ADVANCE SOCIAL SCIENCE ARCHIVE JOURNAL
Available Online: https://assajournal.com
Vol. 05 No. 02. April-June 2026.Page# 2367-2382
Print ISSN: 3006-2497 Online ISSN: 3006-2500
Platform & Workflow by: Open Journal Systems

AI, Information Warfare and National Security

Muhammad Nayyer Adnan
M.Phil Scholar in International Relations at Riphah International University, Islamabad.
nayyeradnan9955@gmail.com

Abstract
AI is transforming information warfare by making it faster, bigger, more personalized, and more authentic. Previous discussions of cyberwar were primarily on networks, coercion, and the boundaries of cyber action; the current AI era introduces a second layer where AI language models, synthetic media, automated accounts, recommender systems and data-driven targeting can help change the political and security landscape before the onset of conventional war. In this paper, prominent journal articles related to cyber conflict, disinformation, social bots, deepfakes, misinformation psychology, autonomous military systems, and AI-driven strategic competition are summarized. It does not suggest that AI will supplant the traditional elements of power, diplomacy, intelligence, or military readiness. Instead, it reduces the cost of deception, shortens decision time, and undermines the social underpinnings of national security: public confidence, institutional legitimacy, crisis communication, electoral confidence, and civil-military cohesion. The paper outlines a national security system that is resilient, not censorship. Fast attribution, public communication without fear of censorship, media and digital literacy, government information that is authenticated, accountability of platforms, tools for AI provenance, lawful counterintelligence, crisis hotlines, and institutional safeguards for human judgement. The core principle is that a comprehensive, national security-focused policy framework for AI-driven information warfare is best served by embedding technology policy, democratic accountability, strategic communication, cyber defense, and public trust within it.
This paper features speeches by academics and experts on a variety of subjects related to cyber conflict and national security, particularly in the light of the new technologies that are making a strong impact on the world.

Introduction

Information warfare is nothing new; AI is shifting its nature. Generate convincing text, images, video and audio, in large numbers; deliver that material through automated or semi-automated networks; and customize messages to specific communities nearly in real time, these are all tactics that have been part of statecraft for a long time, but what has changed is the capacity to do it on a scale that was previously unimaginable. The early literature of the information age envisioned a form of conflict where actors, not just those with territorial or mass interests but also those with information interests, would be preferred. (Arquilla & Ronfeldt, 1993) Subsequent cyber-security scholarship, however, warned that cyberspace was not a new arena where traditional instruments of power could be superseded, but rather an instrument that is used in conjunction with other instruments of power, primarily political, economic and military

(Gartzke, 2013; Kello, 2013; Rid, 2012). Information warfare with AI is no different: It is not a magic bullet but can exacerbate existing society and state vulnerabilities.

The threats to national security are not only in the form of misleading information, but the loss of accurate information in the public domain. Disinformation can be defined as the dissemination of information with the intent to deceive, whereas misinformation can be falsely presented or inaccurate, but with no malicious intent (Fallis, 2015; Tandoc et al., 2018). In reality, hostile influence campaigns frequently use a combination of true information, selective framing, manipulated images, invented statements, and emotional appeals. Democratic media systems, and hybrid media systems in particular, are particularly vulnerable as they rely on trust, plural debate and independent verification in open communication environments. The digital age is often described as a disinformation order, in which hostile information is predominantly social and news ecosystems are increasingly polarized by a lack of trust in institutions (Bennett & Livingston, 2018; Freelon & Wells, 2020).

The reason why AI is important is that it affects the cost curve. Language models can generate fluent articles, comments, speeches, emails, and social posts quickly, and image or video systems can create content that feels as if it's real for normal users. Results also indicate that AI's propaganda can be persuasive, and that the content of political messages written by AI may be challenging for recipients and institutions to distinguish from human-produced communication (Goldstein et al., 2024; Kreps & Kriner, 2024; Kreps et al., 2022; Spitale et al., 2023). In parallel, research on misinformation reveals that misinformation has the ability to spread rapidly, that some people are more likely to be exposed to it, and that social bots can amplify low trustworthy information (Grinberg et al., 2019; Shao et al., 2018; Vosoughi et al., 2018). AI thus helps to boost not only supply but also potential reach.

The following paper makes three arguments. The first is that AI-powered information warfare is a national security issue, not just an opinion-wars issue. It can mislead leaders in times of crisis, sway public opinion before an election, demoralize the military and erode trust in official warnings. Second, AI is not a threat to the human factor. Human operators select goals, story lines, targets, time and escalation levels; AI provides speed, scale, and plausible output. Third, it is not a technical solution or a blanket ban on information but an approach which integrates into a national resilience model, verified information, accountable institutions, cyber defense, media literacy, lawful intelligence and human-centered strategic judgment.

Method and Scope

The method that is used in this paper is narrative literature review. It compiles peer reviewed journal articles from security research, political communication, computer science and psychology and AI governance. This is not about quantifying one campaign or creating a predictive model, it is about cross-disciplinary connections between studies of different campaigns. This is suitable because the information warfare is a systems problem, which can be addressed with the help of artificial intelligence. It entails technologies that create content, platforms that help distribute it, institutions that react to it, and publics that interpret it.

The scope is intentionally defensive and policy oriented. The paper does not offer any operational guidance in deception, targeting or manipulation. Instead it spells out procedures that national security institutions must comprehend to safeguard the citizens, democratic processes, crisis stability and legitimate governance. Existing literature covers cyber conflict and strategic stability, AI and military decision-making, deepfake detection, social bots, diffusion of fake news, misinformation correction, and behavioural inoculation. These fields clearly demonstrate the need for examining technological capability and social vulnerability in relation to one another.

