SPECIAL
COMPETITIVE
STUDIES
PROJECT

Centre for
Emerging Technology
and Security

RESEARCH REPORT

Applying AI to Strategic Warning

Nandita Balakrishnan, Anna Knack, and Timothy Clancy

APRIL 2025

Table of Contents

Executive Summary

This joint report from the Special Competitive Studies Project (SCSP) and The Alan Turing Institute's Centre for Emerging Technology and Security (CETaS) assesses the potential of current artificial intelligence (AI) systems to make assessments about geopolitical events and the path ahead for applying AI to strategic warning for national security and intelligence.

Strategic surprise–an unforeseen event or development often driven by an adversary–can jeopardize human lives and impose substantial security and financial costs. Consequently, national decision-makers are continuously seeking ways to improve their detection of changes to the geopolitical picture, adversarial activities and intentions, and exogenous strategic shocks that may impact their core interests. The rapid improvements in AI systems' ability to digest and analyze enormous amounts of data, and reports of commercial AI systems successfully predicting events leading up to the invasion of Ukraine, have increased interest in using AI to improve early warning and assessments about geopolitical events.

Furthermore, strategic AI competition between Western democracies and the so-called Axis of Disruptors (China, Russia, North Korea, and Iran) has intensified amidst the emergence of Chinese AI models, which claim to be closing the gap on U.S. dominance in AI technology while using fewer resources. Against this backdrop, cultivating a performant AI system that gives decision-makers more time to respond to crises and effectively allocate resources would be a significant contribution to decision advantage.

Our research has found that, while there is currently no AI system that can reliably and consistently predict geopolitical flashpoints or forecast their implications with high accuracy and precision, the state of the technology is changing rapidly, and the advent of Artificial General Intelligence,[1] which some experts predict could arrive within two-to-three years, could change the playing field. The next generation of AGI systems could provide a considerable uplift to strategic warning in several ways in the medium term, giving decision-makers more time to pre-empt or prepare to respond to crises. The two most promising use cases identified in this research are using AI to:

1. Track conflict risk indicators more accurately, by leveraging increased quantities and types of data.

2. Identify possible outcomes and scenarios immediately after a shock or trigger happens. This could be particularly valuable on regions or topics that usually receive scant attention from intelligence services.

[1] Artificial General Intelligence (AGI) is used to describe an AI system that is at least as capable as a human at most tasks. Meredith Ringel Morris, et al., Position: Levels of AGI for Operationalizing Progress on the Path to AGI, arXiv (2024).

Currently, the two main challenges inhibiting AI from making compelling predictions about geopolitical events are:

1. **Data scarcity and inconsistency** because geopolitical events themselves are relatively rare and their triggers can be random; however, it is difficult to gather enough reliable information for robust predictive models. In some regions, data may not be collected at all or may not exist in a form that is suitable for training an AI model. Historically, conflict data has been gathered in an unsystematic way and is not consistently parameterized by conflict risks, conditions, triggers, and tipping points.

2. **Modeling the decisions of individuals** whose intentions may sometimes be random, impulsive, or opportunistic, and which are challenging to deduce even with high-quality classified intelligence. This includes decisions by state leaders, their advisers, dissenters, individuals with deceptive intentions or the actions of individuals that lead to tipping points like the fruit seller who set himself on fire and triggered The Arab Spring.

Building on the findings of this project, we recommend the U.S. and UK national security communities embark on an ambitious, collaborative three-phase project that tackles the critical step of addressing systemic data infrastructure challenges and collating disparate model outcomes that will make an eventual leap to an integrated simulation of all possible risk and stabilization factors more likely to succeed. Many of the technical challenges will be overcome not by more AI research and development but by an ambitious project that radically tackles the IC's data challenges. This project should enable the national security community to leverage best-in-class AI models and build confidence in system reliability without being delayed by technical bottlenecks and unfeasible costs.

- **Phase 1: Establish AI training and testing of data foundations** by leveraging all relevant sources of geopolitical event data. Establish shared standards to ensure consistency in geopolitical event risk data collection and assessment, including non-traditional data sources and quantifying complex behavioral and decision-making data such as decision makers' and groups' intentions, grievances, and biases.

