ARTIFICIAL INTELLIGENCE AND INFORMATION WARFARE: A DANGEROUS WEDLOCK

Col Gaurav Soni & Mr Dhruv Swarnakar

Abstract

This paper aims to understand the newfound and potentially menacing relation between Artificial Intelligence (AI) and Information Warfare (IW). It previews as to how AI could fundamentally change IW making it highly incisive and accurate. While control of information and spread of disinformation shall continue to remain the central theme of IW, in the times of tomorrow, artificially generated disinformation campaigns will exponentially enhance the speed, intensity and most importantly the accuracy of such operations. It reads into the known global instances of AI powered IW to include a case study of the recently concluded Taiwan Elections. It not only covers as to how AI is creating new content at unprecedented speeds but also how it sparks a viral chain of distribution reaching billions of users at an incredible pace. The paper covers as to how AI takes course corrective measures towards the information content by continuous feedback and refinement making the information campaign an assured success. A unique aspect discussed in the paper is how AI is able to weave the data into making the human mind believe the information by factoring-in a touch of authenticity. Aspects like Truthiness, Cognitive Fluency Bias, Hashtag Creations, AI Engagement, AI Simulation, Generative Adversarial Network (GAN), Social Media Analytics, Sockpuppet accounts, Click Farms etc have been elaborated in the paper. The paper throws light on global instances of AI powered IW campaign levying special focus on China's

demonstration of offensive and defensive IW campaigns. Towards its conclusion, it aims to recommend suggestions at the concept level especially applicable to democratic countries likes India in order to battle the rising 'Monster Duo' of AI and IW.

INTRODUCTION

On March 16, 2022, through a video message, Ukraine's defiant President Vladimir Zelensky asked his forces to lay down their arms in the ongoing Russia-Ukraine conflict. It took no time for the video to be circulated to millions of users across the world, and soon enough, global strategists were set to wonder whether the conflict was coming to an end. It took more than 24 hours of restorative actions at the hands of Ukraine to convince its own army and the world that the video was 'deepfake' and was artificially crafted and inseminated in social media. So much so was the impact that some of the news channels had begun running the story that Zelensky had fled Kyiv. Even prominent new channels like 'Ukraine 24' found their website's homepage defaced with hackers having inserted the fake video in the opening webpage. Immediate rebuttal came in from President Zelensky himself who negated release of any such video through Facebook. While Russian social media such as Vkontake, continued to toss the video repeatedly as far and wide as possible, majority of the speedy restoration actions were undertaken by western social media giants like Youtube, Twitter and Meta which reacted favourably by quickly removing the video. Meta's security head Nathaniel Gleicher, Twitter representative Trenton Kennedy and YouTube's Ivy Choi did not lose any time in calling the video 'synthetic' and in 'violation of their policies'.[1] While a supportive Social Media, as in the instant case, could arrest the alleged 'deepfake' video in time, an adversarial information campaign duly powered by Artificial Intelligence(AI) could have changed the outcome of the conflict. It may have to believed that the brewing concoction of AI with far reaching and powerful social media is heralding a change in the design and conduct of modern wars and this, if tapped, has a potential to bring about incalculable damage to the adversary.

This paper aims to understand the newfound and menacing relationship between AI and Information Warfare (IW). It previews how AI could fundamentally change IW, making it highly incisive and accurate. It reads into the global instances of AI-powered IW and briefly covers China's progress in the domain. Towards its conclusion, it aims to recommend enactments at the concept level, especially applicable to democratic countries like India, to battle the rising 'Monster Duo' of AI and IW.

IS DISINFORMATION ENOUGH FOR INFORMATION WARFARE?

The Media-(Dis)information-Security document Defence-Education-Enhancement-Programme (DEEP) of NATO identifies gaining of information advantage over adversary as the key objective of information operations. In doing so, it outlines that, own information-space needs to be protected while destroying and disrupting adversary's information and its flow.[2] Use of information as a tool to warfare is not a novel concept and adequate instances of its application can be drawn out even from the earliest battles and epics known to mankind. Since then, the concept of IW has undergone numerous changes in the manner in which own information-space is protected and that of the adversary denied. Considerable evolution did take place post the World Wars, in a manner that its application could now be through newly evolved domains such as Electromagnetic, Economic, and Cyberspace. However, the generation of content, its speed of generation and application, trial and testing, validation, intensity of effort and accuracy of IW targeting remained limited in scope and execution. Fast forwarding today, and specially so in the last decade, the character of information operations is once again undergoing a metamorphosis. Though the domains of application of IW may remain unchanged but the manner in which information will be synthesised and applied is set to undergo an overhaul with the manifesting of Large Language Models (LLM). One such example is the Open AI's Chat GPT. While control of information and spread of disinformation shall continue to remain the central theme of disinformation, in the times of tomorrow, artificially generated disinformation campaigns will exponentially enhance the speed, intensity and most importantly the accuracy of such operations.[3] A case study of operations of a Chinese

Origin Information Technology (IT) Cell called 'Storm 1376' also known as 'Spamouflage' or 'Dragonbridge', in the Taiwanese General Elections elaborates on how AI can execute a coordinated, expeditious and highly accurate IW campaign.