Many of the empirical studies are conducted on specific platforms, languages, countries and time periods which is a limitation of the literature. The results of one election or one conflict or social network are not automatically translatable to all societies. Therefore, the paper does not take the stance of predicting evidence, but of suggesting it as a tool for risk assessment. A national security framework should build on global research and local data, local languages, legal protection and ongoing evaluation. Sources were selected on the basis of their relevance to artificial intelligence, information warfare, cyber conflict, misinformation, deepfakes, social bots, or strategic stability, and were prioritized if the source was cited or was a high impact journal. This places the paper within the scholarly evidence, yet provides for the synthesis of policies between disciplines. It also steers clear of the technical discussion of AI as a mere technology and explores the political, psychological, legal and military systems in which an AI-generated manipulation creates security effects.

Conceptual Foundations

The key starting point for an analysis is to distinguish between three intertwined domains: information warfare, cyber conflict, and AI-enabled manipulation. Information warfare is the use of information, communication systems and perception management to influence an opponent's will, decisions, or cohesion. Cyber conflict is the activity of interacting with or through digital networks: espionage, disruptions, sabotage, signaling. AI manipulation relates to creating, choosing, targeting, translating, imitating or exaggerating content with AI systems. These domains are becoming more and more interwoven, but they are not the same. Cyber operations can either steal data or disrupt systems without convincing anyone; disinformation can be spread without hacking; AI can help both by generating content; segmenting audiences or automating dissemination.

The historical cyber debate is helpful for the reason that it curtails claims of novelty. Rid (2012) suggested that cyber operations do not typically resemble the classic criteria of war alone, and Gartzke (2013) suggested that cyberspace is more apt to be a force that reinforces the pre-existing power structures rather than a force that changes them. In another perspective, Kello (2013) suggested that digital activity can generate "meaningful harm" below the "war level," in what he termed a "strategic gray space.Kello (2013) focused on a different aspect, positing a "strategic gray space" in which digital activity can have meaningful harm without being a conventional war. Nye (2017) presented a non-nuclear model of cyber deterrence that he called “denial, punishment, entanglement, and norms.” The arguments are relevant to AI, since influence operations are also gray: They can be very serious without being readily considered an armed attack.

Legal and conceptual definitions are also called into question by AI-powered information warfare. One fake tweet, automated account or fake article may seem insignificant. If a coordinated campaign is used, including stolen documents, AI-generated commentary, microtargeted narratives, and bot amplification, it can impact diplomacy, elections, and crisis stability. The strategic question is not if one piece of content is enough, it is if a campaign is enough to shift the information landscape in which decision makers and citizens make choices. This aligns with previous studies which suggest that fake news can't be limited to falsity, as it also involves source deception, intent, genre imitation, and platform dynamics (Molina et al., 2021; Tandoc et al., 2018).

Definitions have policy implications as well. States should not cast too wide a net or they run the risk of stifling legitimate dissent, investigative journalism, satire or political disagreement. On the other hand, if they narrow it down too much, they may be missing out on coordinated manipulation using half-truths instead of blatant lies. Communication scholars highlight the

political nature of disinformation, which is communication within institutional dynamics, media incentives, and platform governance (Bennett & Livingston, 2018; Freelon & Wells, 2020), and Fallis (2015) underscores intentional misleading as a key component of disinformation. A national security strategy should then be able to differentiate between hostile deception and normal pluralism, and bolster free discussion and limit covert manipulation of the public sphere.

AI as a Production, Targeting and Tempo Multiplier.

Production is the first big impact of AI. In the old days before the proliferation of generative systems, there were writers, designers, translators, editors and account managers in charge of influence campaigns. Those roles do not disappear when the use of AI takes the place of the time and skill required to create believable content in various languages and styles. AI-generated text has been the subject of research, with the results indicating that machine-generated text can mimic media formats to create content that can be convincing to large numbers of users (Kreps et al., 2022; Spitale et al., 2023). Recent experimental research suggests that AI-produced propaganda can be very persuasive, and human curation, or faster editing, can boost its efficacy even more (Goldstein et al., 2024). It also signals the possibility that the biggest challenge for bad actors could move from content production to building credibility, operational security, and distribution infrastructures.

The second one is personalization. Operators may use AI systems to help them segment audiences, find out what they are complaining about, create different versions of a story, and determine which version is accepted. This is not to say that all campaigns will have a perfect targeting. It implies that bad guys can conduct more experiments for less money. As for political representation research, machine-generated messages to legislators demonstrated that AI can get pretty close to constituent messaging without it being obvious enough to be detected even when examined carefully (Kreps & Kriner, 2024). In terms of national security, the same ability can be applied as a deluge of artificial public pressure against ministries, embassies, the media, and crisis-response groups.

The third is tempo. The initial hours following a crisis are prime targets for information warfare, as they are when facts are not complete and emotional responses are high. In the midst of a journalist, official or investigator's verification, AI can feed the story writer with several stories.Even before journalists, officials or investigators have finished their verification, AI can feed the story writer with several stories. Authorship of disinformation, which was once difficult to create and easy to identify, becomes more easy and difficult due to generative AI, according to Feuerriegel et al. (2023). The situation becomes particularly bad in the times of terror, boundary conflicts, ethnic tension, public health crisis or election disputes. Hostile narratives can fill the void when official information does not come or when it comes late or is contradictory.

The fourth effect is multilingual reach. The use of AI translation and style transfer can help to reduce the language barriers to influence between regions and diaspora communities. A campaign may utilize the same strategic claim at the local level, where it is translated, adapted, and positioned in the context of language, idiom, and identity. This poses a problem of defense for the states that have multilingual societies and overseas populations. There is no single sentence in public communication that can convey all the information that is required; it needs to be timely, local, verifiable and shareable. It is not a story competition. Citizens' capacity to see authoritative content before being overwhelmed by manipulated content as the explanation.

But AI-generated content isn't necessarily effective. The exposure to misinformation is not uniform across individuals and audiences are not passive recipients. Research on fake news consumption and sharing indicates that it is consumed and shared by relatively small groups of users (Guess et al., 2019; Grinberg et al., 2019). This discovery must not cause panic. The issue is

not that all citizens will accept all synthetic claims, but that credibility will be put under strain by precisely targeted influence, in the way it can drain verification resources, set an agenda, and produce a sense of pressure or give an opponent plausible deniability. The scene of AI does not mean the scene of risk, but rather the scene of increasing risk, in terms of scale, speed and ambiguity.So, AI does not lead to the scene of risk, it leads to the scene of increasing risk, on scale, on speed, on ambiguity.