- **Phase 2: Nurture a suite of best-in-class models that assist analysts**, while prototyping an initial generalizable model of how geopolitical events of interest materialize. The suite of government and industry models would assist in analytic tradecraft and be trained on different types and classifications of data to circumvent major data fusion, traceability, AI security, governance, and cost challenges that directly pursue an integrated simulation of the world will likely entail. The outputs of models will be data points for the analyst rather than a source of finished intelligence, and data points will be compared and triangulated. The models would be continuously retrained and re-weighted while methods of scientifically validating and measuring data quality are developed.

- **Phase 3: Develop an integrated AI-based simulation platform** for strategic warning. The platform should represent geopolitical risk and stabilization factors in high fidelity with high-quality datasets, augmented by synthetic data, and the simulation will run these factors thousands of times to understand the different ways conflict might erupt.

There are three key policy considerations associated with this proposed initiative:

- **Cost.** This will be an expensive endeavor, with initial costs for set-up equipment, data engineering, workforce training, and supporting infrastructure. However, these costs must be balanced against the opportunity cost and potential consequences of not adopting AI for strategic warning.

- **AI sovereignty.** There will be benefits and trade-offs around whether such a tool is pursued unilaterally or multilaterally. The involvement of allies and partners could reduce duplication, spread risks, and reduce individual costs, but it could also introduce delays and challenges in harmonizing regulatory frameworks and compliance requirements.

- **Industry collaboration.** Policymakers must consider whether they wish to prioritize autonomy by building such systems in-house, which would incur more direct research and development costs, or sacrifice a degree of control and autonomy by contracting out R&D processes to trusted industry partners.

We assess that radical enhancements are possible but will only be achieved by starting to address the upfront challenges and costs of the systemic data infrastructure challenges in the UK and U.S. national security communities. Unlocking the transformative potential that AI offers for strategic warning will not be achieved by procuring the lowest cost yet technically feasible tool. This will be an expensive, time-consuming, and politically sensitive project. However, it could prove pivotal in maintaining decision advantage over aggressors in future conflicts or geopolitical crises.

Scope Note and Methodology

Conducted through a collaboration between the Special Competitive Studies Project (SCSP) and the Alan Turing Institute's Centre for Emerging Technology and Security (CETaS), this project seeks to examine the extent to which artificial intelligence (AI) and intelligent automation can enhance the intelligence community's ability to forecast key geopolitical events of interest, such as leadership decisions, political transitions, and military aggressions. By improving forecasting capabilities, intelligence analysts will be better equipped to provide policymakers with the most timely and actionable intelligence. To achieve this goal, the project is structured around three main objectives:

- **Technological Assessment:** Establishing the current technological state of play in the field of AI and geopolitical forecasting to highlight existing models in use and under development. An expanded version of this chapter, which details specific technological limitations and opportunities, was published in November 2024.[2]

- **Technical Challenges:** Identifying the AI technical challenges that must be overcome to progress from current applications toward an ideal single-integrated platform.

- **Cost-Benefit Analysis:** Exploring both the quantifiable and non-quantifiable costs of developing such a system and contextualizing these costs against the potential benefits.

To address these research questions, the research team conducted a comprehensive review of academic and gray literature from the past three years, establishing the current technological state of play to apply AI to strategic warning and assessing the current data infrastructure for existing tools. This review also outlined the criteria for optimal use cases, ideal dataset requirements, technical challenges, potential solutions, policy challenges, and foreseeable costs. In parallel, the team evaluated private sector AI capabilities and cutting-edge AI-driven scenario modeling tools to understand their application in strategic warning. Semi-structured interviews were then conducted with UK and U.S. government producers and consumers of conflict and political instability modeling, as well as with academic experts, OSINT analysis providers, and technical experts from the industry.

[2] Anna Knack & Nandita Balakrishnan, The State of AI for Strategic Warning, Centre for Emerging Technology and Security (2024).

Introduction

In the fraught aftermath of the 9/11 attacks against the United States, the world witnessed how a few missed warning signs can claim thousands of lives, trigger decades of conflict, and alter the national security landscape for a generation. Intelligence failures[3] will always be a risk for any state's intelligence community (IC) because human analysts are pressed for time, working with limited data, or can be wedded to existing mental models of how the world works.

