119-134

ARTIFICIAL INTELLIGENCE DISRUPTION IN INFORMATION WARFARE AND INFLUENCE OPERATIONS

Gp Capt R K Dogra

"There are known knowns. These are things we know that we know. There are known unknowns. That is to say, there are things that we know, we don't know. But there are also unknown unknowns. There are things we don't know, we don't know."

- Donald Rumsfeld

Abstract

In this new age of hybrid warfare, nations have resorted to the use of Information Warfare (IW) as an unconventional method to impose their National Will on an adversary with high anonymity and without violating international laws on sovereignty. Information security as part of the National Security Strategy is an important step towards waging the war in info space, the fifth dimension of Warfare. Information warfare has lately transformed into Information Warfare and Influence Operations (IWIO) in the comprehensive international security scenario. IWIO is witnessing rapid changes due to the emergence of disruptive technologies, especially the ones affecting cyber and digital spaces. Artificial Intelligence (AI) is one such disruptive technology that has used incursions in the internet to become a permanent fixture in the daily lives of all humans. AI not only augments the offensive capabilities of non-state, malicious actors in the digital space but also weakens the defensive networks of our society against information manipulation. This realisation is increasingly pushing international institutions and their member states to question the harmful effects of AI, particularly in the present geopolitics, marked by a resurgence of information warfare. AI thus has a huge capability to influence IWIO. This paper will study how AI is disrupting the IWIO domain, how large state actors like the US, Russia, and China are using AI for IWIO and how India should respond to the use of AI in IWIO.

INTRODUCTION

IWIO are essentially intended to position your message across or to prevent your adversary from doing so in your domain. IWIO is not just about developing a coherent and convincing storyline but also involves a multitude of Psy Ops like confusing, distracting, dividing and demoralising the adversary. Influence can be mostly exerted using information.[1] However, technologies in the cyber world have enabled nations and non-state actors to influence specific audiences in more intrusive ways that may steal, destroy, inject, compromise or change information by having hidden access to information systems and networks. AI is one such technology that has the potential to incapacitate an adversary's information grid by intruding into its physical and cognitive worlds.

Many government officials, military leaders, and researchers acknowledge the evolution of information war into IWIO.[2] It has become a double-edged sword, equally important for the powerful states as well as technically poor states, non-state actors, and individual experts in software. The importance of IWIO to international security can be judged by the fact that it can shape global and domestic narratives that affect stability within states, international alliances, and the survivability of governments and leaders. As per the military and security pundits, 'Information superiority could be the single most decisive factor in present and future warfare'. IWIO is evolving at a rapid pace due to transformations in the technologies that have occurred in recent years that influence the IWIO domain. AI developments have a significant impact on IWIO because they not only enhance the speed and effectiveness of IWIO operations but also shape the specific IWIO tactics that can be employed.[3]

This astounding growth of AI is well acknowledged by the world with the public release of ChatGPT and is further supported by facts indicating how AI is increasingly becoming an integral part of government and public sector infrastructure. Along with the rapid digitalisation of government organisations and private businesses, the expansion of AI is seen in all spheres of life, enhancing opportunities and productivity for individuals and businesses all around the globe. AI represents an exciting new domain in technology with the potential to revolutionise access to data, information and interactions with the world. However, this increasing exposure to AI has also brought more risks of exploitation and manipulation. AI has functional weaknesses and, due to its novel nature, highlights the urgency of addressing its exploitation within an established and legalised framework. This paper will present IWIO in the broader context of information warfare. The second part of the paper gives a comprehensive analysis of the marriage of AI and information and researches the significant risks this relation poses to any country's security and sovereignty. The third part of the paper presents how countries like the US, China and Russia are using AI in the IWIO domain. The fourth segment of the paper deals with suggested countermeasures that may help detect or reduce the role of AI in IWIO. In the final segment, AI and IWIO will be discussed in the security context of India.

CONCEPT OF INFORMATION WARFARE

By definition[4], the core weapon and target in IW is 'information'. It is the product that has to be manipulated to the advantage of those trying to influence events. The means of achieving this influence are multiple. Sympathisers of IW can attempt to directly change the data or to deprive adversaries of access to it. The methodology used in data or information collection, storage, and dissemination can be compromised. Using more subtle techniques, the way the data is interpreted can be changed by altering the context in which it is viewed. Thus, the range of activities in the context of information warfare is unlimited. However, the first thing to be established is the nature of information itself.

