information Operations (2012).

[14] G. S. Jowett, V. O'donnell, Propaganda & persuasion, Sage publications, 2018.

[15] C. Ivan, I. Chiru, R. Arcos, Hybrid Security Threats and the Information Domain: Concepts and Definitions, in: Routledge Handbook of Disinformation and National Security, Routledge, 2023, pp. 9–19.

[16] D. Selvage, Operation "Denver": The East German ministry of state security and the KGB's AIDS disinformation campaign, 1985–1986 (Part 1), Journal of Cold War Studies 21 (2019) 71–123. Publisher: MIT Press One Rogers Street, Cambridge, MA 02142-1209, USA journals-info ....

[17] D. Selvage, Operation "Denver" The East German Ministry for State Security and the KGB's AIDS Disinformation Campaign, 1986–1989 (Part 2), Journal of Cold War Studies 23 (2021) 4–80. Publisher: MIT Press One Rogers Street, Cambridge, MA 02142-1209, USA journals-info ....

[18] T. Rid, Active measures: The secret history of disinformation and political warfare, Profile Books, 2020.

[19] I. V. Pasquetto, A. F. Olivieri, L. Tacchetti, G. Riotta, A. Spada, Disinformation as Infrastructure:

Making and maintaining the QAnon conspiracy on Italian digital media, Proceedings of the ACM on Human-Computer Interaction 6 (2022) 1–31. Publisher: ACM New York, NY, USA.
[20] D. Fallis, What is disinformation?, Library trends 63 (2015) 401–426. Publisher: Johns Hopkins University Press.
[21] H. P. Randolph, D. Labriny, A. DiOrio, Historical Disinformation Practices: Learning From The Russians, in: Routledge Handbook of Disinformation and National Security, Routledge, 2023, pp. 59–83.
[22] R. E. Guadagno, K. Guttieri, Fake news and information warfare: An examination of the political and psychological processes from the digital sphere to the real world, in: Research anthology on fake news, political warfare, and combatting the spread of misinformation, IGI Global, 2021, pp. 218–242.
[23] R. E. Petty, J. T. Cacioppo, The elaboration likelihood model of persuasion, in: Advances in experimental social psychology, volume 19, Elsevier, 1986, pp. 123–205.
[24] R. B. Cialdini, R. E. Petty, J. T. Cacioppo, Attitude and attitude change, Annual review of psychology 32 (1981) 357–404. Publisher: Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA.
[25] R. B. Cialdini, Pre-Suasion: A Revolutionary Way to Influence and Persuade, first edition ed., Simon & Schuster, 2016.
[26] R. B. Cialdini, Influence, New and Expanded: The Psychology of Persuasion, first edition ed., HarperCollins Publishers Inc, 2021.
[27] R. E. Guadagno, Psychological processes in social media: Why we click, Academic Press, 2025.
[28] L. Bittman, The KGB and Soviet Disinformation: An Insider's View, pergamon-brassey's ed., Washington, 1985.
[29] H. Zheng, D. H. Goh, E. W. J. Lee, C. S. Lee, Y. Theng, Understanding the effects of message cues on COVID-19 information sharing on Twitter, Journal of the Association for Information Science and Technology 73 (2022) 847–862. Publisher: Wiley Online Library.
[30] X. Lin, P. R. Spence, K. A. Lachlan, Social media and credibility indicators: The effect of influence cues, Computers in human behavior 63 (2016) 264–271. Publisher: Elsevier.
[31] J. Roozenbeek, S. Van Der Linden, The fake news game: actively inoculating against the risk of misinformation, Journal of risk research 22 (2019) 570–580. Publisher: Taylor & Francis.
[32] R. E. Guadagno, R. B. Cialdini, G. Evron, Storming the servers: A social psychological analysis of the first internet war, Cyberpsychology, Behavior, and Social Networking 13 (2010) 447–453. Publisher: Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA.
[33] B. Lilly, Russian information warfare : assault on democracies in the cyber wild west, Naval Institute Press, 2022. URL: https://books.google.com/books/about/Russian_Information_Warfare.html?id=ek7TzgEACAAJ.
[34] C. Wittenberg, A. J. Berinsky, Misinformation and its correction, Social Media and Democracy: The State of the Field, Prospects for Reform (2020) 163–198. doi:10.1177/1529100612451018, publisher: Cambridge University Press ISBN: 9781108890960.
[35] J. S. Nye Jr, Deterrence and dissuasion in cyberspace, International security 41 (2016) 44–71. URL: https://direct.mit.edu/isec/article/41/3/44/12147/Deterrence-and-Dissuasion-in-Cyberspace, publisher: MIT Press One Rogers Street, Cambridge, MA 02142-1209, USA journals-info ....
[36] J. Pamment, B. Palmertz, Deterrence by Denial and Resilience Building, in: Routledge Handbook of Disinformation and National Security, Routledge, 2023, pp. 20–30.
[37] S. J. Terp, P. Breuer, DISARM: A Framework for Analysis of Disinformation Campaigns, Proceedings - 2022 IEEE International Conference on Cognitive and Computational Aspects of Situation Management, CogSIMA 2022 (2022) 1–8. doi:10.1109/COGSIMA54611.2022.9830669, publisher: Institute of Electrical and Electronics Engineers Inc. ISBN: 9781665483308.
[38] B. E. Strom, A. Applebaum, D. P. Miller, K. C. Nickels, A. G. Pennington, C. B. Thomas, MITRE ATT&CK: Design and Philosophy (2020).
[39] D. Bianco, Enterprise Detection & Response: The Pyramid of Pain, 2014. URL: http://detect-respond.blogspot.com/2013/03/the-pyramid-of-pain.html.

[40] R. R. McCrae, P. T. Costa, Validation of the five-factor model of personality across instruments and observers., Journal of personality and social psychology 52 (1987) 81. Publisher: American Psychological Association.
[41] NATO, Cognitive Warfare: Strengthening and Defending the Mind - NATO's ACT, 2023. URL: https://www.act.nato.int/article/cognitive-warfare-strengthening-and-defending-the-mind/.