The stakes of strategic surprises speak to the primary role of the intelligence analytic community: to give policymakers strategic warnings about key events of interest. For U.S. policymakers, their primary national security concerns are turned towards understanding what is happening in the Axis of Disruptors (China, Russia, Iran, and North Korea) because the actions and policies of these adversaries have the highest impact on U.S. national security priorities. These policymakers turn to their IC to give strategic warnings as early as possible about the possibility of Chinese troops moving in the Taiwan strait, Russia escalating (or de-escalating engagement) in Ukraine, the threat of a nuclear launch from North Korea towards Seoul, or the possibility of a political transition in Iran.

For UK counterparts, their analysts must thoroughly assess how middle-power states and internal conflicts could evolve, ensuring they anticipate and address potential impacts on the nation's broader security landscape. Given the UK's status as a global financial hub with strong security partnerships, the ascendance or destabilization of middle-power states, particularly in the Commonwealth, can directly impact British interests and regional stability. Moreover, aggressive actions by China, Russia, Iran, or North Korea in these states could amplify threats such as conflict spillover, terrorism, or unchecked proliferation, all of which pose significant concerns for the UK's foreign policy and homeland security.

For analysts on both sides of the Atlantic, the goal of monitoring adversarial activities, intentions, and capabilities is to give policymakers as much time as possible to craft a response, whether that is executing a deterrence strategy, shaping diplomatic engagement, or taking robust offensive or defensive measures. Early warning enables the government to allocate resources effectively and strengthen alliances while also safeguarding U.S. and UK interests and global stability. Such foresight not only reduces the risk of strategic surprise but also helps maintain a competitive edge in a rapidly changing geopolitical landscape.

Right now, this approach is primarily human-executed; while some AI tools are used for bespoke purposes,[4] they have not been adopted on a large scale. The benefit is that such assessments are easily explainable, and the chain of accountability is transparent. However, this comes at the expense of speed and the limited volume of intelligence sources that a human

[3] The 9/11 Commission Report, National Commission on Terrorist Attacks Upon the United States at 8 (2004).
[4] Frank Konkel, The U.S. Intelligence Community is Embracing Generative AI, Government Executive, (2024); Rapid Explanation, Analysis and Sourcing Online, IARPA (2024).

analyst can manually process. Analysts could misinterpret data (e.g. signals or classified intelligence), misjudge the data's credibility, and even fail to incorporate relevant data altogether, often due to a lack of knowledge of a data source's existence. Furthermore, an analyst can struggle to manually incorporate both quantitative and qualitative data; as a result, there can be restrictions to conclusions they can draw from data, particularly when it comes to identifying anomalies or causal links. This is all assuming that analysts have access to all the relevant data that they need and that an IC has access to all the personnel that it needs, which is increasingly untrue.

Today, the stakes of those potential failures could not be higher. The attack surface has increased with threats from not only state adversaries but also technologically sophisticated non-state actors who have a wide range of tools that can be deployed against U.S. and UK interests. While conventional attacks are always a threat, gray zone or hybrid conflict, as well as cyberattacks against critical infrastructure or financial systems, can be just as grave a threat.[5] Therefore, it is imperative that analysts are fully equipped with the tools necessary to give warnings on all these possible threats. This is where AI tools, especially those that help humans process and analyze data faster, could be indispensable in helping U.S. and UK analysts and policymakers maintain their decision advantage.

While the two ICs might already be experimenting with adopting AI tools into their workstreams, they must think about longer-term benefits and capabilities as this technology evolves. Intelligence analysis offers both an easy access point[6] and an opportunity for high impact. The ICs should be looking to leverage any tool, system, or technology that allows for quicker and more comprehensive indicators and warnings of geopolitical events of interest. Research and development both in academia and the commercial sector demonstrate this is a growing area of interest, making it a prudent avenue to explore, especially in the context of a human-machine team. AI has the potential to help policymakers spend more time on the strategic—rather than tactical—aspects of decision-making.