Synthetic Media, Deepfakes and the Erosion of Evidence

The most visible type of AI-powered deception is deepfakes and synthetic media. Audio clones can impersonate officials, military leaders, journalists or community leaders. Video manipulation can put a public figure in a scene he or she never actually existed in, or change the meaning of actual video. Using image generation, fake evidence of violence, disaster, corruption, or foreign involvement can be created. Deepfake technology has progressed in both generation and detection, as evidenced by surveys, with generation capabilities constantly progressing and detection being a moving target (Mirsky & Lee, 2021; Nguyen et al., 2022; Tolosana et al., 2020). The strategic threat of synthetic media isn't just about being successful in forgery. Even fake news is fake news can lead to misinformation, a delay in denying that information, and the polarisation of the audience that believes or does not believe it. Even when they are not convinced, deepfakes can create uncertainty and distrust in news, as demonstrated by Vaccari and Chadwick (2020). Weikmann and Lecheler (2023) highlight the specific influence of visual disinformation, as images and videos may be interpreted as evidence, rather than arguments. This is especially important in emergencies such as a visual information spread faster than official verification.

The second risk is the liar's dividend: When synthetic media is widespread, actual evidence will be branded as bogus. Kietzmann et al. (2020) highlight the direct deception hazards posed by deepfakes and general trust issues faced by organisations and publics. In the field of national security, this may create a lack of accountability for real wrongdoing, weaken judicial mechanisms, and have a negative impact on documenting war crimes. The public sphere can be commodified when each and every video can be said to be a construct, and each and every denial can be said to be a cover up.

Social identity also influences synthetic media. A fabricated clip is hardly ever assessed based on how well the technical aspects are done, it is rather assessed in terms of set beliefs, grievances, and group allegiances. This ties in with the study of misinformation psychology. Claims that align with expectations are more likely to be accepted by people, but analytic thinking skills and awareness of the sources can help to improve one's ability to discern (Pennycook & Rand, 2019). Thus, a strictly technical defense is not enough. Detection, provenance and watermarking are useful but need to be supplemented with public knowledge, swift official clarification and independent verification.

The policy question is how to not be complacent and overreactive. States should be able to confirm media authenticity in a timely manner, but not blanketly deny the veracity of journalistic and civil society evidence because of the threat posed by deepfakes. A credible response needs to define verification standards, implement chain of custody protocols, enable independent fact-checking and have legal liability around harmful impersonation. The aim is not to prevent the spread of falsehoods, but to cut down the gap between falsehoods and public interpretations. It is in that window when panic, retaliation, mob pressure and diplomatic miscalculation are most probable.

Platforms, Bots and Networked Amplification

Information warfare is effective when information comes to the distribution medium. Previous research on fake news, bots and network diffusion also demonstrates that in some instances low-credibility content can be spread faster on a digital platform. In a dataset of tweets, Vosoughi et al. (2018) documented that misinformation is more likely to be shared than the truth, particularly when it comes to political information. Shao et al. (2018) revealed that social bots were more heavily involved in the dissemination of low credibility content, particularly during early amplification. According to Ferrara et al. (2016), social bots are automated accounts that can act as users and engage in proportionate interactions. In the age of AI, these are key elements as the generative systems can provide the content that automated networks distribute.

But the empirical situation isn't just one where misinformation makes it to everyone. The fake news consumption and sharing was found to be concentrated by Guess et al. (2019) and Grinberg et al. (2019). This is of importance in national security policy. Public alarm regarding misinformation can lead to over-censorship or a general mistrust of the public, whereas targeted analysis can identify high-risk channels, communities, and moments. If these influential accounts, pages or groups are included in politically active networks, or unintentionally reinforced by journalists and officials, their narratives can influence.

Coordinated inauthentic behavior frequently involves automation with a human operator. Linvill and Warren (2020) reported that the troll-factory can fabricate local political identities, capitalise on social divisions, and switch genders. Russian trolls on Twitter and YouTube during the 2016 U.S. presidential election were not operating in a parallel universe, but rather were engaging with existing political communication (Golovchenko et al. 2020). In the field of Web-based political discourse, Stella et al. (2018) have reported that bots can amplify negative and inflammatory messages. In the studies that have been conducted, it has been shown that there is no need for hostile operations to create all the grievances; they can intensify these.

In addition to the two complications, AI makes two more difficult. One is that automated accounts are now capable of generating a wider range of language, which makes repetitive signals more difficult to spot as they were once.One is that the automated accounts now have the ability to generate a more diverse range of language, making the repetitive signals a little trickier to spot than they were before. Second, the synthetic personas can only interact with users for longer periods of time, which is much more effective in shaping an agenda, burying users, and infiltrating communities. Detection should go beyond comparing content to include analysis of behavior, mapping infrastructure, cross-platform coordination, and transparency around account provenance. Shu et al. (2017) and Zhou and Zafarani (2020) demonstrate that detecting fake news cannot depend on a single signal, but must be based on content, social context and propagation features.

The national security issue is obvious, distribution networks are strategic terrain. States need to be vigilant against hostile influence operations, but regulation and protections of civil liberties are crucial. The counter-disinformation strategy should be coordinated deception, foreign covert influence, impersonation and incitement and manipulation of public systems. It shouldn't be done to label criticism as threat. A resilient state promotes the flow of open debate, uncovers hidden infrastructure, diminishes the ability to amplify manipulated content and allows citizens to verify claims against authoritative public information.

