Artificial Intelligence and Its Impact on National Security

Sqn. Leader Chhavi Prasad (Retd.)[1], Akash Das[2]

Abstract

This paper explores the growing impact of Artificial Intelligence (AI) on National Security, a field where its potential for both disruption and enhancement has become increasingly evident. While AI has transformed industries such as retail, telecom, and logistics, its influence on national security has raised both opportunities and risks. A compelling example of this is the ongoing Russia-Ukraine war, where AI-driven technologies have been deployed both as tools of aggression and defence. AI has enabled cyberattacks, disinformation campaigns, and deepfake manipulation, threatening state stability. At the same time, it offers solutions for improving intelligence gathering, countering disinformation, and war simulation using AI and strengthening cybersecurity defences. The paper examines the role of Generative AI (GenAI) and Predictive AI in shaping national security, focusing on their applications in countering false narratives, streamlining intelligence analysis, and optimising defence logistics and bringing more real-life simulation in war gaming. It also addresses concerns such as the weaponisation of AI in cyberattacks, and on usage of AI-enabled models to prevent honey trapping, and the challenges surrounding the trustworthiness of AI models in sensitive contexts. Drawing on case studies from countries like India and recent conflicts, the paper highlights efforts to counter AI-driven threats, including the development of indigenous AI systems to protect against foreign influence. The need for ethical AI frameworks, data privacy, and robust security measures to safeguard national cyberspace infrastructure is emphasised. Ultimately, this paper provides a comprehensive view of AI's dual-edged impact on national security, recognising its capacity to both safeguard and undermine state interests, and proposes strategies to harness its potential while mitigating risks.

Keywords: Artificial Intelligence (AI); Generative AI (GenAI); Predictive AI; National Security; Large Language Model (LLM), Retrieval Augmented Generation (RAG), Machine Learning (ML)

[1] Retd. Sqn. Leader Indian Air Force

[2] Data Scientist Oracle Corporation

Usage of Generative AI to combat disinformation

Disinformation campaigns can play a great role in creating instability within a country. In the context of India, ethnicity-related violence in Manipur[3] was being promoted by a disinformation campaign. Similar incidents have happened all over the world, where fake news, deepfake videos have been used to create chaos.

So, the need of the hour is to fight this disinformation campaign. Disinformation campaigns can be addressed across two fronts:

1. Having strict compliance and ethical AI-based frameworks in place, so that when using Generative AI (GenAI), fake news cannot be produced.
2. Another way is confronting the narratives by having counter counter-narrative with correct facts and figures in place.

The second approach is of utmost importance, as even with the compliance in place, there can be some leakage. Counter-narrative building must be done at a very high pace with the right facts and properly researched content. So, to achieve that, we must get this process automated. Hence best way to combat GenAI-related disinformation is by using GenAI to build a counter-narrative with the right facts and figures.

The approach for this is by using **Retrieval Augmented Generation (RAG).** An approach where the Large Language Model is integrated with an external data source, which can be domain-specific. For any press brief post incident, various data sources are being used, like on-ground reports, reports filed by the security forces, intelligence inputs, etc.

Similarly, for RAG, we can use all this data. All this data can be passed into a centralised database, which serves as a single source of information. This information is then fed into the LLM model, which will embed it and store it in a vector database.

By passing the right prompt, a very accurate well well-researched summary with precise facts and figures can be presented at a much lesser time and with much more accuracy than a human being.

Reactive approach is a way where entire web sources like social media, news channels will be crawled, and in case any information is present that is sensitive in nature, it will be compared using the state-owned Generative AI model. If the facts don't match, then this news can be flagged, and counter counter-narrative can be built. This entire process can be automated, thus cutting down on time and effort (Figure 1).