INFORMATION AND IWIO

IW is a complex phenomenon, and a plethora of definitions exist for this term. Academicians and researchers have explained IW as 'the deliberate manipulation or use of information by one party on an adversary to influence the choices and decisions the adversary makes for military or strategic gain'. Whilst detailed, this definition signifies the fundamental elements of IW, namely the intentional and targeted aim to influence an adversary’s decision-making process through information. With the ever-expanding impact of technology on wars and the extent of usage of information as an element of war, information warfare has stepped up as information operations, and these operations can fall within the realms of psychological operations, operations security, military deception, and electronic warfare. The meaning of IWIO can be well explained by Lin’s definition, which is the “deliberate use of information (whether true or false) by one party on an adversary to confuse, mislead, and ultimately to influence the choices and decisions that the adversary makes”.[5]

ALL WATCHED OVER BY MACHINES OF LOVING GRACE

By Richard Brautigan 1967

I like to think
(and the sooner the better!)

of a cybernetic meadow where
mammals and computers live
together in mutually programming
harmony like pure water touching
clear sky.

I like to think
(right now, please!)

of a cybernetic forest
filled with pines and electronics
where deer stroll peacefully past
computers as if they were flowers
with spinning blossoms.

I like to think (it has to be!)
of a cybernetic ecology where we
are free of our labors and joined
back to nature, returned to our
mammal brothers and sisters, and
all watched over

by machines of loving grace.

ROLE OF AI IN INFORMATION WARFARE

As we transcend into the era of digitalisation, we encounter a definite shift in the way we manage the complex systems that govern our administrations, enterprises, and personal lives. Undoubtedly, AI has become a crucial element in this evolution. Global access to the internet was at first driven by an independent vision;[6] there are now serious and multiple concerns regarding the increasing incursion of AI into our world. It is now established that this technology is becoming a permanent fixture in our daily lives, as indicated by the availability of freely accessible tools such as image generators and conversational agents like chatbots. It is shocking to witness the rapid growth of AI Large Language Models (LLMs), such as OpenAI's ChatGPT, Meta's Llama and Google's Bard. Trained on an enormous collection of open-source data collected from the internet and containing many billions of parameters, the capability of LLMs to generate coherent, well-structured, and persuasive sentences resembling human writing has alarmed experts. Consider the term cognitive fluency bias, an extensive subject in academic AI research in recent times. Cognitive fluency[7] bias - when people mistakenly equate polished presentation for authenticity - can mislead. This bias is deeply rooted, often influencing one's perceptions and decisions without their conscious knowledge. Cognitive fluency bias is especially prone to 'truthiness', truthiness is 'how smart, sophisticated people use unrelated information to decide whether something is true or not'. Truthiness illustrates how high-quality presentation whether through well-crafted text or compelling visuals can make statements appear more truthful. In Newman's words, 'when things feel easy to process, they feel trustworthy'.

Malevolent minds can exploit AI-generated content to take advantage of cognitive fluency bias and truthiness, thereby highly influencing people's intuitive thinking. These 'gut feelings'-the cognitive mechanisms for rapid and mostly accurate decision-making are rooted in the brain's evolved heuristics for judgments. This logical presentation style of AI-generated content projects a feeling of intelligence and aligns with the heuristic to accept certain statements at face value without extensive scrutiny to differentiate factual data from fiction. This use of AI for disseminating false information that bypasses mindful scrutiny is worrisome, more so because AI language models can be used to prepare fictional or fake messages intentionally targeting large segments of the population. These AI-generated, repetitive messages cause it to appear more reliable, a phenomenon called the 'illusory-truth effect'.