AI-POWERED IW CAMPAIGN ON TAIWAN ELECTIONS: A CASE STUDY

In January 2024, Taiwan conducted its 16th presidential elections. The election remained narrowly contested between the Democratic Progressive Party (DPP), headed by its candidate William Lai, and the Kuomintang (KMT) party led by Yu-ih, with the third party being the Taiwan People's Party (TPP), run by Ko Wen-je. While DPP and TPP have a democratic approach, KMT is believed to be enjoying China's backing. A report released by Microsoft Threat Intelligence published on Microsoft's Official website brought out that beginning in November 2023, China-based Group 'Storm 1376' artificially generated thousands of memes denigrating images of DPP and TPP in the forthcoming elections.[4] In December 2023, a deepfake video of a woman claiming to be the mistress of DPP candidate William Lai was pumped into social media and made viral by AI tools through Search Engine Optimisation (SEO). Also, emerged in December the 'Spring Breeze Files', which alleged that William Lai was an informant acting against Taiwan.[5] These files were artificially amplified on social media like Twitter, Facebook and Japan's social media app 'Line'. Sensational hashtags were artificially generated using crawling software, which could feel the pulse, diction and vocabulary of the voter. These artificially generated hashtags exponentially increased the speed and spread of the viral files. In January itself, a series of videos titled 'Secret History of Tsai-Ing Wen' wherein artificial anchors virtually read out a 300-page document alleging a number of frauds on the then president Tsai-Ing Wen to include falsified academic credentials, finances and personal life.[6] These videos were artificially generated using 'Capcut opensource AI software' designed by China-owned firm ByteDance, which is also the parent company of Tik-Tok.[7] One of the most remarkable AI-generated IW campaigns emerged right on the day of the election when an AI audio recording depicting

Foxconn's head, Terry Gou, was seen backing the KMT candidate Hou Yu-ih. It is imperative to understand that Foxconn is the most revered private company in Taiwan and is responsible for its semiconductor prowess.

The campaign, as above, is not only related to the 'Creation of Disinformation' but has far more to do with its distribution, marketing, feedback-based improvement and precise and timely delivery to the targeted audience. Some of the advanced processes which were seen to be manifesting in the campaign are discussed as follows:

- **Creation of Content - Deepfake Videos**. An AI model or software is created based on complex algorithms called Variational Autoencoders (VAEs) or Generative Adversarial Network (GAN). Once the software or the model is ready, a very large set of images, videos and voiceovers of the target individual is then imported onto the model.[8] This large dataset is used to train the model to the desired degree of accuracy. Desire propaganda scripts is then inserted in a manner as if the target individual himself is delivering the script. Chinese software like Capcut can facilitate its editing in a smooth manner by providing unique features like chroma key (green screen), keyframe animation and motion tracking, which refine an AI-generated deepfake video. Keyframe Animation, for instance, gives a much finer control over the target object's movement and allows for smooth transitions, while Motion Tracking helps synchronisation of the movements.

- **Creation of Distribution Media - Virtual Accounts Using Virtual Phone Numbers.** Almost all social media platforms now necessitate linking of the account with a personal mobile number. An AI-based web engine can create bots which are 24 x 7 active on the internet, creating virtual phone numbers. Alternatively, apps like 'Call Hippo' can be used to generate virtual phone numbers. These phone numbers can then be used to create virtual email accounts and, consequently, social media accounts. Once such an account is created, bots pretending to be human

entities can be designed to be part of large-scale fake Whatsapp groups, Telegram Groups or any closed group on the internet. Activities of a normal, legitimate group, like Casual Chats, Display Profiles, etc, are artificially undertaken by these member bots. In a process called 'AI Engagement' or 'AI Simulation', the bots can even be trained to like, comment or share chats, thus earning the trust of other human members. At the opportune time thereafter, the bots pump deepfake videos and falsified information on these groups as part of coordinated IW campaign.[9]

- **Social Media Analytics - Understanding Public Sentiment**. Firstly, a large data set is extracted with the help of member bots from groups on social media. This is integrated with various data analytics AI software like Microsoft Power BI, IBM Watson, Sprout Social, Tableau AI, etc. Data which is analysed includes comments, reviews on products, likes, dislikes, watch times, repeat watching frequency, sharing patterns, etc. Even emojis and the undertone in the language of comments are analysed. Bots can even incite discussions and artificially generate pointed questions to know and better understand the minds of human participants.[10] Such analysis gives out some key outputs like the potential favourable or non-favourable target audience, key issues of contention and issues which can be sensational in nature. Even individual target personnel can be identified.

- **Hashtag Creation - Setting a Chain Reaction**. A hashtag first indexes a post and then incorporates it as part of a much larger audience. Thus, it is not merely a private group agenda but allows the subject to attract participation from an audience spread across the globe. AI can create sensational hashtags by first undertaking social media analysis and understanding of common public sensations and then undertaking Natural Language Processing (NLP) to create the most apt hashtags. Some of the most commonly available AI hashtag generating software like 'Ahref' and 'Hootsuite' can craft hashtags that can propel a subject exponentially and thus hog the attention space.[11] Once artificially

pushed to the extent of being 'Trending', it thrusts itself into the mind space of a considerably large section of society. Such can be the negative impact of hashtags that, on one occasion, the #Pizzagate, resulted in the spreading of false rumours of child trafficking, public mobilisation, and even firing incidents in the US.[12]

- **Optimisation, Refinement and Testing**. AI tools undertake an iterative process to closely monitor their actions during all of the processes as above to continuously identify problem areas, undertake remedial actions and thus artificially correct and refine the processes. Unlike human processing, which entails considerable processing and decision delays, AI-based refinement procedures enhance targeting accuracy multi-folds in near real-time.[13]