[3] https://economictimes.indiatimes.com/news/india/how-fake-news-and-videos-inciting-violence-in-manipur/articleshow/102065845.cms

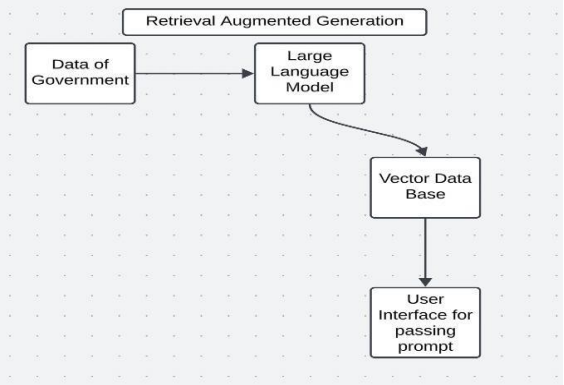

Figure 1: AI-Based Counter-Disinformation Model*. This figure illustrates how Generative AI, combined with Retrieval Augmented Generation (RAG), can be used to identify and counter disinformation campaigns by analysing multiple data sources and generating fact-based narratives. (Created by the author.)*

Generative AI in Streamlining Intelligence Analysis

In case of any country, terrorist attacks can be prevented effectively with fast and accurate intelligence input analysis.

Intelligence related data might come from various sources like Human Intelligence, Signal Intelligence, information gathered across various border check posts or change in patterns across certain geographies within the country or outside the country, all this can be an actionable insight. Now for an intelligence analyst the task is to combine all this data and discover the hidden pattern within this to derive the conclusion.

In many cases, like the 26/11 or 9/11 terrorist attacks, it has been noticed that although a lot of information was present, the dots were never connected, and hence the attack was not prevented[4].

In such cases, today's Generative AI can play a key role in converting, detecting, and summarising intelligence data. First, let's understand what Generative AI (GenAI) is. GenAI is a process of Artificial Intelligence where based on User prompt information is being generated. Below snapshot is one such example (Figure 2).

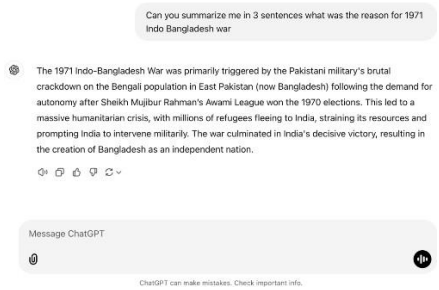

Figure 2: Snapshot of Chat GPT. *This figure represents the process of using Generative AI. Here, the author tries to demonstrate that when an input is being passed as a prompt, the content against that prompt was generated by summarising the results from all the available sources. (Created by the author.)*

This same approach can be taken for intelligence analysis. In the case of GenAI, there are 2 major components: Prompt and LLM. Prompts are for taking user inputs, and Large Language Models (LLM) are to capture relationships amongst the data in the form of word embeddings and then using attention mechanism to detect the relevant data amongst the data sources, summarizing them and then transforming them into meaningful context and presenting it in front of the user. Now, LLMS mostly use web data. But for an organisational-specific task, we need to use the data specific to that organisation. Data is being fed into the

[4] https://ctc.westpoint.edu/improving-indias-counterterrorism-policy-after-mumbai/ , https://www.brookings.edu/articles/9-11-and-the-reinvention-of-the-u-s-intelligence-community/

LLM, which is known as **Retrieval Augmented Generation (RAG).**

In the case of Intelligence analysis, the data captured can be an image, an audio clip, or data captured using Human intelligence. All of this can be fed or passed to the LLM model. The LLM model then converts this into embeddings and stores them in a vector database. Now, when an analyst passes a prompt like "Can you spot me suspicious audio clips in the last 2 months originating from Kashmir?" Then it will be an easy approach to fetch those records, but also fetching that data will be much faster, efficient as compared to previous approach, which might have involved either writing a SQL query or looking for data based on the STD code, etc. Here, due to the capability of the LLM to summarise, it can fetch records of calls which are present in the dataset captured using triangulation, fetch records based on the STD code, and based on the satellite phones, etc. Today, there are Gen AI models that are so advanced that they can do in-depth analysis of audio and images. If such different kinds of models are being chained, then, like a human analyst, it can give insights based on a combination of satellite imagery, signal intercepts, and human inputs.

The analysis in this case will be much more insightful and can derive many more hidden patterns due to the neural networks, which could have been missed by human beings.

Cyber Attacks and the Impact of AI

Worldwide cyberattacks have always been a menace. Whether it's a DDOS attack or a breach like Target Corporation, cyberattacks lead to vital information leakage, which can cause financial losses, but more importantly, it can impact national security as bad actors are getting access to critical applications.

In India in 2024, 1.2 billion cyberattacks occurred. In the case of India's financial sector, 16 million cases have been dealt with by the Computer Emergency Response Team (CERT), which was just 53000, seven years back in 2017.[75]

The famous ransomware attack at AIIMS, Delhi or a similar attack with small Indian cooperative banks are amongst famous ones.