Within the current technological state of play for geopolitical strategic warning, there is no unified AI-powered system that can perfectly predict outputs or forecast security implications of interest, even for the Axis of Disruptors, where the most attention is being paid. However, investments in the technological developments already underway will allow the ICs to potentially make significant strides in improving crucial components of the strategic warning process. For example, Rhombus Power's deployment of AI to make reportedly accurate predictions of Russia's invasion of Ukraine and the U.S. military's Raven Sentry predictions of Taliban attacks in Afghanistan demonstrate the clear advantages AI could have in making geopolitical forecasts.[7] With the technological landscape impacting the threat landscape, the traditional scope of strategic warning is changing. Moreover, the ICs will need to start

[5] Sean Monaghan & Tim McDonald, Campaigning in the Grey Zone, RAND Corporation (2024).

[6] The Future of Intelligence Analysis: U.S.-Australia Project on AI and Human Machine Teaming, Special Competitive Studies Project (2024).

[7] Frank Bajak, US Intelligence Agencies' Embrace of Generative AI is At Once Wary and Urgent, The Associated Press (2024); How America Built an AI Tool to Predict Taliban Attacks, The Economist (2024).

incorporating technological vulnerabilities and opportunities in strategic warning assessments, especially under the expectation that adversaries will make similar investments. To be able to effectively understand how to give strategic warnings about this technology, the ICs need to be using this technology.

Chapter 1: Can Current AI Systems Make Assessments about Geopolitical Events of Interest?

For analysts and policymakers, an AI-powered geopolitical prediction tool would likely serve two distinct but not mutually exclusive purposes. First, it could allow analysts to better focus on key countries of concern, such as the Axis of Disruptors by augmenting the close monitoring already done by human analysts. For example, a 24/7 monitoring system that gave real-time assessments about Chinese or Russian troop activity or potential political instability would allow analysts to get insights to policymakers even quicker. A second opportunity would be to apply geospatial AI innovation to monitor the rest of the world, quantify causal mechanisms for conflict,[8] or to look back at historical conflict data to identify conflict risk indicator patterns that human analysts may have missed.

Together, this could contribute towards improved early warning that would help enable the IC to focus human personnel and other resources on their most pressing priorities and reallocate resources as necessary when the system detects anomalies in less closely observed countries. In either case, an AI-enabled tool would be most helpful if it could help analysts better predict discrete events, such as an impending coup or military attack, and forecast the implications of those events, such as assessing how the economy and geographic neighbors will respond.

Political scientists have long used quantitative models and simulations to predict the propensity for political stability or instability, which have allowed them to identify critical correlative patterns, including those that occur across different regions and time.[9] Analysts have used early machine learning tools to incorporate more data and thereby sharpen their assessments, as seen in the ViEWS competition,[10] which tested models of conflict in Africa against one another. As computing power has expanded, these models have been able to

[8] Global Urban Analytics for Resilient Defence, The Alan Turing Institute (2024).

[9] Michael Mobius et. al., AI-Based Military Decision Support Using Natural Language, Institute of Electrical and Electronics Engineers (2023); Scotty Black & Christian Darken, Scaling Artificial Intelligence for Digital Wargaming in Support of Decision-Making, NATO (2024); Joost van Oijen & Pieter de Marez Oyens, Empowering Military Decision Support Through the Synergy of AI And Simulation, NATO (2023); Cordis Europa, Using Machine Learning to Identify Political Violence and Anticipate Conflict, (2023); Mirco Musolesi & Akin Unver, Computational Modelling of Civil Wars, The Alan Turing Institute (2024).

[10] The Future of Intelligence Analysis: U.S.-Australia Project on AI and Human Machine Teaming, Special Competitive Studies Project (2024); The Pilot Early-Warning System (ViEWS) Views Forecasting (last accessed 2025).

incorporate even more data, address missing data issues, and allow for more sophisticated techniques that have even lent themselves to some causal inferences.[11]

There is currently no AI-powered system that can consistently predict with high accuracy geopolitical flashpoints or forecast their implications. There are AI-powered open-source intelligence platforms that fuse text data or satellite imagery for real-time monitoring of geopolitical events in order to identify emerging threats. Some platforms, which are used both by ICs and the commercial sector, scan social media to track potential flashpoints of violence, but their primary role is to give contemporaneous and immediate updates to users rather than to make actual predictions of events. However, when it comes to predictive analytics, there are several models and tools available that address essential components.