Cognition, Trust and Societal Resilience

The information warfare that is done with the help of AI is not just about the technology but also involves human cognition. Misinformation endures because people also develop mental models, remember what sounds familiar, and even stick with corrected info. Lewandowsky et al. (2012)

make the point that correcting someone is more impactful when it offers an alternative explanation instead of just stating that a claim is incorrect. This is important for communicating with the public in times of crisis. If a citizen is denied, he or she may have a gap in his/her story that an adversary can exploit.

Analytic thinking will make you more resistant to falsehoods. Pennycook and Rand (2019) discovered that the susceptibility of partisan fake news is more due to lack of reasoning than to motivated reasoning. This does not imply that ideology is not an issue, but rather one that can be addressed by the promptings of accuracy, reflection, and source checking. Behavioral science can foster truth and limit misinformation when interventions are experimented with and tailored for platforms, according to Lorenz-Spreen et al. (2020). Roozenbeek and van der Linden (2019) demonstrated that inoculation via a game could induce resistance to common misinformation techniques, including impersonation, emotional manipulation, polarization, conspiracy framing, trolling and deflection.

The results are in line with the resilience model. The role of public education is to not only inform citizens what to think but also to show them how to think. It should educate citizens on how to manipulate, how to check their sources, how to look at visual evidence, and how to hit the reset button before sharing crisis information! The goal is to equip students to reach civic competence rather than to control their beliefs. From a national security point of view, a public that is able to identify manipulation is a strategic good. It makes it less attractive to engage in hostile influence operations and less reliant on emergency censorship.

The issue of trust still remains challenging. According to Lazer et al. 2018, fake news is a scientific and civic issue that must be tackled from a multidisciplinary perspective. When people don't trust in any institutions, even if it was right information, it doesn't work. Hostile narratives are more convincing if officials communicate slowly, selectively or defensively. Routine transparency, data access, responsive government and the public disclosure of government errors, then, must be pursued in order to foster trust prior to a crisis. The most effective counter-disinformation system is a state that has a regular body of information that is believable.

Journalists, researchers and civil society verifiers also must be protected as part of psychological resilience. Governments do not like independent scrutiny, but verified information is more credible. If all that is permitted to say is from state sources then hostile actors can spin all corrections as propaganda. Media, universities, courts and official bodies working together on facts, the manipulation is more difficult. The institutional pluralism can serve as a defensive network in the AI era.

AI, Cyber Conflict and Military Decision-Making

AI also impacts the nation's security in fields such as military decision-making, cyber operations, and strategic stability. AI in war: The speed, autonomy, prediction, and the uncertainty of the debate. Horowitz (2018) suggests that AI can impact international competition and balance of power by altering the distribution of military capabilities and the importance of data, talent, and organizational adoption. Payne argues that AI is a potential revolution in strategic affairs, as it could change the way command, control, targeting and the human element in war is done (2018). Similarly, Ayoub and Payne (2016) associate AI with strategy, noting that smart systems impact the way decisions are made both operationally and politically.

Central security risk isn't just autonomous weapons, it's compressed decision time. However, Horowitz (2019) suggests that lethal autonomous systems can become a deterrent and a risk of instability if they speed up the escalation of conflict. If leaders are subject to "opaque systems" or fall for machine-generated indicators in the heat of crisis, AI can have an impact on nuclear risk and crisis stability, Johnson (2019a) notes. Goldfarb and Lindsay (2022) provide a corrective

by stating that the use of AI makes human judgment more important because there is still room for interpretation, values, and contextual choice when it comes to prediction. Safeguarding the nation therefore becomes an organizational challenge: how to leverage AI without compromising the accountability or strategic judgment.

Autonomy also entails some normative and legal issues. Autonomous weapon systems can have an impact on strategic stability in a variety of ways, including altering perceptions of offense/defense and reducing human control, warns Altmann and Sauer (2017). Bode and Huelss (2018) analyse international norms challenged by autonomous weapons, and Maas (2019) discusses arms control challenges in the context of military AI. These debates relate to information warfare, as misinformation, miscommunication, or manipulation may impact automated decisions, and the use of AI-based command structures could amplify the effects of misinformation.

The relationship between AI and Cyber is particularly crucial. Artificial Intelligence can exacerbate security dilemmas in combination with cyber operations, enhancing offense, defense, reconnaissance, deception, and decision support (Johnson, 2019b). It's not just code that determines the cyber offense-defense balance, warns Slayton (2017). Smeets (2018) also points out that cyber weapons are ephemeral: vulnerabilities can be fixed, tools can be published, capabilities can rapidly become obsolete. This means that any cyber advantage that can be achieved with the help of artificial intelligence (AI) might exist but be precarious, leading states to be secretive, exploit the technology quickly and continuously adapt to it.

Empirical research on cyber conflicts also debunks hyperbole. Valeriano and Maness (2014) discovered that disputes in the cyber domain are generally small and controlled compared to expectations of worst-case scenarios. In the cases of Ukraine and Syria, Kostyuk and Zhukov (2019) identified few indications that cyberattacks were decisive in determining battlefield violence. However, this doesn't imply low importance. While the outcome of a cyber or information operation cannot be guaranteed, the influence it exerts on morale, logistics, diplomacy and public perception remains. AI adds to this significance by enabling actors to leverage freely on stolen information, craft compelling narratives, and stage cyber events alongside influence operations.

The key word for defense planners is "restraint. AI can enhance intelligence analysis, anomaly detection, logistics, language translation and crisis monitoring. It can also lead to automation bias, escalation risk and brittleness. There must be explainable procedures, red-teaming, audit trails and fail-safe communication channels for human commanders and civilian leaders. Scarcely as important as capability is signaling, doctrine, and trust that machines won't initiate uncontrolled escalation in strategic stability.

National Security Implications

There are multiple ways in which AI-driven information warfare impacts national security. The first is electoral legitimacy. AI-generated content has the capability to mimic local voices, create candidate statements, overload complaint systems and influence communities with suppression narratives. Not only are there changed votes, but there is a belief that no vote is trustworthy. The other way is crisis escalation. Publics and Leaders can spiral toward retaliation before they are verified in the face of fabricated videos or bogus casualty reports, or impersonated official messages. The third pathway is military morale and civil-military trust. A lack of cohesion can result from synthetic leaks, fake orders, and targeted rumors during conflict situations.