This realisation of illusory truth or fabricated realism is prompting international institutions and their member states to caution the countries against the harmful effects of AI[8], especially in a geopolitical context marked by the increased use of information warfare. AI gives an edge to the offensive capabilities of states and malicious private actors engaged in cyber and information warfare. Armed with tools and technology to generate false information, AI can contribute hugely to the intensification of information fog. While safeguards exist to restrict these tools from accessing requests that are considered to be unsafe and malicious, there are loopholes in their moderation. It is possible to 'jailbreak' ChatGPT, which means bypassing the restrictions and safeguards imposed by OpenAI. Indeed, the AI tools being developed and deployed by these malign actors hold adequate potential to offer narratives tailored to match specific cultural and linguistic contexts, targeting diverse sociological and ethnic groups to ensure better responses. With the help of AI, thus, it becomes possible to produce low-cost, high-quality propaganda narratives, doing away with the need to hire humans who are domain experts.

The ability of ChatGPT to code in different programming languages helps in the creation of online sites and networks of automated accounts (botnets), which in turn can be directed to intensify the formulated narratives, a phenomenon known as astroturfing. This practice is likely to attract malicious activities from state and private actors in two different ways. It can significantly increase the creation of botnets, which will be active on social networks, more importantly when social platforms such as Metaverse and Twitter are doing massive layoffs of those who were engaged in moderation tasks. Secondly, cybercriminal operations, such as phishing, will increase, enabled by ChatGPT, which may help in the rapid generation of fake emails with a convincing level of realism capable of bypassing traditional anti-spam filters.

EMPLOYMENT OF AI IN IWIO BY THE US, CHINA AND RUSSIA

THE US

In the preceding paragraphs, we have seen how AI can enhance IWIO by increasing the pace of IWIO operations and how it can be applied to a wide range of IWIO applications. Three major powers of the world, the US, China, and Russia[9] are applying AI in their IWIO strategies and tactics in unique and actively impactful ways.

The United States, though late in waking up to the advantages of AI in the IWIO domain, is applying AI in its overall defensive IWIO strategy through numerous techniques. The US is utilising AI in many governmental and military areas in identifying and countering IWIO threats spread online and over social media. The US is deploying specific AI applications to identify particular texts, themes, images, and videos that are part of foreign governments' and non-state actors' IWIO operations. The objective is to contain threats that have plans to spread misinformation and propaganda, intensify polarisation and division within the population, and may cause discontent with the elected government.

US plans to use AI to scan through large amounts of data to identify misinformation, propaganda, and intentionally divisive content that is intended to sow discord among domestic and friendly overseas groups. Emphasis is also being placed on using AI technology to counter AI-driven deepfake technology that could be misused by adversaries for IWIO operations directed at the US and organisations like NATO. Additionally, AI is being used to protect critical infrastructure from cyber-attacks. Machine Learning programmes are being deployed to sift through large amounts of data for any indication of possible attacks and thereafter generate AI programmes to defend against Cyber-attacks.

CHINA

China incorporates Mao's notion of the People's War, which consists of employing overwhelming Cyber-attacks combined with online disinformation. IWIO has to be the central component of China's military strategy, given that China secretly acknowledges that it cannot match US military spending. There are confirmed reports that China has employed IW simulation training for over a decade, and IW units specialised in psychological warfare are embedded within its security structure. Furthermore, it is important to note that operations in Cyber-space are an important component of China's IW strategy. An interconnected network of Chinese online influencers who reinforce Chinese narratives in IWIO-targeted countries is an example of this cyber force.

China plans to utilise emerging new technologies like AI in a wide variety of sectors and regions, including 'disruption through trade wars, information manipulation in cyberspace, and military integration of disruptive technologies'. To achieve its goals, China created the Strategic Support Force (SSF) in 2015. China is reportedly using AI under its cognitive warfare tactics to attempt to manipulate public opinion in Taiwan regarding its plans for reunification. This is being done in part through AI-powered programmes and bots that target Taiwanese citizens through the spread of misinformation and propaganda on social media. China has also mounted significant international IWIO operations ranging from AI bots generating misleading content on social media to altered videos depicting the treatment of Uyghurs in China.

RUSSIA

Russia is currently lagging behind the United States and China in terms of incorporating AI technology into its digital security infrastructure. The recent Russia-Ukraine war, however, has forced Russian think tanks to aggressively use AI as war tactics. Russia has demonstrated a concerted effort to further develop its already advanced IWIO tactics with the help of AI technologies.[10] Russia's internet-sponsored propaganda manufacturing facilities are now equipped with AI-powered deepfake technology that can create more realistic false narratives by constructing fake images and even video clips.