For the purposes of this study, we focus on models that aim to predict or forecast geopolitical events; however, it is worth noting that there are similar tools under development that focus on the scenarios and the implications of geopolitical events instead.

In both the commercial and academic sectors, we are seeing three broad categories of models that profess to model geopolitical events as a dependent variable:

1. **The prediction of geopolitical risk:** These models draw on large datasets to compute a quantitative risk score of political instability for a given geographic region, usually a state. The product of such models is often what is known as a heat map, which helps analysts make comparisons across regions and pinpoint areas of focus by assessing the conditions that make a geopolitical event more likely. The claimed benefit of such models is that they can be globally focused, and make longer-term predictions (with some models on the market claiming to predict conflict risk two years in the future),[12] while still helping users detect essential shifts. However, it can be challenging to articulate how these quantitative outputs are meaningfully additive from the qualitative assessments human analysts or subject matter experts in the field are already able to provide.[13] For example, if an AI model predicts relative stability in Denmark two years out and relative instability in Niger in that same timeframe, most policymakers would likely say they did not need the model to make those assessments. Some experts argue that if we are monitoring early warning for tactical changes, then ground sentiment and human experts could more efficiently identify the risk of a novel conflict outbreak than AI can at present.[14]

2. **The forecasting of specific shorter-term outcomes of interest:** Drawing on lessons from models like economic and weather forecasting for which there is rich data that

[11] Rachel Myrick & Cheng Wang, Domestic Polarization and International Rivalry: How Adversaries Respond to America's Partisan Politics, The Journal of Politics (2024).

[12] Author interview with industry participant (November 2024); Author interview with academic participant (October 2024); Author interview with industry participant (September 2024), Author interview with academic participant (October 2024).

[13] Author interview with academic participant (October 2024).

[14] Author interview with academic participant (October 2024).

can help analysts make forecasts of an event over a slightly longer timeframe, some experts were optimistic. They believed existing models could help forecast uncertain outcomes of more expected events in situations where they had better access to robust, labeled data. One illustrative example of this is election forecasting using socioeconomic indicators coupled with polling data. Such models also draw on historical data to forecast the implications of these outcomes. These models are susceptible to fluctuations in input data, which is why they tend to perform in a shorter time window, such as weeks, as opposed to months or years.[15]

3. **The prediction of specific location of events:** One of the key questions analysts are trying to address is the exact location and timing of a particular event. Improvements in data collection, such as imagery data coupled with social media and traditional media information, have led to improvements in predicting the geolocation of political activity. An everyday use of such models is giving better surveys of where troops are located in order to make assessments about where they might go next. One of the key benefits of such models is that with improved data quality, they allow analysts to speak to subnational events of interest. These models are also increasingly being deployed to address geopolitical tail risk– rare but highly impactful events– but they come at the expense of time. These models have the shortest window of time – often kept to a week or less – and are currently deployed on very targeted geographic areas of interest, such as the war in Ukraine and the Taiwan Strait.[16] The overall benefit of using AI for all three model types is clear. The explosion of Big Data, especially through web scraping and language translation, has expanded the capabilities beyond what a human analyst could manually do. In addition, satellite data and mobility tracking can help analysts detect micro-level changes that precede more significant instability events. The increasing amount of data supports the improved detection of anomalies,[17] time forecasting with more replicability, and high-resolution monitoring.[18] One of the conventional criticisms of global coverage models is that important nuances between cultural contexts might get lost, for example, how the Communist legacies of China, Russia, and North Korea impact them differently.[19] However, as these models have become increasingly sophisticated, it has become possible to account for as much nuance as data is available and for analysts to extract specific variables that are driving changes. In short, analysts do not need to know in advance what indicators to feed into the model but can allow these models to help them understand which indicators to prioritize. In addition, as data at the subnational level

[15] Author interview with industry participant (November 2024), Author interview with academic participant (October 2024); Author interview with academic participant (October 2024); Author interview with academic participant (October 2024).

[16] Author interview with industry participant (September 2024); Author interview with industry participant (October 2024).

[17] Author interview with industry participant (October 2024); Author interview with academic participant (October 2024).

[18] Author interview with industry participant (November 2024).