The fourth way is Diplomacy Coercion. AI can create false narratives to demonize a state abroad, incriminate it with fake atrocities, or leverage against its negotiating partners. Economic security is the fifth pathway. Misinformation about banks, energy systems, food supplies or public health

can lead to a panic or disruption of the market. The sixth pathway is 'internal cohesion'. By using manipulated evidence, influence campaigns can make ethnic, sectarian, regional, and ideological divide appear even larger than it is. In none of these ways is the entire citizenship deceived. It's sufficient to cause confusion on key audiences at key times.

AI is also an attribution threat. There is a suspicion from the state of foreign involvement but no evidence that can be released to the public. Influence infrastructure can operate inter-jurisdictional on a private network. Domestic actors can unwittingly enhance foreign-origin stories. This complexity could make it hard to act or make impetuous accusations. The literature on cyber deterrence cautions that there are challenges with attribution, proportionality, and signaling in digital conflict (Nye, 2017; Slayton, 2017). AI-driven information operations have these issues and inherit content ambiguity. The false story can be foreign, domestic, commercial, ideological and opportunistic.

National security agencies must have a multi-layered answer. Intelligence services to monitor hostile infrastructure and foreign coordination. Government systems should be secured, a check should be made for impersonation and assistance provided with forensic analysis should be provided by cyber agencies. Information ministries/public communication offices should release timely and evidence-based updates. Election bodies should pre-bunk manipulation claims prior to polling. Impersonation, harassment and incitement should be dealt with by the courts and regulators in accordance with due process. Education systems should be developing literacy over a long term. The problem belongs to no only one institution.

There has to be a way to keep the risk of the securitization under control. To regard all misinformation as a security threat is a negative impact on rights and legitimacy. Allowing state censorship of legitimate criticism in the name of counter-disinformation can reinforce the distrust used by hostile actors. The best one is evidence based and behavior based: identify coordinated inauthentic activity, foreign covert operations, fake official communications, fake identities, incitement to violence, manipulation of critical systems. Guard freedom of expression, freedom of the press, freedom of political opposition and freedom of public grievance. Resilience requires legitimacy.

Operational Scenarios for National Security Planning

The literature can be turned into readiness through scenario planning. A first scenario is an election in which there is a contest. Ballots have been lost, a candidate dropped out, or polling stations are closed, are claims made by AI-generated posts. Synthetic audio seems to be of an election official conceding that there is election fraud. Automated accounts push the claim into journalist inboxes, messaging groups. The answer must not be prepared for the polling day. Election bodies should set up official communication channels, create pages to address rumors, work with the media and share swift evidence-based corrections, but remain politically neutral.

The second scenario is a border or military crisis. A fabricated video is one that seems to depict an attack, desecration or surrender. Hashtags are calling for instant reaction before the issue is verified by military and diplomatic avenues. Here, the risk is escalation. Statements from the military and foreign ministry, back-channel verification with adversaries if possible, and public language should be done in a disciplined way during crisis communication. Their aim is to make it difficult for hostile forces to take control of decision time. The information created by AI cannot be trusted until it has been validated by forensic and contextual investigations.

A third is some type of terrorist violence or internal unrest. The attackers or sympathizers could exploit AI to inflate the casualty figures, assume the identity of law enforcement, blame a different sect or motivate revenge. The information environment can be exacerbated by a slow, opaque or abusive state response. Security agencies should provide accurate facts, preserve the

dignity of victims, refrain from premature attributions, and correct errors in a public manner. The other key elements of lawful investigation and transparent accountability become counter-disinformation strategies by themselves in the sense that they lessen the credibility of rumor and manipulation.

A fourth possibility is economic panic. Synthetic messages can be designed to imitate banks, energy regulators, health authorities or food-supply agencies. If the public does not have reliable channels of verification, even a false claim can lead to queues, hoarding, withdrawals or market volatility. Governments and regulated sectors should keep up maintained alert systems, open dashboards, media relationship and rehearsed denial protocols. Economic security thus needs to be backed up with communication security, financial reserves and cyber protection.

The fifth case is the isolation of the enemy by way of narrative attack. AI-generated articles, falsified documents, and fake expert commentary may allege to a state of bad faith, human rights violations, treaty breaches, or covert aggression. Denial is not the only defense. States must be documentarily prepared: with archived evidence, transparency of data, credible spokespeople, independent observers if available and speedy diplomatic action. In international politics, information warfare has the goal of inducing partners to put on their guard. With a prepared evidentiary record, that room for doubt is minimized.

The government and the defensive architecture.

The first step in creating an effective national framework would be to have verified, publicly available information. Governments need to provide official data in an official, timely, machine-readable and archived format, so citizens and journalists can verify claims in a timely manner. Official channels will provide short updates, evidence links, media assets and contact points in cases of emergency for verification. This diminishes the vacuum of which hostile narratives thrive. It also provides journalists and citizens a shared fact to anchor onto in times of uncertainty. It also promotes what could be termed information sovereignty, or the ability of the state to supply trusted information on the state's public affairs without cutting citizens off from international information.

Second, state require AI age authentication. Official speeches, orders, emergency alerts and diplomatic statements should be signed with cryptographic signatures, verified accounts and be issued across websites, social channels, and traditional media. While media provenance standards and proof of media can only help to make things more expensive and to establish the process of verification, they are not a solution to all deception. Deepfake detection is not a truth machine; it's a forensic support tool. However, the detection tools should be evaluated independently, due to the co-evolution of generation and detection (Mirsky & Lee, 2021; Nguyen et al., 2022).