This number is going to increase in future.

AI coming into the scenario; cyberattacks become much more dangerous. AI assisted cyberattacks known as offensive AI can lead to faster and deeper breaches. Hence damage will be much more.

Earlier the cyberattacks were being done mostly by lot of human efforts. Human efforts included analysis of the software vulnerabilities by using different network scanning tools and web scanning tools. This would make the process slower. Even phishing emails were generated manually. Thus, DDOS attack, phishing, unauthenticated access all that could have been detected using rule-based detection engines and prevented.

[75] https://www.darkreading.com/cyber-risk/india-s-critical-infrastructure-suffers-spike-in-cyberattacks

But with AI, the attacks are becoming much more sophisticated. For any cyber-attack, there are various phases. The first phase is the reconnaissance, where the hacker explores the application and the network to understand the vulnerability. Today, with tools that utilise neural networks, exploration and learning happen much faster. Thus, resulting in faster and more efficient reconnaissance. The second phase is the actual attack. This attack can be by Brute force to bypass authentication. Today, there are tools that have been trained on passwords from previous breaches and hence can generate much closer passwords to bypass the authentication. Similarly, image detection and AI are being used to bypass the CAPTCHA mechanism. For phishing attacks, the email generated using Generative AI looks very real and authentic. Thus, misleading the user, who in turn ends up clicking the phishing links. Similarly, the DDOS attacks can be automated using AI.

The third phase for any cybersecurity attack is evading detection. For any attack there are identification signatures. Today with the help of AI the identification signatures can be changed very fast thus evading detection (Figure 3).

So, when Generative AI is available as an open-source technology and similarly lot of machine learning models are available in open source, they can be utilised for breach; this becomes a scary space. Similarly, lot of datasets to train these models are equally available.

This establishes the fact that not only attack count will go high as even novice hackers can try a breach; but also, the damage will be much higher in magnitude and preventing it would be a challenge. Thus, thinking holistically with lot of foresightedness in cybersecurity space is need of the hour.

As a country cybersecurity framework must be brought in at the national level. The framework must have strict compliance when it comes to cyber infrastructure, and there must be a stress for indigenous infrastructure in place so that other hostile state actors lack the advantage of supplying a faulty piece of hardware that can be breached. Security testing of applications must be rigorously done, and AI must be adopted for testing that software so that we can simulate attacks. There should be a stress on having own datacentres and data encryption, so that no data goes out of the country and security is guaranteed. This approach will ensure fewer attacks like Man in the middle attack (MiM). When such an ecosystem is being built, which promotes strict compliance, strict security testing and usage of secure indigenous systems, then we can expect a sanitised cyberspace within the country.

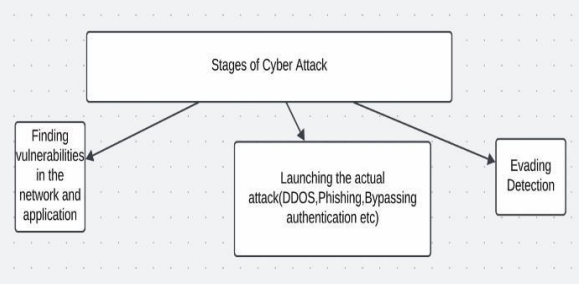

Figure 3: AI-Powered Cyber Attack. *This figure demonstrates 3 stages of AI based cyber-attack. This includes network analysis, launching of attack, evading detection post attack (Created by the author.)*

Usage of AI to Prevent Honeytrapping

Honey trapping is one of the oldest trades in spy craft. Back from ancient India, Chanakya utilized Vishkanyas for the purpose of leaking sensitive information related to enemy.

India has been victim of honeytrapping. Famous cases of honeytrapping involve Madhuri Gupta case, Brahmos engineer being honey trapped and a senior DRDO official being honey trapped which lead to leakage of sensitive vital information.

Now counter-intelligence teams can be successful in detecting such compromised individuals and take legal actions; but the damage is already done. Hence it should be prevented using early detection approach.Artificial Intelligence can play a great role. Whenever any user exists over any social media, he has a particular pattern which identifies his behavior. Such behavior-based segmentation is widely used in ecommerce websites for recommendations.