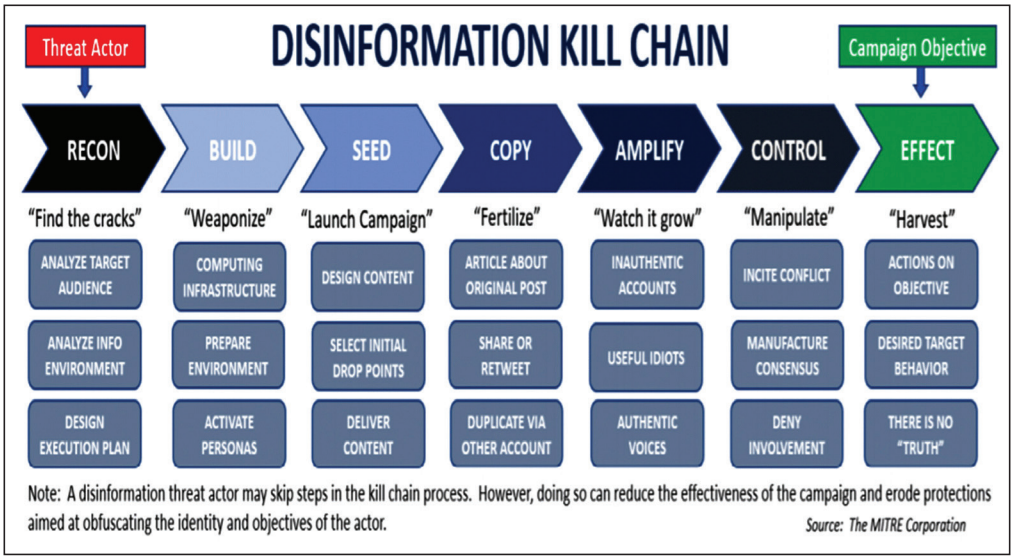

Figure 1: Disinformation Kill Chain Adopted in IW. **Source:** The Mitre Corporation

Unlike earlier elections where conventional IW was a standard tool for meddling with election outcomes, the 16th Taiwanese elections saw a never before well-coordinated AI-powered IW campaign at the hands of adversaries. In this campaign, the focus shifted from the mere creation

of disinformation to the processes through which this disinformation was executed, making it highly precise and effective. Such a shift in the IW is best understood in the language of business management. In business, the Product Concept applies to the generation of quality and cost-effective products. On the contrary, the Marketing Concept targets the market and aims to integrate profitability with manufacturing by suitably modifying production, processes, delivery and communication.[14] Before the onset of AI, information campaigns were largely focused on the product, i.e. Disinformation. However, with the coming of AI, IW does not just focus on Disinformation but its quick generation, mass distribution, targeted approach and feedback-based iterative modifications applicable both to the product, i.e. disinformation and the process in a highly automated, sharp, incisive and ruthless manner. The advantages of incorporating AI in IW are not limited only to the creation of the product i.e. Disinformation and processes enhancements. In fact, the most remarkable advantage is its ability to package the Disinformation in a manner most trustable by human minds. This is discussed in subsequent paras.

Tinkering the 'How' of What We Believe. AI-generated IW manipulates two of the many fundamental qualities of human thought. First is Cognitive Fluency Bias wherein humans tend to believe what is easy to process or 'be understood' by them[15] and second is Factor of Truthiness, which tends to accord authenticity to a well-presented, convincing and gripping visual, text or speech. Given the vast data set available to AI models and its ability to modify them and selectively embed them with evidence, AI can produce information that is highly polished and well presented in a considerably short time. Given the constraints of time in a strained scenario, human-made presentations can appear to be 'messy', 'disorganised' and even 'delayed when pitched against outputs of AI. Since AI presentations would be easy to process, Cognitive Fluency Bias causes human minds to believe them more vis-à-vis human-made presentations. Further, the 'Factor of Truthiness' invokes authenticity in AI presentations, which is not likely to undergo scrutiny by human 'gut feeling' as it is well crafted and seemingly duly evidenced.[16] Thus, AI-generated content can not only be fake but also be so engineered

that it even aims to tinker with the human subconscious, which is the key element in crafting our initial beliefs and proclivities. Once such beliefs are instated, they are reinforced using 'filter bubbles', which, in essence, are a set of contents shown over and over to a user based on his initial choices and inclinations. Similar to the concept of YouTube recommendations, which display videos similar to users' earlier choices, such filter bubbles over a period of time, build a tunnelled vision and a belief system that is resilient and defiant even if conflicting arguments are logical and duly supported by evidence.[17] Action based on beliefs and not based on objective facts. Once the belief system is concretised, humans tend to act according to belief and not based on objective facts or even visual reality. Defined very popularly by the term Post Truth and adopted by the Oxford Dictionary in 2016 as its word-of-the-year, in this phenomenon, individuals prioritise their firmly instilled beliefs even if they face contrarian real evidence. Decisions and Actions that follow are drawn from these beliefs.[18]

RECENT USAGES OF AI IN IW: GLOBAL SCAN

Some of the recent cases of aggressive uses of AI-integrated IW undertaken by nations across the world are briefly discussed in subsequent paras.