[19] Author interview with academic participant (October 2024).

gets better, the model types that have been traditionally focused on country-level outputs could give more refined outputs.[20]

MODEL OUTPUT	TIME WINDOW
Political instability score	Years
Uncertain outcome of certain event	Months
Location of events	Days to weeks

Another promising avenue of development is the use and incorporation of generative AI (GenAI) in this field, especially with scenario generation.[21] As mentioned with intelligence failure, one of the key challenges for analysts is understanding what to do with the detection of an anomaly, especially if an existing mindset stands counter to the new data. GenAI is less susceptible to the same cognitive biases of humans and there is extensive work being done to reduce the impact of hallucinations.[22] The incorporation of GenAI in these models could help analysts with scenario development and hypothesis testing and even policymakers with decision support to ensure users know how to contextualize and act on the outputs.

Use by an Intelligence Analyst

The models focus on providing assessments of geopolitical outcomes, but how then do analysts turn that into strategic warning? When analysts have the capacity to deliver quicker, more comprehensive assessments about events of interest, you move from the tactical to the strategic. Some interviewees highlighted the goal of strategic warning, noting that analysts need to be able to inform policymakers when our strategic landscape has changed rather than simply predicting a specific event's time or place.[23] In its immediate use case, analysts can

[20] Author interview with industry participant (November 2024).

[21] Author interview with industry participant (November 2024); Author interview with academic participant (October 2024); Author interview (2) with industry participant (October 2024); Author interview with industry participant (October 2024).

[22] Sebastian Farquhar, et al., Detecting Hallucinations in Large Language Models Using Semantic Entropy, Nature 630 at 625-630 (2024).

[23] Author interview with academic participant (October 2024).

leverage AI in the collection and synthesis of data[24] to highlight anomalies that need more attention.

Some interviewees were deeply skeptical that AI would ever make accurate predictions of conflict years in advance or actually be able to identify the moment a conflict would break out[25] because of two key challenges: data limitations and the challenge of predicting leaders' decision-making.

As a result, outputs are likely to be a source of data as opposed to finished intelligence,[26] but there can be an upside as it leverages many pieces of information that a human analyst would otherwise manually input while moving faster. Human analysts will always be required to validate information and demonstrate knowledge of the provenance of information,[27] but if humans can play a role later in the process rather than having to have touchpoints throughout, it can free up personnel to achieve other goals.

Challenge 1: Data Scarcity, Inconsistency, and Fusion

AI systems designed for geopolitical forecasting face a range of technical challenges, many of which stem from fragmented data infrastructures and persistent human biases. Take the example of weather forecasting. One of the reasons weather forecasting is much easier to do than geopolitical forecasting is because of the rich data available. Even the rare events in weather modeling, like cyclones, are merely extensions of normal storms about which much is known.[28] Rare geopolitical events are much more nuanced, and far less linear relationships drive them. In some relevant regions, which could nevertheless have enormous conflict escalation repercussions, the U.S. and UK IC may not be gathering data at all. Very little data on factors that protect from risks or that stabilize after a crisis is also systematically gathered with AI training data in mind.[29]

Overreliance on limited or historical datasets can lead to overfitting and missed anomalies,[30] especially as most publicly available models lack access to sensitive or classified information. These constraints are compounded by short forecasting windows, meaning that recent geopolitical shifts may not be reflected in training data. In addition, many existing tools fail to capture key social or inter-state factors, often relying on incomplete or outdated information that can reinforce analyst biases.

[24] Author interview (2) with industry participant (October 2024); Author interview with academic participant (October 2024); Author interview with academic participant (October 2024).

[25] Author interview with industry participant (October 2024); Author interview with academic participant (October 2024).

[26] Author interview with industry participant (September 2024).

[27] Author interview with industry participant (October 2024); Author interview with academic participant (October 2024).

[28] Author interview with academic participant (October 2024).

[29] CETaS workshop (September 2024).

[30] Scotty Black & Christian Darken, Scaling Artificial Intelligence for Digital Wargaming in Support of Decision-Making, NATO (2024).

Evaluating and comparing different systems also remains challenging due to inconsistent performance metrics.[31] The more disparate datasets are integrated, the more likely we are to encounter traceability, AI security, and data validation challenges as different models are layered with increasing opacity regarding the assumptions and limitations behind different models.