Such patterns can be used for anomaly detection. In case of social media genuine users with no intent to cause any espionage and unwanted accounts with sole aim of connecting with someone and then developing relationships to extract information will have different behavior.

By using clustering models, we can segregate these accounts into different clusters based on their behaviour. These segregated clusters can be studied further, and exploratory data analysis can be performed on them to understand their behaviour. Based on these findings, we can come up with features that can be fed now into supervised learning models. Now these models are trained based on these features and labelled data marking which communication or account type is a honeytrapping account; these models can now be employed to detect anomalies and flag communications that are suspicious. Data from previous honeytrapping cases can also be used to enhance model accuracy.

Another approach is based on the usage of AI-based honeypots (Figure 4). A **honeypot** is a decoy system or resource set up to attract, detect, and study unauthorised access attempts.

Similar honeypots can be used to attract those actors who are involved in honeytrapping. The hostile elements can end up believing that this account belongs to a person of importance. They might initiate honeytrapping conversations which will help intelligence agencies to identify them and isolate them.

By using Generative AI we can create authentic identity which will lure the attackers. The interaction can be made more realistic using GenAI. The learning from these conversations can be faster by using machine learning. This learning can now be fed into ML models which are being used to detect anomalies in the communication content. Thus detection and isolation of the hostile elements are being achieved faster.

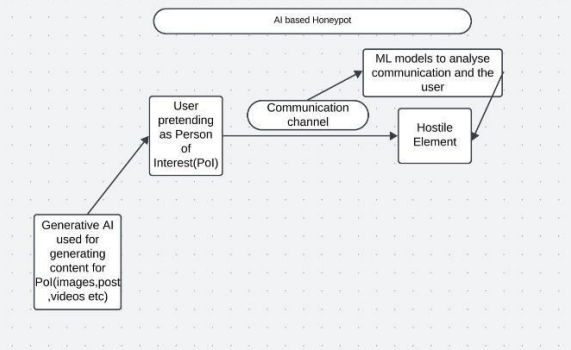

Figure 4: AI-Powered Honeypot. *This figure demonstrates the usage of Generative AI to create Honeypots (Created by the author.)*

Trustworthy AI systems and Data privacy

With usage of AI across every sector, there is a dire need to have AI systems which have a certain level of trustworthiness. Trustworthy AI are the systems which are explainable, fair, interpretable, robust, transparent, safe and secure. These qualities create trust in AI systems.

In past there has been certain cases where AI models have shown bias or inaccuracy. For example, Amazon recruitment system used one ML model which preferred male over female.

AI models are being used for Governance systems and critical applications, where fairness is of utmost importance. If AI models are not robust and fair, then certain areas concerned with national security can be impacted. Like Cybersecurity AI based detection engines can fail, misinformation campaign can be promoted using Generative AI, Autonomous weapon systems can be sabotaged causing collateral damage, detection systems can be erroneous thus attacks won't get prevented, governance systems like traffic management can malfunction causing accidents, intelligence gathering and analysis can be wrong, economic systems like stock prediction can go wrong causing financial instability.

Thus, ensuring Trustworthy AI must be part of national security strategies. AI models can be compromised by two forms of attacks.

Evasion attack*:* The Model is misled during runtime, like wearing spectacles can cause failure of the image detection system.

Poisoning attack: Data is being corrupted so that the model gets trained on wrong data; eventually causing wrong predictions or content generation.

When it comes to countries like the USA or European countries, a lot of steps have been taken to ensure trustworthy AI models. In the case of USA, a framework has been defined which is known as NIST AI Risk Management Framework to ensure trustworthy AI models. In case of European countries European AI model has been defined to ensure trusted AI system.

India does not have a single comprehensive AI framework in place, although initiatives have been taken by NITI Aayog. Indian AI framework needs to stress on below points:

1. Robustness of the AI systems: The AI model must be prone to small perturbations and hallucinations.
2. Data Diversity: Any AI model must be trained with data across all segments, so that results are fair.
3. Explainability: Behind any prediction from AI model; there should exist some form of explanation. This is lacking specially for deep learning models. Some common frameworks which can be adopted are Local Interpretable Model-agnostic Explanations (LIME) and SHAP (Shapley Additive explanations).
4. Evaluation: Evaluating AI model output by subject matter experts.
5. Data Privacy: Data privacy is an important aspect for any AI model. If data privacy is breached poisoning attack can be performed. In case of European Union a much matured framework exist to protect the data of European citizens. But in India Digital Personal Data Protection Act (DPDP Act) is concerned of data processing within India.