Russia exploits information ecosystems by interjecting dis/ misinformation (partially attained through cyber-attacks) and fake news stories that a majority of those exposed to, believed true at the time. Russia's perspective of IWIO covers a wide space of technology, where jamming electronic communication and limiting access to the electromagnetic spectrum, cyber-espionage, and Distributed Denial of Services (DDoS) attacks are no different from (and work in tandem with) using trolls and bots to spread dis/misinformation, establishing pro-Russian media outlets, or supporting local sympathisers to propagate favourable messages.

COUNTERMEASURES FOR MITIGATING AI RISKS IN IWIO

Monitoring the Information Environment

Recognising the severity of the threat of AI in the IWIO domain calls for enhancing capabilities to monitor, analyse, characterise, evaluate, predict, and visualise the information environment. This guidance aligns with the Observe and Orient stages of the Observe-Orient-Decide-Act (OODA)[11] loop. The observation here represents the important first step in the early detection of degradation attempts and issuing of warnings about potential disinformation campaigns. Orientation involves understanding the complex interplay between cognitive biases and the information produced by AI language models. Observation and orientation together lay the foundation for informed decision-making and planning effective action against suspected malign players.

One such example of information-sharing initiatives is the European External Action Service's Rapid Alert System (launched in 2019), which highlights the importance of international collaboration in addressing disinformation threats. Host of pacts between participating countries of various organisations like QUAD, BRICS, AUKUS, etc, can formulate guidelines and policies to monitor and share activities related to IWIO attempts from non-state actors.

Issuance of Advanced Warnings

The important task of issuing timely warnings constitutes an important countermeasure against malign information operations. It is observed that preemptive warnings about possible disinformation attacks significantly reduce the risk of people getting influenced by these attacks. The warnings alert audiences about potential attempts to misinform them, which in turn leads to critical evaluation of the information being encountered. When there is a preemptive warning, the cognitive bias of perceived truthfulness is attenuated to scepticism, forcing the audience to think more.

In the process of issuing effective warnings, attention should be given to not only exposing false narratives but also endorsing true ones. There is a need to continuously adapt to the rapidly evolving information environment. This is possible if there is the ability to promptly detect and respond to emerging disinformation campaigns. Leveraging AI and machine learning technologies can assist these efforts by monitoring the information environment, identifying threats, and swiftly issuing relevant warnings.

Partner Information Operations

In the face of pervasive AI-driven threats, developing partnerships, planning and conducting collaborative operations have become more crucial. By sharing partner nations' capabilities, countries can significantly strengthen their capability against disinformation and subversion attempts. This partnered approach recognises the complex and dynamic nature of the information environment. A core objective should be to enhance partner nations' ability to execute successful information operations independently. Equipping these forces with the knowledge, strategies and tools to operate effectively within the information environment will not only counteract disinformation but will also raise global awareness about the nuances of IWIO.

The Concept of Cyber Teammate

The cyber teammate[12] is a concept application of AI technology for cyber defence and countering information attacks. It is software-based, and it provides a unique and otherwise lacking sense of the environment to the cyber warriors. With the availability of ML algorithms, excellent computing speeds and availability of data, the cyber teammate will build its logic of cyberspace and will support the cyber teams by giving them a clear perspective and understanding of the conflict. A cyber teammate will essentially collect data across open sources, translate and synthesise it in short paragraphs, provide a status update and isolate the tactical and technical information relevant for combat. It would also have the capability to identify the effect of data (information) on/of cultural influence through social media by automating part of the translation process and simulating more credible inputs by process of exploration to conclude better on the impact of inflammatory posts. The cyber teammate will be deployed passively; as and when it assesses mischievous massive amounts of open written text, it will be able to detect changes in tone or style and link together differences in the text to further generate a warning or caution.

FORMING LEGAL REGULATIONS

REAIM Summit

Advances in AI and consequent misuse have sparked increasing calls for setting up regulatory oversight. Raising the slogan 'Responsible AI for Safer Tomorrow', 80 countries across the world participated in the first REAIM[13] summit in the year 2023, and 61 endorsed the non-binding agreement, including major nations like the US, UK, China, Japan, Australia etc. The first meeting of the REAIM took place in Hague in 2023. Following the summit, the US launched its 'Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy'. The UN has also called for initiating a legally binding treaty by 2026 to ban the use of Lethal Autonomous Weapon Systems (LAWS).