Broader problems—such as data governance, security, and affordability[32]—further hinder the creation of integrated data infrastructures. Lack of suitable data to train AI models is particularly common in organizations applying AI for the first time or to a new use case.[33] Data on conflict or instability risk is not collected with AI training in mind, so data engineers may have issues with the quality of testing data.

Challenge 2: Modeling the Decisions of Individuals

Predictions of specific "trigger" events will always be challenging, especially for events in which the element of surprise is paramount.[34] For example, you might be able to assess the overall likelihood of a coup or an international war occurring based on overall risk factors coupled with factors that suggest an event could be imminent (i.e., troop locations) but face difficulty in predicting the exact event for two reasons. First, although the CIA has made attempts to use AI to help its analysts better understand and anticipate leadership decision-making,[35] it is challenging to get into the head of a leader ideally, even with very good, classified information, although digital twins could get us very close. Second, it is not possible to ideally account for random,[36] impulsive, or opportunistic decisions by individuals. This could also include influential factors such as dissenting voices in a government, advisers for a state or military leader, or psychological instability. Individuals with deceptive intentions are also difficult to capture as training data for a model. Furthermore, the decisions of individuals who may not appear sufficiently high priority to track but who could act in ways that trigger regional stability (e.g., the self-immolating fruit vendor who triggered The Arab Spring) also cannot feasibly be captured in geopolitical modeling tools. Therefore, while an AI model can identify when a situation or country is highly volatile, it will face challenges predicting the exact moment an event of interest will occur.

[31] Mayank Kejriwal, Link Prediction Between Structured Geopolitical Events: Models and Experiments, Frontiers in Big Data at 860-896 (2021).
[32] Roshanak Rose Nilchiani, et al., Joint All-Domain Command and Control (JADC2) Opportunities on the Horizon, Acquisition Research Program (2023).
[33] James Ryseff, et. al., The Root Causes of Failure for Artificial Intelligence Projects and How They Can Succeed, The RAND Corporation (2024).
[34] Author interview with academic participant (October 2024); Author interview with academic participant (October 2024).
[35] Scott Nover, Can the CIA's AI Chatbot Get Inside the Minds of World Leaders?, GZeroAI (2025).
[36] Author interview with industry participant (October 2024); Author interview with academic participant (October 2024); Benjamin Jones & Benjamin A. Olken, Hit or Miss? The Effect of Assassinations on Institutions and War, American Economic Journal: Macroeconomics at 55–87 (2009).

Despite these challenges, it is essential to remember these technologies are rapidly maturing. Next-generation AI tools will not only monitor broader swaths of activity to detect anomalies early but also generate new scenarios that analysts may not have considered, leading to more comprehensive and nuanced threat assessments. Most importantly, it offers a time advantage: even a brief earlier alert can be the difference between proactive action and reactive response – an edge that becomes more vital as adversaries also invest heavily in AI. The growing involvement of commercial innovators also means intelligence services need not build everything from scratch, allowing them to integrate and fine-tune AI systems more quickly. Over the short term, AI can augment human analysts as an additional source of intelligence, and in the longer term, this human-machine teaming will steadily refine warning mechanisms, producing faster, richer, and more reliable insights for decision makers.

Chapter 2: What is the Path Ahead?

The previous chapter covered the landscape of current tools utilizing AI for strategic warning in the present day. This chapter looks to the future and presents the study team's analysis of the path toward AI-enabled strategic warning tools that will bring the most benefit to the IC.