6. Thus the very same data can be exploited by the hostile nations to run analysis and target certain segments with misinformation.
7. Development of indigenous systems: For trusted AI systems, data must be protected, shielded from analysis, data integrity must be maintained, data should lie within India, and procedural bias should not be introduced during model development. To ensure all this, we need indigenous systems. Open source or foreign technology is prone to breaches and misuse of data. We need data centres within our country; we need indigenously developed hardware for such data centres so that any form of malicious patches is not being deployed. For the interoperability of data amongst various government organisations, it will be best to have our own fully developed cloud framework to ensure data integrity during transmission. We need to take giant steps towards software development, too. LLM models that exist today on the internet are black box in nature and lack explainability. Whether these models internally have some sort of bias is not clear to us. In such cases, to combat disinformation, we should have our own exclusive data sources and also own indigenously developed AI models or GenAI models.

Indian Defence Logistics and AI Integration

The Indian defence services have traditionally relied on a variety of SAP platforms for the management of logistics or inventory needs, like the ILMS (Integrated Logistics Management System) for the Navy, IMMOLS (Integrated Material Management Online System) for the Indian Air Force, and the Inventory Management System of the Indian Army. In these platforms, tremendous amounts of data are generated with respect to procurement behaviour, stocks, maintenance plans, etc. This data could be considered as a robust base for training ML models, which will enable predictions and better intelligent decision making to optimise defence logistics. There are numerous challenges with logistics. For example, during Operation Snow Leopard, Israeli missile systems deployed in high-altitude detachments required missile replenishment from the parent base within a designated period because they can only remain on the launcher for a limited number of days. After that, the missiles must be sent back to the base for testing. This task can only be accomplished using C-17 or IL-76 aircraft; hence, the Air Force needs to optimise air movements every month. The goal is to ensure that missiles are collected from the parent base, transported to high-altitude locations, and the older missiles are brought down for testing, all with minimal movement. There are no automated systems in place to manage this process. If the unit's operations head forgets to raise the air requisition on time, the movement may not happen as scheduled. This type of coordination is time-consuming and inefficient. If AI systems are implemented, the entire workflow can be streamlined. Due to the lack of such automation, missiles often exceed their allowed deployment period. AI has the potential to transform inventory management by automatically determining the reorder levels and quantities. Like Amazon or Target, which use ML for such scenarios, it is possible to produce a demand forecast that is accurate based on snag trends with seasons and usage, to support critical material inventory without excessive accumulation. It could introduce new search mechanisms for procuring materials or performing orders from uncertain

vendors. Sophisticated platforms like AnyLogistix, capable of returning defence-like dashboards, easily integrate into real-time supply chain control systems. Thus, dynamic routing of freight and predictive maintenance for cost-effective and timely logistics support.

But before it can be fully leveraged by the Indian defence services, there is a critical need to address basic deficits in digital literacy. A significant amount of procurement and logistics continues to be paper-based, and even in digital systems, the use of facilities is not maximised due to an insufficient knowledge of basic tools like Microsoft Excel. Without addressing these deficiencies, the introduction of AI introduces the risk of inefficiencies. Hence, digital literacy must be addressed.

Figure 5: AI-Integration for Logistics and Strategy. *This figure demonstrates the usage of AI for logistics planning and War gaming (Created by using Generative AI, ChatGPT.)*

Military War Games and Strategy Development

The Indian Army has a longstanding tradition of conducting war games to enhance strategic planning and operational readiness. These exercises, ranging from battalion-level simulations to large-scale operations like Operation Brasstacks, have been instrumental in refining military tactics and doctrines. However, what often follows these exercises are tabletop discussions that lack in-depth analysis. Instead of conducting live drills, participants rely on photography to simulate outcomes. For example, during a simulated attack on an airfield, defenders would send interceptors to engage attackers. Capturing an opponent on camera signifies a successful hit. Finally, an umpire evaluates the plans, their execution, and the evidence to declare the winner. Gen AI into an online war-gaming platform can revolutionise risk management by creating immersive, dynamic simulations that mirror complex real-world scenarios. Generative AI enables these situations to be responsive and interactive, changing according to players' choices and adding new factors, thus checking improvisational skills and critical thinking. Players can play the roles of decision-makers, advisors, or operators and are challenged using AI as both a partner and adversary.