- Moldova, which is a neighbouring state to Ukraine, has been adopting a pro-west stance in the Russia-Ukraine conflict. In a deepfake video, Maia Sandu, the west-friendly President of Moldova, was seen backing a Russia-friendly political party and was shown to propose her resignation.[19]

- In June 2024, Tech firm Open AI alleged that Israel's AI firm STOIC may have launched an AI campaign in India in an attempt to meddle with Indian elections. It was alleged that the AI campaign called Zero Zeno praised the Congress Party and undertook an anti-BJP stance. One of the typical modus operandi of STOIC is to create fictional persons and generate artificial biographies to garner voters' attention.[20]

- AI-powered 'Shallowfake' videos can artificially slow down or increase the speed of video. In May 2019, US House of Representatives Speaker Nancy Pelosi's video was 'Shallowfaked', making her speech appear drowsy and drunk. Such videos can significantly dent the reputation of the target individual.[21]

- Venezuela's government used the services of the private company 'Synthesia' to generate pro-government AI-generated news through news channels which never existed and were artificially created.[22]

- Political parties explore new ways to build emotional connections as part of their voter outreach. In Indonesia, an AI app launched by a presidential candidate enabled users to artificially create a joint selfie with the leader, thus building a personal connection.

- Another form of large-scale AI-powered IW is the generation of 'Robocalls' faking the voice of popular personalities. In the run-up to primary elections in the US New Hampshire, people received a large number of phone calls from robots impersonating US President Joe Biden.[23]

CHINA'S CONCEPTUAL PURSUITS AND GIANT LEAPS IN AI-ENABLED IW: A BRIEF OVERVIEW

Control of Information Space has remained an approach at the heart of China's Strategic pursuits. Chinese Defence White Paper of 2004 modified China's erstwhile approach of 'local wars under modern, high-tech conditions' to 'local wars under informationised conditions', which was further modified to 'Winning Local Wars under conditions of informationisation'.[24] By 2015, China had further tweaked its military strategy to 'winning informationised local wars'. This was also the time when the PLA Strategic Support Force (PLASSF) had begun to take shape, and concepts of 'Integrated Network-Electronic Warfare (INEW)' were introduced.

As LLM models such as Chat GPT have enabled generative AI solutions offering automated content creation and processing, China has made constant efforts to maintain an incisive edge in the domain. At the grassroots level, its Academy of Military Sciences, PLA Academy of Electronic Technologies, Military Strategy Research Centre and the Academy of Xian for Politics have undertaken research studies and have developed themselves into pockets of excellence. China's Military Strategy of 'Intelligentised Warfare' is a four-pronged approach achieved through information-processing, quick decision-making, cognitive warfare and use of swarms. To execute these, China has identified that AI must form the core of Intelligentised Warfare.[25]

At the concept level itself, China has not shied away from accepting AI as a driving tool to Modern Warfare. In 2017, PLA's 'New Generation AI Development Plan' categorised AI as a strategic Initiative and identified the need for a Whole-of-Nation approach in the domain. In 2019, President Xi Jinping, while addressing the Collective Study Session of CCP's Politburo, identified the need to guide public opinion by using AI in IW domains of news collection, its production, distribution, and timely feedback and suggested creation of AI Editorial Departments.[26] In May 2021, he further identified the need to create an opinion of the external public favourable to China. During the 'Internet Civilisation Conference' conducted in 2022, Ye Zhenzhen, a CCP secretary, expressed that big data and AI help understand citizens better and can contribute towards China's leadership. He referred to the development of projects focussed on cognitive computing in order to guide public opinion and even public values and called them 'national weapons in the digital era'.

CHINA'S OFFENSIVE APPLICATION OF AI

- In August 2023, a forest fire engulfed the areas of Maui in Hawaii, USA, after which Storm 1376 flooded the internet with AI-generated files, memes and articles blaming the US government for testing a 'Weather Bomb' leading to the fire incident. Interesting to note is that the flames were artificially shown to be engulfing cities and suburban towns to cause outrage in the local populace with an aim to create an outrage in the public.[27]

- Soon enough, in November, after a train derailment in Kentucky, Storm 1376 artificially drew similarities of the incident with Pearl Harbour or the 9/11 Terrorist Attack and called it a US government vendetta.[28]

- The Microsoft Threat Intelligence Report released in 2024 highlighted that a large number of 'Sockpuppet' accounts operating on social media are linked to the Chinese Communist Party. Sockpuppet accounts are virtual accounts wherein bots may pretend to act as citizens of the country and participate in actual political discussions, thus swaying common public thoughts.

- In August 2023, Chinese Storm 1376 undertook a powered IW campaign pointed towards nuclear water discharge into the Pacific Ocean by Japan and attributed this act to US-Japan making attempts towards Water Hegemony. Realistic-looking Twitter accounts, captivating AI memes and visuals were used to garner public attention.[29]

- In the run-up to American elections, several deepfake videos emerged depicting US President Joe Biden giving out 'transphobic' viewpoints in an attempt to undermine his popularity in the transgender community.

- Meta, the parent company of Instagram and Facebook, revealed that over 8000 fake Chinese accounts were deleted, which may have been originally based either in China or in China's 'Click Farms' in Brazil and Vietnam.[30] Earlier, a Click Farm typically employed a large number of persons to click on online content, feigning large traffic flow and thus enhancing popularity statistics. With AI now in the offing, Click Farms are now easily being 'cultured' and upscaled by employing artificial bots.