Third, platform accountability should be directed towards transparency and coordinated manipulation. Platforms should make data available to researchers and regulators relating to political advertising, automated behavior, state-connected networks, takedowns and recommendations effects, with protections for privacy. Research on bots and trolls has demonstrated that network behavior can often provide more evidence of manipulation than content alone (Ferrara et al., 2016; Linvill & Warren, 2020; Shao et al., 2018). It is governments' responsibility to ask for transparency, but not use platforms as tools of political censorship.

In the fourth, national security institutions should develop analytic capacity. Media can be clustered and analyzed using AI tools that allow for coordinated timing, comparison of media artifacts, and monitoring of multilingual spaces. Human analysts, however, have to make sense of the political background and determine credibility and local grievances. Goldfarb and Lindsay (2022) caution in their warning that prediction is not judgment. If a counter-disinformation cell

is not knowledgeable about the region, legally savvy, and has no communication skill, it can mistake dissent as foreign influence or overlook covert operations disguised within the body of what looks like legitimate complaints.

Fifth, try to practice a crisis communication. In the event of a terrorist attack, a border incident, a cyber disruption or a public health emergency, response agencies should be aware of who is responsible for verifying information, who is speaking publicly, how corrections are made to information, how evidence is preserved and how counterfeit official messages are corrected. It's not only about speed, it's about credibility. A quick denial is worse than an update that comes at a later time but is based on evidence. The corrections should provide an explanation of what is known and what is unknown as well as when the next update will be available.

6. Cooperation between nations is an urgent need. Information warfare spreads across borders via platforms, cloud, finances and diasporas powered by AI. States should cooperate and share resources, including mutual legal assistance, cyber incident channels, standards for synthetic media, and exposure of foreign influence operations. While military AI arms control and confidence building dialogues should discuss information integrity and crisis stability alongside weapons platforms (Altmann & Sauer, 2017; Maas, 2019), it is crucial to avoid overlooking the other side of the coin: information integrity and crisis stability in the context of military AI.

Policy Recommendations

There are seven policy pillars to build a national security strategy for AI-enabled information warfare. First, create a legally created information integrity center under civilian supervision. It should coordinate, research, issue early warnings, and provide support to public communication; it shouldn't have absolute censorship authority. Second, develop a timely verification system for fake media featuring forensic labs, independent experts and official representatives. Third, verify government information by cryptographic signature, verified domain, public archives and emergency broadcasts redundancy.

Fourth, improve the capability of open source intelligence and counterintelligence on coordinated foreign influence. The emphasis must be networks and funding, infrastructure and impersonation, covert coordination, rather than common opposition. Fifth, develop digital skills in schools and for civil service training. The threat of deceptive AI has made it imperative for citizens, journalists, judges, police officers, soldiers, and diplomats to acquire the skills to deal with it in a responsible way. Sixth, mandate transparency of platforms regarding state-affiliated influence operations, automated amplification and political advertising data. Seventh, maintain accountability and rights through judicial reviews, Parliamentary oversight, transparency reports and rights of appeal concerning content-based actions.

Human-centric AI is the main principle that must be used for defense and intelligence. AI can be used to aid in anomaly detection, translation, triage, and pattern recognition, but critical national security matters need human leadership. This is especially true in military and cyber situations with the possibility of false inputs that could lead to escalation. Adversarial examples, manipulated data, spoofed communications and synthetic media should be all used in testing command systems. Red team exercises must be a mix of cyber intrusion, public disinformation, diplomatic pressure and domestic rumour to simulate real hybrid attacks.

When communicating with the public, the rule should be for speed; and evidence. When reassuring the public is required, officials should not use vague denials, partisan language, and avoid unnecessary secrecy. A good strategy is to publish with authority facts, to describe the process of checking, to correct errors publicly, and to make updates regularly. This method is in line with research on misinformation interventions highlighting that factual alternatives and trusted sources enhance outcomes of misinformation corrections (Lewandowsky et al., 2012;

Lorenz-Spreen et al., 2020). The message has to be multi-lingual and accessible, and localised to the media preferences.

To build resilience to democratic threats, states need to respond to grievances that are targeted by hostile narratives. Disinformation campaigns don't work in isolation. They get lodged in social mistrust, corruption, inequality, regional dissension, discrimination, or security lapses. Hence, combating information warfare is inseparable from governance reform. Good public services, equitable policing, responsible institutions, and effective government makes less juice for the bad guys. Security is not a military or intelligence matter in this sense, it's a legitimacy matter.

States should work to establish norms against synthetic impersonation of leaders during an emergency, against AI generated fake emergency notifications and against manipulation of nuclear or military command communications for international stability. These norms will not prevent all forms of conflict, but they will help attribute, collectively condemn, and respond proportionately. Human control, auditability, testing, and communication in a crisis are essential elements of military AI governance. The goal is to make sure that AI doesn't exacerbate informational noise into strategic misjudgment.

Limitations and Future Research

There are a number of restrictions which should be taken into account in future research and policy development. First, the information warfare produced by AI will be language-dependent, platform-dependent, level of literacy-dependent, context of media trust-dependent and so on, and will be different according to the language, the platform, the level of literacy, the trust with media, and the political context. English-language experiments and U.S.-centric datasets can be useful, but they do not reflect multilingual societies, encrypted messaging systems, and areas where television, radio, religious networks, and regular people (influencers) set the tone. The national security institutions should therefore provide support for local-language research and exchange anonymized data with universities, with privacy protection.

Second, the detection research is adversarial. Unfortunately, as deepfake and language-generation tools get better, detectors trained on older versions can become less effective. Identifying that content was AI-generated is not the only way the defense should rely on. It should also explore what is happening with the account, what the content is and why, when the content was created, who is funding the content, and the overall narrative purpose. This is still applicable if content is technically hard to classify.

Thirdly, governments need to assess the effectiveness of their own response measures. Takedown, labelling, arrests, public rebuttals and media campaigns can have unintentionally ‘counterproductive' effects, such as the creation of martyrs for hostile actors, heightened curiosity, or loss of trust in those campaigns when they are unfair. Policy needs to be trialed, reviewed and amended. To be effective, counter-disinformation must be legitimate.