EU AI Act

Another notable example of these emerging regulations for AI is the proposed EU AI Act[14], a landmark legislative effort poised to shape the future of AI governance in Europe. The Act’s risk-based framework segregates AI systems into four categories: prohibited, high-risk, limited-risk, and minimal-risk. Each category is subject to varying degrees of regulatory scrutiny and compliance requirements. This structured approach reflects a conscientious effort to harmonise AI advancement with public safety and ethical standards.

US Blueprint for an AI Bill of Rights

White House has also come up with 'Blueprint for an AI Bill of Rights', which talks about five principles that should guide the design, use, and deployment of automated systems to protect the American public in the age of artificial intelligence. The 'Blueprint for an AI Bill of Rights'[15] is a guide for a society that protects all people from these threats—and uses technologies in ways that reinforce the US public’s highest values. From principles to practice, the bill is a handbook for anyone seeking to incorporate these protections into policy and practice, including detailed steps toward actualising these principles in the technological design process. These principles help provide guidance whenever automated systems can meaningfully impact the public’s rights, opportunities, or access to critical needs.

Likewise, NITI Aayog in India has also issued regular directions for responsible use of AI by companies in India. Research is also going on to explore the use of AI in countering AI-generated misinformation or data.[16] Some of these concepts are AI Firewalls, Guardrails, Watermarking and content detection.

VULNERABILITY OF INDIA TO AI AND IWIO

India is undoubtedly one of the fastest-growing markets for social media users. However, due to a lack of awareness, laws and mechanisms to check the spread of rumours, fake news and manipulated videos, it is easy to manipulate the Indian population. Pakistani state-run agencies are increasingly using cyberspace for the collection of sensitive information and the spreading of misinformation. At the same time, Chinese intrusion into the lifestyle of Indians is unprecedented. From malicious hardware infused cheap mobiles to data-gathering apps of Chinese origin, the Indian public has at present no safeguard against the Chinese hidden information war. India is the biggest user of smartphones in the world. These mere statistics make Indian citizens prone to misuse of AI by state and non-state actors. With smartphones aggressively advertising the use of AI, the danger is even further growing. To add to the woes is the monopoly of apps and social media giants who are sending sensitive data and even selling it for their gains. Indian citizens, its soldiers and their families are at the greatest risk of this misuse of AI by Chinese and Pakistani agents.

RECOMMENDATIONS FOR INDIAN SECURITY LANDSCAPE

Through programmes such as Satyamav Jayate, Pradhan Mantri Gramin Digital Saksharta Abhiyan, and National Digital Literacy Mission, India has taken preventive measures to counter disinformation and increase digital literacy.[17] One of the best examples of this was during the COVID-19 pandemic, wherein the government established a WhatsApp chatbot and a fact-checking unit under the Press Information Bureau. However, the government needs to revamp the information and broadcasting network to reach the last man in explaining the good and bad of AI. This can be done through schools, panchayats and social media advertisements.

On the IW front, the Indian Army created the position of Director General Information Warfare two years ago to monitor propaganda from China and Pakistan. These initiatives, however, are primarily serving as fact-checks and not addressing the larger objective of dominating the ‘war of narratives’. Technologically superior and coherent efforts are required to win this propaganda warfare. IW sections set up at the unit level must educate troops regularly on identifying and scrutinising malicious AI-enabled information warfare.

Countering the anti-India narratives can be effective only in the short term. The Government of India will need to adopt a coordinated and Whole of Nation approach involving all ministries in need of the hour to combat the IWIO that targets the Indian minds, especially the youth, to generate communal discord and discontent against the government.

The Ministry of Defence must invest aggressively in technologies that can identify and counter the AI in IWIO intended for its forces. In this regard, it must convince the Government of India to force social media giants like Metaverse and Twitter to set fact-checking and filter AI-polluted information aimed towards India. Lastly, Government of India must formulate tough laws for foreign firms found involved in fabricating disinformation, deep-fakes and exploiting online social space to incite people to destabilise the chosen government or influence leadership. As covered earlier, India should become part of larger world forces making efforts to counter the malicious use of AI.