- China does not fall short of running an IW campaign, even at individual levels. Chinese immigrants who are now US citizens are targeted by AI bots sending out countless messages and flooding their social media accounts. Jiyayang Fan, a Chinese-US citizen,

received a barrage of AI-generated demeaning messages calling her a traitor and homophobic.[31] Even an AI-generated hashtag, #TraitorJiayangFan, was first made to trend with the help of AI tools and then retweeted between 12,000 users.

CHINA'S DEFENSIVE APPLICATION OF AI

China's internet firewall and CCP's absolute control of the in-house narrative are well known. Despite these guards in place, China has maintained an active defensive AI-powered IW campaign under:

- **Search Engine Optimisation (SEO).** China's surveillance bots maintain a constant vigil over search queries. When searched for sensitive topics like 'Uyghur' or 'Xinjiang', Chinese state media showing positive news rank best in SEO results thus returning as the top pages on Google or Bing.
- **Astroturfing Marketing.** Astroturfing Marketing is a business term wherein inauthentic messages are subtly pushed to make them appear authentic. China undertakes large-scale coordinated campaigns posting AI-generated posts supporting CCP's policy in the form of fake people's interviews depicting wide grassroots acceptance.[32]
- **Drowning Conversations.** A large quantum of artificially generated messages and articles are created to drown out anti-China conversations on the internet.
- **Multi-Channel Networks (MCN) of Social Influencers.** While YouTube is banned in China, some Youtubers are permitted and contracted out using a MCN agreement with the CCP. These are popular influencers, often women, who bring the best picture forward and are allowed to be monetised on YouTube. Comments generated on their videos are artificially reposted for increased visibility through AI models. On the other hand, non-MCN channels are not allowed to be monetised, thus drawing away the inspiration to create even neutral content.[33]

- The Chinese government has created a large number of surveillance chatbots that trap an online query or discussion on Tienmann Square and prevent the display of any information on the subject.
- China has also outsourced UK-based firm Synthesia, which has artificially created a Western media news channel 'Wolf News' singing praises for CCP. This was primarily targeting the in-house Chinese population towards strengthening their belief in the CCP.

The application of AI by China for powering the new generation of IW needs a detailed study and can only be covered as a brief overview in this paper. It should, however, be understood that, having identified the burgeoning role of AI in future IW, China, incited by its desire to control the global narrative, will continue to hone its AI prowess.

GLOBAL EFFORTS TO COUNTER AI-POWERED IW

Given its propensity and accuracy, IW powered by AI can considerably undermine governments, regulatory mechanisms and belief systems of a large mass of people. The threat posed by AI-powered IW has drawn the attention of governments across the globe.

- The European Union (EU) is amongst the forerunners in countering IW powered by AI. In 2019, the EU's Rapid Alert System was launched, which was aimed to enable common situational awareness by sharing information and jointly mitigating disinformation between stakeholder countries. EU has legislated that media platforms are required to give assurance on curbing Disinformation on Election proactively. Very soon, it is likely that legislation will require social media platforms to identify deepfake content and 'label' them prior to circulation. EU's AI model Project InVID is an intelligent Web Crawler which undertakes video fragmentation, annotation detection, video forensic investigation, contextual analysis and web intelligence analysis for filtering out authentic content to be provided to news media. This content can then be handed over to responsible media houses and journalists

for publishing on their news platforms. It also introduces a User Generated Content (UGC) Verification App, which can identify and authenticate a valid user who can then post authentic data only on the web.[34]

- Social media firms have also evolved their own mechanisms to thwart IW campaigns. Google has developed a Convolutional Neural Network (CNN) called EfficientNets, which can analyse fake images on the web. Microsoft has developed 'Video Authenticator' to give a percentage basis veracity to a video. Microsoft has also initiated Project Origin partnering with Radio Canada, BBC and The New York Times, which aims to verify online content by attaching digital certificates to files which can be plugged in with the user's browser extension. Wall Street Journal and Associated Press have also come together towards Trusted News Initiative with a plan to create similar labelling or certification of online content. This plug-in, called Newsguard, can be installed with various online browsers.
- China has been moving towards criminalisation of deepfakes as the Cyberspace Administration of China has called it a threat to China's security and social stability.
- Necessary changes in the legislation of democratic governments are a much-desired reform to battle IW. Singapore, in 2021, passed a Foreign Interference Countermeasure Act, which allows its bureaucrats to take suo motto cognisance of Foreign Interference and thus undertake suitable countermeasures.
- The US is unequivocal about its prioritisation of AI. The 2023 National Defence Authorisation Act (NDAA) brought in a $20 billion hike in spending on AI. Further, the US Third Offset Strategy (TOS), which is a highly tech-oriented innovative programme, has identified AI to be the number one on the list of priorities. In 2018, the US announced the formation of the Jt AI Centre (JAIC) to progress AI as a tool for future warfare. In 2019, the US Department of Defence proposed its 'Defend Forward' Strategy, which aimed to disrupt malicious IW away from the US mainland.

DoD's DARPA has also undertaken the Media Forensics (MediFor) for large scale threat detection and Semantic Forensics (SemaFor) large-scale characterisation of flagged content in order to reach its originator.[35] In 2020, JAIC adopted IW as one of the key objectives of its AI campaign. By 2022, JAIC was merged with the Chief Digital & AI Office (CDAO), which is not only limited to the military but adopts a whole-of-government approach for data analytics and AI strategy. One of the most remarkable steps has been the raising of the Global Engagement Center (GEC), which is a US Government body operating under the Bureau of Public Affairs, with an objective to counter foreign disinformation campaigns by undertaking inter-agency coordination. One of the main objectives of the GEC is also to collaborate globally with like-minded nations towards such initiatives. In its Special Report on China's IW campaign released in 2023, the GEC identified that China's Information Manipulation is a 'challenge to the integrity of global information space'.[36]

KEY CHALLENGES: BATTLING AI POWERED IW

A report published by Davos has identified disinformation facilitated by AI as the most grave short-term threat to mankind. In the ever-evolving technological scenario, battling AI-based IW can be immensely challenging. Some of the key characteristics that give AI-powered IW a definite edge in the 'cat and mouse' game are discussed in subsequent paras.