There are some areas for further research including crisis simulations, cross-platform influence pathways, audio AI in low-literacy contexts, diaspora targeting and encrypted messaging networks, and how disinformation and real grievances intersect. But, at the heart of it, the question is what can trusted institutions do when they have to communicate truth in the face of false content?

Conclusion

AI is not the first technology to transform information warfare and it likely won't be the last. It's relevant because of the scale, speed, realism, personalization and automation. Such abilities can enhance the integrity of the governance in the context of serving the public, translation, data analysis and emergency response. They can also help enlist the support of hostile actors to make

up evidence, mimic communities, flood institutions, and undermine trust. The national security challenge is to secure the public sphere while preserving its openness to give it legitimacy.

The conclusion of this literature review is a balanced one. While cyber and AI capabilities are very strong, they are not always able to function without consideration of politics, institutions or human judgment (Gartzke, 2013; Goldfarb & Lindsay, 2022; Valeriano & Maness, 2014). Exposure to and sharing of misinformation can be uneven, and social patterns are likely to influence exposure (Guess et al., 2019; Guess et al., 2019; Vosoughi et al., 2018). The combination of forensic capabilities and public literacy and institutional credibility is a key resistance to the damage deepfakes cause to trust (Vaccari & Chadwick, 2020; Weikmann & Lecheler, 2023). Military AI can help to speed up the conflict, but human judgment, norms and organizational discipline will still be the ultimate deciding factors (Horowitz, 2019; Johnson, 2019a; Payne, 2018).

A state that is ready for AI-powered information warfare will not use censorship as a first resort nor a first solution. It will disseminate trustworthy information, validate official communication, uncover hidden manipulation, ensure independent verification, inform citizens, strengthen cyber security systems, regulate platforms transparently and hold humans accountable for making decisions about security. A national security system based on trust, evidence, legality, and strategic prudence is the best form of protection against deception by AI.

References

Altmann, J., & Sauer, F. (2017). Autonomous weapon systems and strategic stability. Survival, 59(5), 117-142. https://doi.org/10.1080/00396338.2017.1375263

Arquilla, J., & Ronfeldt, D. (1993). Cyberwar is coming! Comparative Strategy, 12(2), 141-165. https://doi.org/10.1080/01495939308402915

Ayoub, K., & Payne, K. (2016). Strategy in the age of artificial intelligence. Journal of Strategic Studies, 39(5-6), 793-819. https://doi.org/10.1080/01402390.2015.1088838

Bennett, W. L., & Livingston, S. (2018). The disinformation order: Disruptive communication and the decline of democratic institutions. European Journal of Communication, 33(2), 122-139. https://doi.org/10.1177/0267323118760317

Bode, I., & Huelss, H. (2018). Autonomous weapons systems and changing norms in international relations. Review of International Studies, 44(3), 393-413. https://doi.org/10.1017/S0260210517000614

Fallis, D. (2015). What is disinformation? Library Trends, 63(3), 401-426. https://doi.org/10.1353/lib.2015.0014

Feuerriegel, S., DiResta, R., Goldstein, J. A., Kumar, S., Lorenz-Spreen, P., Tomz, M., & Prollochs, N. (2023). Research can help to tackle AI-generated disinformation. Nature Human Behaviour, 7, 1818-1821. https://doi.org/10.1038/s41562-023-01726-2

Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The rise of social bots. Communications of the ACM, 59(7), 96-104. https://doi.org/10.1145/2818717

Freelon, D., & Wells, C. (2020). Disinformation as political communication. Political Communication, 37(2), 145-156. https://doi.org/10.1080/10584609.2020.1723755

Gartzke, E. (2013). The myth of cyberwar: Bringing war in cyberspace back down to earth. International Security, 38(2), 41-73. https://doi.org/10.1162/ISEC_a_00136

Goldfarb, A., & Lindsay, J. R. (2022). Prediction and judgment: Why artificial intelligence increases the importance of humans in war. International Security, 46(3), 7-50. https://doi.org/10.1162/isec_a_00425

Goldstein, J. A., Chao, J., Grossman, S., Stamos, A., & Tomz, M. (2024). How persuasive is AI-generated propaganda? PNAS Nexus, 3(2), pgae034. https://doi.org/10.1093/pnasnexus/pgae034

Golovchenko, Y., Buntain, C., Eady, G., Brown, M. A., & Tucker, J. A. (2020). Russian trolls on Twitter and YouTube during the 2016 U.S. presidential election. The International Journal of Press/Politics, 25(3), 357-389. https://doi.org/10.1177/1940161220912682

Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., & Lazer, D. (2019). Fake news on Twitter during the 2016 U.S. presidential election. Science, 363(6425), 374-378. https://doi.org/10.1126/science.aau2706

Guess, A., Nagler, J., & Tucker, J. (2019). Less than you think: Prevalence and predictors of fake news dissemination on Facebook. Science Advances, 5(1), eaau4586. https://doi.org/10.1126/sciadv.aau4586

Horowitz, M. C. (2018). Artificial intelligence, international competition, and the balance of power. Texas National Security Review, 1(3), 37-57. https://doi.org/10.15781/T2639KP49

Horowitz, M. C. (2019). When speed kills: Lethal autonomous weapon systems, deterrence and stability. Journal of Strategic Studies, 42(6), 764-788. https://doi.org/10.1080/01402390.2019.1621174

Johnson, J. (2019a). Artificial intelligence and future warfare: Implications for international security. Defense & Security Analysis, 35(2), 147-169. https://doi.org/10.1080/14751798.2019.1600800

Johnson, J. (2019b). The AI-cyber nexus: Implications for military escalation, deterrence and strategic stability. Journal of Cyber Policy, 4(3), 442-460. https://doi.org/10.1080/23738871.2019.1701693

Kello, L. (2013). The meaning of the cyber revolution: Perils to theory and statecraft. International Security, 38(2), 7-40. https://doi.org/10.1162/ISEC_a_00138