Real-world data derived from economic reports, weather reports, or news reports can be utilised to design the game's sequence. E.g., a simulated cyberattack may lead to financial destabilization, and participants may be asked to devise solutions to reduce the damage.

Due to utilisation of AI in such platforms; actions taken can be evaluated and the same feedback can be utilised for training the model further. Generative AI being integrated into such model allows user to experience very real time intricate scenarios and also allows multiple participants thus enhancing strategy making skills and interoperability and coordination.

Conclusion

Artificial Intelligence is rapidly transforming national security, acting as both a shield and a sword. On one hand, AI-driven technologies help nations strengthen their defences, detect cyber threats, and streamline intelligence analysis. On the other hand, they introduce new risks such as deepfake-fuelled disinformation, AI-powered cyberattacks, and sophisticated espionage tactics.

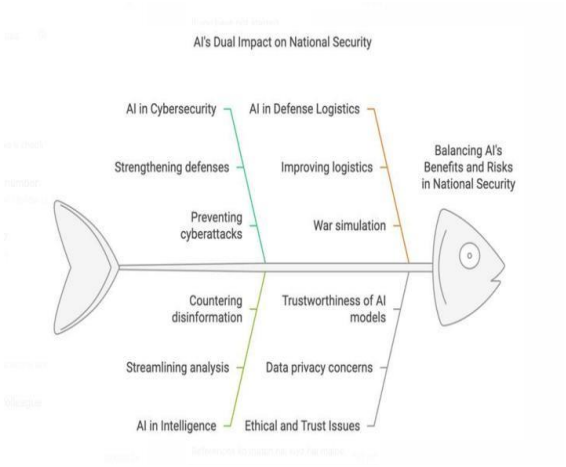

Figure 6: Conclusion (*Created by using Generative AI, ChatGPT)*

The challenge is clear: while AI can protect a nation, it can also be weaponised against it. As seen in conflicts like the Russia-Ukraine war and domestic security incidents in India, AI's role in shaping modern warfare and information control is undeniable. Governments must act swiftly to harness their potential responsibly while building strong safeguards to prevent misuse. The path forward lies in creating indigenous trustworthy AI system, ones that are ethical, transparent, and resilient against manipulation. India, like many other nations, must invest in indigenous AI development, enforce strict cybersecurity policies, and prioritise digital literacy across defence sectors. AI alone isn't a solution; it's a tool that must be wielded with foresight and responsibility. By fostering collaboration between policymakers, technologists, and security experts, we can ensure that AI remains a force for protection rather than a source of vulnerability. In the end, the true measure of AI's impact on national security will not be just in its capabilities but in how wisely we choose to use it.

References

Blessing or curse? A survey on the Impact of Generative AI on fake news https://arxiv.org/abs/2404.03021

Generative AI and Disinformation: Recent Advances, Challenges and Opportunities https://edmo.eu/wp-content/uploads/2023/12/Generative-AI-and-Disinformation_-White-Paper-v8.pdf

Retrieval Augmented Generation for Large Language Models https://arxiv.org/abs/2312.10997

Impact of Artificial Intelligence on the future of Cyber Security https://mecsj.com/uplode/images/photo/The_Impact_of_Artificial_Intelligence_on_the_Future_of_Cybersecurity.pdf

Advancing Cybersecurity: a comprehensive review of AI-driven detection techniques https://mecsj.com/uplode/images/photo/The_Impact_of_Artificial_Intelligence_on_the_Future_of_Cybersecurity.pdf

Honeymodels: Machine Learning Honeypots https://arxiv.org/abs/2202.10309

TrustLLM: Trustworthiness in Large Language Models https://arxiv.org/abs/2401.05561

How Amazon uses AI to make its Transportation network flow https://freight.amazon.com/newsroom/2024-amazon-ai-network?ref=E_CO_R4S_2024-amazon-ai-network_WEB_AR_

Digitisation of Supply Chain https://balloonone.com/resources/amazon-whitepaper/

Artificial Intelligence in War Gaming: An Evidence-based assessment of AI applications https://cetas.turing.ac.uk/sites/default/files/2023-06/cetas_research_report_-_ai_in_wargaming.pdf