- **Accessibility and Affordability.** Modern Open Source AI tools, despite being new in the market, are remarkably cheap and accessible, thus lowering the entry barrier for most users.[37] Softwares like DeepFaceLab allow the creation of Deepfake videos at almost no cost. This allows for large-scale participation and a humongous generation of data available to the target audience. Even if a counter-campaign is run, it is likely to be overwhelmed by the sheer propensity of data.

- **Policy Structure - Very Large Online Platforms (VLOP).** Some of the biggest VLOPs, such as Google, Facebook, and Twitter,

came together to sign an understanding wherein they assured the prevention of the usage of AI tools from interfering in elections. However, some platforms like Telegram, having end-to-end encryption, have refrained from such understandings and thus are not bound to check or control AI-generated deepfake data. It is, therefore, that Telegram remained one of the most extensively used services for the flow of AI-powered IW during the Russia-Ukraine conflict.[38]

- **Undermining Trust in Truth.** Excessive flooding of information of a variety puts the information consumer in a situation of dilemma as to what to believe and what not. This allows for a window of opportunity even to a mischief-maker as he can brand any information on the internet as AI-generated, thus whisking it away as mere propaganda. Such a window, popularly termed as Liar's Dividend, allows even a real culprit to capitalise on the dilemma so created and deny even real evidence.[39]

- **The Streisand Effect.** In most cases, the first response of the government is to ban or remove content from social media to prevent its proliferation. However, psychologists believe that it gives rise to what is known as the Streisand Effect which, on the contrary, intrigues the common public and further motivates them to consume the content.[40] An official rebuttal by an authentic source may provide a better alternative.

- **Freedom of Press and Freedom of Expression.** Especially applicable in democratic countries like the US and India, AI-powered IW finds adequate windows of operation under the protection of freedom of the press or that of expression.

SUGGESTED COUNTERMEASURES BY DEMOCRACIES: AI-POWERED IW

Very Large Online Platforms (VLOPs) are the prime movers of the mass population. There is a need to draw operating guidelines and regulations necessitating these platforms to inform the government and public at large on foreign IW campaigns. Adequate transparency norms have to be instituted. Detection, deepfake labelling and responsibility for verification of authentic information before the public's consumption has to be mandated upon the VLOPs.

There is a need for a clinical approach in the identification of foreign IW campaigns. Foreign agencies instigating such vendettas have to be brought under regulation or duly censored before the public's consumption. There is also a need to develop advanced AI software and tools that can detect and counter malicious IW campaigns. Web browser plugins or government-developed apps hosted on official websites should be developed to detect such discrepancies.

The concept of Cognitive Innoculation entails that suitable pre-emptive warning be given to the public en masse about the possible intent and approach of the adversary's IW campaign. This can remarkably reduce the extent of impairment likely to be caused by a malicious IW campaign.[41] Such warnings will have to be predictive in nature and will, therefore, have to use AI-based apps or software to detect anomalies and subsequently inform the public.

A quick rebuttal in the form of a correct narrative has to be built up. However, given the speed and intensity of AI-powered offensive IW, such rebuttals will have to be artificially generated and processed for wider consumption. Stronger regulations carefully crafted to be able to delineate freedom of expression from intentful anti-national IW

campaigns are the most important steps towards preserving security while also retaining the democratic character.

A battle in the information space needs an integrated approach. Such integration is at two levels. Firstly is the global collaboration of countries and agencies, which need to combine vigilance with response for an accurate and quick response in arresting the disinformation campaign. The second is integration at the national level. Specialised and empowered mission-based agencies adopting a whole of government approach towards data integration and analytics, detection and mitigation of IW campaigns will have to be raised to address disinformation drives.

CONCLUSION

The modern world security calculus is set to undergo an unprecedented tumult with the coming of new-age IW campaigns powered by AI. With its astute abilities to remain in control of information battlespace, AI can lead to large-scale population mobilisation, thus leading to a state of disharmony and perpetual distrust even amongst unsuspecting players. In particular, are the vulnerable democratic nations where the will of the people sets the national agendas. With nations like China, which not only limit IW to the military but apply a whole-of-society approach in times irrespective of conflict and peace, it is this will of the people and their natural belief system which shall be at the centre of the AI-powered IW campaign of the future. Battling IW powered by the ever-evolving and maturing AI will, therefore, have to be at the core of all efforts of responsible governments.

★★★

Col Gaurav Soni is presently posted as Directing Staff in the Junior Command Wing, Army War College. The officer is from Artillery, has served in Srilanka and has commanded his regiment along the Indian borders. The Officer is a PhD in Defence and Strategic Studies and is a PG Diploma in AÍ and Machine Learning.