Kietzmann, J., Lee, L. W., McCarthy, I. P., & Kietzmann, T. C. (2020). Deepfakes: Trick or treat? Business Horizons, 63(2), 135-146. https://doi.org/10.1016/j.bushor.2019.11.006

Kostyuk, N., & Zhukov, Y. M. (2019). Invisible digital front: Can cyber attacks shape battlefield events? Journal of Conflict Resolution, 63(2), 317-347. https://doi.org/10.1177/0022002717737138

Kreps, S., & Kriner, D. L. (2024). The potential impact of emerging technologies on democratic representation: Evidence from a field experiment. New Media & Society, 26(12), 6918-6937. https://doi.org/10.1177/14614448231160526

Kreps, S., McCain, R. M., & Brundage, M. (2022). All the news that's fit to fabricate: AI-generated text as a tool of media misinformation. Journal of Experimental Political Science, 9(1), 104-117. https://doi.org/10.1017/XPS.2020.37

Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J., & Zittrain, J. L. (2018). The science of fake news. Science, 359(6380), 1094-1096. https://doi.org/10.1126/science.aao2998

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

Linvill, D. L., & Warren, P. L. (2020). Troll factories: Manufacturing specialized disinformation on Twitter. Political Communication, 37(4), 447-467. https://doi.org/10.1080/10584609.2020.1718257

Lorenz-Spreen, P., Lewandowsky, S., Sunstein, C. R., & Hertwig, R. (2020). How behavioural sciences can promote truth, autonomy and democratic discourse online. Nature Human Behaviour, 4, 1102-1109. https://doi.org/10.1038/s41562-020-0889-7

Maas, M. M. (2019). How viable is international arms control for military artificial intelligence? Three lessons from nuclear weapons. Contemporary Security Policy, 40(3), 285-311. https://doi.org/10.1080/13523260.2019.1576464

Mirsky, Y., & Lee, W. (2021). The creation and detection of deepfakes: A survey. ACM Computing Surveys, 54(1), Article 7. https://doi.org/10.1145/3425780

Molina, M. D., Sundar, S. S., Le, T., & Lee, D. (2021). Fake news is not simply false information: A concept explication and taxonomy of online content. American Behavioral Scientist, 65(2), 180-212. https://doi.org/10.1177/0002764219878224

Nguyen, T. T., Nguyen, Q. V. H., Nguyen, D. T., Nguyen, D. T., Huynh-The, T., Nahavandi, S., Nguyen, T. T., Pham, Q.-V., & Nguyen, C. M. (2022). Deep learning for deepfakes creation and detection: A survey. Computer Vision and Image Understanding, 223, 103525. https://doi.org/10.1016/j.cviu.2022.103525

Nye, J. S., Jr. (2017). Deterrence and dissuasion in cyberspace. International Security, 41(3), 44-71. https://doi.org/10.1162/ISEC_a_00266

Payne, K. (2018). Artificial intelligence: A revolution in strategic affairs? Survival, 60(5), 7-32. https://doi.org/10.1080/00396338.2018.1518374

Pennycook, G., & Rand, D. G. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. Cognition, 188, 39-50. https://doi.org/10.1016/j.cognition.2018.06.011

Rid, T. (2012). Cyber war will not take place. Journal of Strategic Studies, 35(1), 5-32. https://doi.org/10.1080/01402390.2011.608939

Roozenbeek, J., & van der Linden, S. (2019). Fake news game confers psychological resistance against online misinformation. Palgrave Communications, 5, Article 65. https://doi.org/10.1057/s41599-019-0279-9

Shao, C., Ciampaglia, G. L., Varol, O., Yang, K.-C., Flammini, A., & Menczer, F. (2018). The spread of low-credibility content by social bots. Nature Communications, 9, Article 4787. https://doi.org/10.1038/s41467-018-06930-7

Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. ACM SIGKDD Explorations Newsletter, 19(1), 22-36. https://doi.org/10.1145/3137597.3137600

Slayton, R. (2017). What is the cyber offense-defense balance? Conceptions, causes, and assessment. International Security, 41(3), 72-109. https://doi.org/10.1162/ISEC_a_00267

Smeets, M. (2018). A matter of time: On the transitory nature of cyberweapons. Journal of Strategic Studies, 41(1-2), 6-32. https://doi.org/10.1080/01402390.2017.1288107

Spitale, G., Biller-Andorno, N., & Germani, F. (2023). AI model GPT-3 (dis)informs us better than humans. Science Advances, 9(26), eadh1850. https://doi.org/10.1126/sciadv.adh1850

Stella, M., Ferrara, E., & De Domenico, M. (2018). Bots increase exposure to negative and inflammatory content in online social systems. Proceedings of the National Academy of Sciences, 115(49), 12435-12440. https://doi.org/10.1073/pnas.1803470115

Tandoc, E. C., Jr., Lim, Z. W., & Ling, R. (2018). Defining fake news: A typology of scholarly definitions. Digital Journalism, 6(2), 137-153. https://doi.org/10.1080/21670811.2017.1360143

Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., & Ortega-Garcia, J. (2020). Deepfakes and beyond: A survey of face manipulation and fake detection. Information Fusion, 64, 131-148. https://doi.org/10.1016/j.inffus.2020.06.014

Vaccari, C., & Chadwick, A. (2020). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. Social Media + Society, 6(1). https://doi.org/10.1177/2056305120903408

Valeriano, B., & Maness, R. C. (2014). The dynamics of cyber conflict between rival antagonists, 2001-11. Journal of Peace Research, 51(3), 347-360. https://doi.org/10.1177/0022343313518940

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

Weikmann, T., & Lecheler, S. (2023). Visual disinformation in a digital age: A literature synthesis and research agenda. New Media & Society, 25(12), 3696-3713. https://doi.org/10.1177/14614448221141648

Zhou, X., & Zafarani, R. (2020). A survey of fake news: Fundamental theories, detection methods, and opportunities. ACM Computing Surveys, 53(5), Article 109. https://doi.org/10.1145/3395046