CONCLUSION

The rapid advancement of AI, its intrusion and its expanding role in modern societies are unstoppable and undeniable. The article was a detailed exploration of AI's complex nature, revealing the risks it pose in the realm of information warfare. This risk is actually real and not hypothetical; it presents a tangible threat in our current digital world, where AI technologies such as Bing Chat and ChatGPT are all the time more intertwined with the internet and easily accessible gigantic pools of open data.

The accessible nature of AI also means that current vulnerabilities in AI systems could be exploited by a wide group of malevolent forces, ranging from radical terrorist networks to antagonistic state actors. Such exploitation could potentially disrupt societal structures and skew public perception, a risk that is amplified considering AI's increasing integration across multiple sectors, including government and private enterprises. Regulatory frameworks like the EU AI Act represent positive steps towards safer management of AI. Yet, they still lack a holistic approach to covering the full spectrum of risks associated with AI, especially against sophisticated adversarial AI tactics. The risk associated with the use of AI in IWIO necessitates a regulatory method that not only channels and controls AI development but also facilitates its harmonious integration and evolution in the face of external threats, while maintaining balance with our complex socio-technological landscapes and human right.

★★★

Gp Capt R K Dogra is serving Aeronautical Engineering officer of IAF. He specialises in aircraft technology, especially the MRO functions. He has served as Chief Engineering Officer of an important base in EAC. He is currently working with LRDE, DRDO and writes on technology and defence related subjects.

NOTES

[1] *Pascal Brangetto, Matthijs A. Veenendaal: Influence Cyber Operations: The Use of Cyberattacks in Support of Influence Operations, 2016 8th International Conference on Cyber Conflict Cyber Power.*

[2] *Lance Y. Hunter, Craig D. Albert, Josh Rutland, Kristen Topping & Christopher Hennigan (05 Mar 2024): Artificial intelligence and information warfare in major power states: how the US, China, and Russia are using artificial intelligence in their information warfare and influence operations, Defense & Security Analysis, DOI: 10.1080/14751798.2024.2321736*

[3] *Ibid*

[4] *Hutchinson, W., and M. Warren. "Principles of Information Warfare." Journal of Information Warfare 1, no. 1 (2001): 1–6. https://www.jstor.org/stable/26485918.*

[5] *Lance Y. Hunter (p-6)*

[6] *Dusan Bozalka: Information warfare in the age of Artificial Intelligence*

[7] *Russel Hanson, Adam R Grissom, Christopher A Mouton: The Future of Indo-Pacific Information Warfare- Challenges and Prospects for the rise of AI (RAND Corporation, 2024)*

[8] *Ibid*

[9] *Lance Y. Hunter (p-14)*

[10] *Lance Y. Hunter (p-22)*

[11] *Russel Hanson (p-5)*

[12] *Guyonneau, Rudy, and Arnaud Le Dez. "Artificial Intelligence in Digital Warfare: Introducing the Concept of the Cyberteammate." The Cyber Defense Review 4, no. 2 (2019): 103–16. https://www.jstor.org/stable/26843895.*

[13] *Prasad, Ashika S . "AI in Warfare: The REAIM Summit and India's Approach – CENJOWS." Cenjows.in, October 15, 2024. https://cenjows.in/ai-in-warfare-the-reaim-summit-and-indias-approach/.*

[14] *Lax, Edwin. "AI Pollution: The Future Threats of Information Warfare." Trendsresearch.org, 2024. https://trendsresearch.org/insight/ai-pollution-the-future-threats-of-information-warfare/?srsltid=AfmBOooaFzjvhv17FTGqrYOqlgXsXusTJ29FJq5vbYysN2VySIfA31OY*

[15] *Whitehouse Website (www.whitehouse.gov), The Blueprint for an AI Bill of Rights*

[16] *Rahul Kapoor, Theresa T Kalathil: AI Regulation In India: Current State and Future Perspectives, Morgon Lewis (Jan 2024)*

[17] *Young Voices. "India's Two-Front Information War." orfonline.org, May 10, 2023. https://www.orfonline.org/expert-speak/indias-two-front-information-war.*