Mr Dhruv Swarnakar is pursuing research in Mechatronics at Manipal Institute of Technology, Udupi. His areas of interests include Mechatronics and Robotics. With keen interests in Weapon Technology and AI, he actively follows defence upgrades across the globe. He is a research member for projects undertaking designing of Bionic Arm for handicapped persons. He has also made significant contributions towards research for Disaster Management and Weather Forecasting using Machine Learning Algorithms.

NOTES

[1] *Simonite, Tom. 2022. "A Zelensky Deepfake Was Quickly Defeated. The Next One Might Not Be WIRED." March 17, 2022. https://www.wired.com/story/zelensky-deepfake-facebook-twitter-playbook/.*

[2] *DEEP. n.d. "What Is Information Warfare?" Accessed September 12, 2024. www.nato.int/nato_static_fl2014 /assets/pdf/2020/5/pdf/2005-deepportal4-information-warfare.pdf.*

[3] *Hunter, Lance , Craig Albert, Josh Rutland, Kristen Topping, and Christopher Hennigan.. 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 40 (2): 235–69. https://doi.org/10.1080/14751798.2024.2321736.*

[4] *"China Tests US Fault-Lines and Ramps AI Content to Boost Its Geopolitical Interests - Microsoft On the Issues." n.d. Accessed September 18, 2024. https://blogs.microsoft.com/on-the-issues/2024/04/04/china-ai-influence-elections-mtac-cybersecurity/.*

[5] *"China Tests US-Fault-Lines and Ramps AI Content to Boost Its Geopolitical Interests - Microsoft On the Issues." n.d. Accessed on September 18, 2024. https://blogs.microsoft.com/on-the-issues/2024/04/04/china-ai-influence-elections-mtac-cybersecurity/.*

[6] *Tsai Yung-yao, Jonathan Chin. 2024. "China Is Posting Fake Videos of President: Sources - Taipei Times." Accessed on September 13, 2024. www.Taipeitimes.Com/News/Front/Archives/2024/01/11/2003811930.*

[7] *Chi Hui Lin. 2024. "How China Is Using AI-News Anchors to Deliver Its Propaganda | Artificial Intelligence (AI) | The Guardian." Https://Www.Theguardian.Com/Technology/Article/2024/May/18/How-China-Is-Using-Ai-News-Anchors-to-Deliver-Its-Propaganda. 2024. Accessed on 14 September, 2024. www.theguardian.com/ technology/ article/2024/may/19*

[8] *Zendran, Michał, and A Rusiecki. 2021. "Science Direct-NC-ND-License https://Creativecommons.Org/Licenses/by-Nc-Nd Peer-Review under Responsibility of the Scientific-Committee of KES International. Swapping Face-Images with Generative Neural Networks for deepfake Technology-Experimental Study." Accessed on 13 September, 2024. https://doi.org/10.1016/j.procs.2021.08.086.*

[9] *Rajvardhan Oak. 2024. "Friend-or-Faux : How Bots Pose Challenges in Social-Media Spaces." Accessed on 09 September, 2024. www.Timesofindia.Indiatimes.Com/Blogs/Cyber-Chronicles/Friend-or-Faux-How-Bots-Pose-Challenges-in-Social-Media-Spaces/. January 17, 2024.*

[10] *Ibid*

[11] *"70 Best Social Media Hashtag AI Tools - 2024." n.d. Accessed September 18, 2024. https://topai.tools/s/social-media-hashtag-.*

[12] *Mark Fisher. 2016. "Pizzagate: From Rumor-to-Hashtag to Gunfire in D.C." December 6, 2016. Accessed on 02 September, 2024. www.washingtonpost.com/local/pizzagate-from-rumor-to-hashtag-to-gunfire-in-dc/2016/12/06.*

[13] *Hunter, Lance , Craig Albert, Josh Rutland, Kristen Topping, and Christopher Hennigan.. 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 40 (2): 235–69. Accessed on 07 September, 2024. https://doi.org/10.1080/14751798.2024.2321736.*

[14] *Philip Kotler. 2000. "THE MARKETING CONCEPT." 2000. Accessed on 07 September, 2024. https://www2.nau.edu/~rgm/ha400/class/professional/concept/Article-Mkt-Con.html.*

[15] *Kliegr, Tomas, Bahník, and Fürnkranz. 2021. "A Review of Possible-Effects of Cognitive-Biases on Interpretation of Rule Based Machine Learning Models." Accessed on 08 September, 2024. Artificial Intelligence 295 (June):103458. https://doi.org/10.1016.*

[16] *Mallory Schlossberg. 2014. "One Of The Best-Moments On 'Colbert Report' Was When He Coined 'Truthiness' In 2005 | Business Insider India." December 19, 2014. Accessed on 11 September, 2024. https://www.businessinsider.in/one-of-the-best-moments-on-colbert-report-was-when-he-coined-truthiness-in-2005/articleshow/45568420.cms.*

[17] *Rhodes, Samuel. 2022. "Filter Bubbles, Echo Chambers, & Fake-News: How Social-Media Conditions Individuals to Be Less Critical of Political-Misinformation." Political Communication 39 (1): 1–22. Accessed on 19 September, 2024. https://doi.org/10.1080/10584609.2021.1910887.*

[18] *Lewandowsky, Stephan. 2019. "The Post Truth World, Misinformation, & Information Literacy: A Perspective From Cognitive-Science." Informed Societies, February, 69–88. Accessed on 06 September, 2024. https://doi.org/10.29085/9781783303922.006.*

[19] *Madalin Necsutu. 2023. "Moldova Dismisses Deep fake Video Targeting President-Sandu | Balkan Insight." December 29, 2023. Accessed on 12 September, 2024. www.balkaninsight.com/2023/12/29/moldova-dismisses-deepfake-video-targeting-president-sandu/.*

[20] *ET Online. 2024. "STOIC-Hits-India with 'Zero Zeno': Israeli Firm Tries to Disrupt Lok-Sabha Elections; Pushed Anti-BJP, pro-Congress Content - The Economic Times." June 1, 2024. Accessed on 06 September, 2024. https://economictimes.indiatimes.com/news/elections/lok-sabha/india/stoic-hits-india-with-zero-zeno-israeli-firm-tries-to-disrupt-lok-sabha-elections-pushed-anti-bjp-pro-congress-content/articleshow/110611373.cms?from=mdr.*

[21] *Kalev Leetaru. 2019. "The Real Danger Today Is Shallow Fakes And Selective Editing Not Deep Fakes." August 26, 2019. https://www.forbes.com/sites/kalevleetaru/2019/08/26/the-real-danger-today-is-shallow-fakes-and-selective-editing-not-deep-fakes/.*

[22] *Jeronimo Gonzalez. 2023. "AI Avatars Are Being Used to Spread Pro-Venezuela Propaganda |Semafor." February 21, 2023. https://www.semafor.com/article/02/21/2023/venezuela-uses-ai-avatars-to-disseminate-propaganda.*

[23] *Casey Tolan, Donie O'Sullivan, and Jeff Winter. 2024. "How a Biden AI Robocall in New Hampshire Allegedly Links Back to a Texas Strip Mall | CNN Politics." February 8, 2024. https://edition.cnn.com/2024/02/07/politics/biden-robocall-texas-strip-mall-invs/index.html.*

[24] *Bath, P. 2021. "China's Military Space Strategy | Vivekananda International Foundation." VIF. Accessed on 10 September, 2024. https://www.vifindia.org/article/2021/june/15/china-s-military-space-strategy*

[25] *Hunter, Lance , Craig Albert, Josh Rutland, Kristen Topping, and Christopher Hennigan.. 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 40 (2): 235–69. Accessed on 09 September, 2024. https://doi.org/10.1080/14751798.2024.2321736.*

[26] *Nathan Beauchamp, Mustafaga, and Bill Marcellino. 2023. "The U.S. Is not Ready for AI Fuelled Disinformation-But China Is TIME." October 5, 2023. Accessed on 06 September, 2024. https://time.com/6320638/ai-disinformation-china/.*

[27] *"China Tests US-Fault-Lines and Ramps AI Content to Boost Its Geopolitical Interests - Microsoft On the Issues." n.d. Accessed September 18, 2024. https://blogs.microsoft.com/on-the-issues/2024/04/04/china-ai-influence-elections-mtac-cybersecurity/.*

[28] *Ibid*

[29] *Ibid*

[30] *"The United States Isn't Ready for the New Age of AI Fuelled Disinformation | RAND." 2023. August 5, 2023. Accessed on 02 September, 2024. www.rand.org/pubs/commentary/2023/10/*

the-united-states-isnt-ready-for-the-new-age-of-ai.html.

[31] *Koh Ewe. 2024. "Microsoft China Uses AI to Sow Disinformation & Discord Around-the-World TIME." April 5, 2024. Accessed on 06 September, 2024. https://time.com/6963787/china-influence-operations-artificial-intelligence-cyber-threats-microsoft/.*

[32] *Rosa Lazarotto, Barbara, and Barbara-da-Rosa. 2023. "The Grass Isn't Always Greener on the Other Side: The Use of Digital-Astroturfing to Spread Disinformation and the Erosion of the Rule of Law." LSU Law Journal for Social Justice & Policy 3:9.*

[33] *Zhao Chenchen. 2022. "China Tightens Multi-Channel Networks Regulations - CGTN." March 18, 2022. www.news.cgtn.com / news/ 2022-03-18/ China-tightens-multi-channel-network-regulations-18tS8nOaUpy/index.html.*

[34] *"EU AI-Act-2024 Regulations & Handling of Deepfakes - BioID." n.d. Accessed September 19, 2024. www.bioid.com/2024/06/03/eu-ai-act-deepfake-regulations/.*

[35] *Wil Corvey. n.d. "Semantic Forensics." Accessed September 19, 2024. www.darpa.mil/program/semantic-forensics.*

[36] *Gec. n.d. "How the PRC Seeks to Reshape the Global Information Environment." Accessed September 19, 2024. www.state.gov/how-the-peoples-republic-of-china-seeks-to-reshape-the-global-information-environment.*

[37] *Hunter, Lance , Craig Albert, Josh Rutland, Kristen Topping, and Christopher Hennigan.. 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 40 (2): 235–69. Accessed September 18, 2024. https://doi.org/10.1080/14751798.2024.2321736.*

[38] *Ibid*

[39] *ibid*

[40] *Jansen, Sue C., Brian M., and Barbra Streisand. 2015. "The Streisand Effect and Censorship Backfire." International Journal of Communication 9:656–71. Accessed on 16 September, 2024. http://ijoc.org.*

[41] *Pilditch, Toby D., Jon R., Jens Madsen, and Sander Van Der Linden. 2022. "Psychological Inoculation Can Reduce Susceptibility to Misinformation in Large Rational Agent Networks." Royal Society Open Science 9 (8). Accessed on 17 September, 2024. https://doi.org/10.1098/RSOS.211953.*