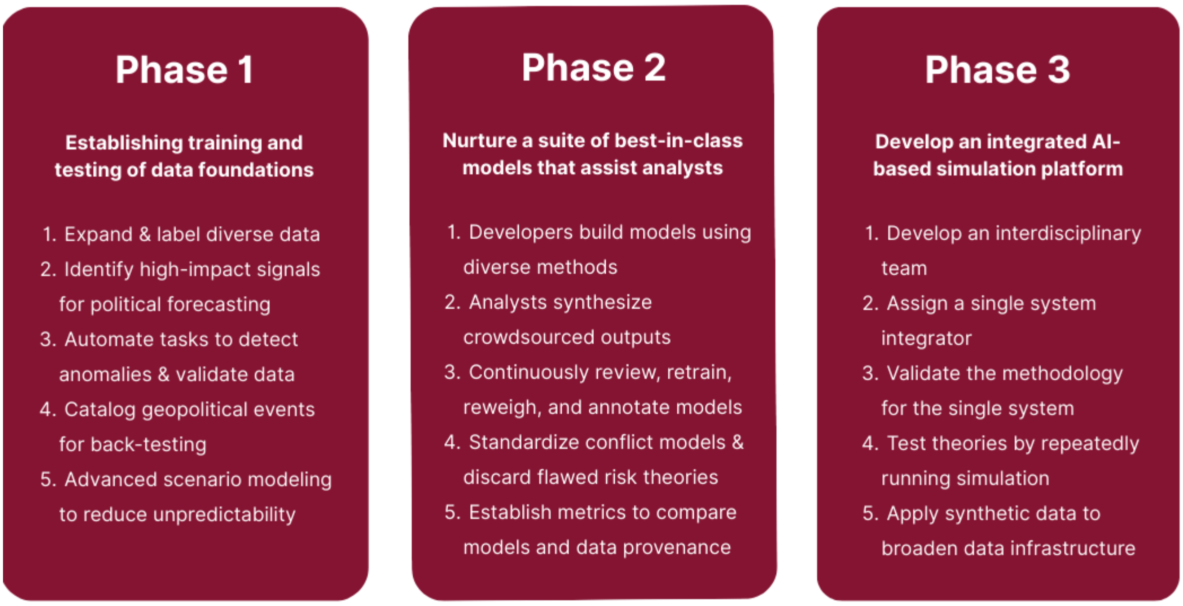
AI may never be 100 percent accurate in its predictions of future events because human behavior is inherently difficult to model, but we can significantly improve the quality and accuracy of the outputs and overall forecasting capabilities of an AI-enabled strategic warning system and secure more time for policymakers to respond to crises. Instead of predicting the time and place where geopolitical flashpoints will occur, experts suggested there would be great utility in helping analysts track conflict risk indicators more accurately by leveraging more data, as well as more rapidly helping analysts identify the possible outcomes of a shock immediately after it happens.[37] Essentially, we can track the tinder and the kindling and predict the path of a forest fire without predicting exactly when the match will strike.

It might be tempting to jump to the idea of a single super-simulation of all the possible factors that could affect conflict and instability or the reverse. However, the simple reality is that while there are some factors that essentially predict conflict across the globe, the factors that ultimately lead to the outbreak of violence in a particular state or region are highly nuanced.

Improvements in conflict modeling tools could be pursued in phases to first address low-hanging challenges while building confidence and setting the stage for a more ambitious project that would work towards a model that comprehensively simulates all possible instability risk and stabilization indicators. This project, broken down in a phased approach,

[37] Author interview (2) with government participant (December 2024).

will enable the IC to make the most of existing models pragmatically and to start small and scale responsibly without being delayed by technical bottlenecks and costs.



Phase 1

Establishing training and testing of data foundations

1. Expand & label diverse data
2. Identify high-impact signals for political forecasting
3. Automate tasks to detect anomalies & validate data
4. Catalog geopolitical events for back-testing
5. Advanced scenario modeling to reduce unpredictability

Phase 2

Nurture a suite of best-in-class models that assist analysts

1. Developers build models using diverse methods
2. Analysts synthesize crowdsourced outputs
3. Continuously review, retrain, reweigh, and annotate models
4. Standardize conflict models & discard flawed risk theories
5. Establish metrics to compare models and data provenance

Phase 3

Develop an integrated AI-based simulation platform

1. Develop an interdisciplinary team
2. Assign a single system integrator
3. Validate the methodology for the single system
4. Test theories by repeatedly running simulation
5. Apply synthetic data to broaden data infrastructure

Phase 1 – Establish Training and Testing of Data Foundations

In Phase 1, the first priority will be to set solid data foundations by filling in missing conflict warning data by collecting them in priority countries where this does not yet occur and systematizing data collection for future AI systems for strategic warning.[38] There are three categories of data of relevance:

1. **Available measured data.** This includes data we already collect (e.g. weather data, stock indices, passive signal emissions),[39] but there is also a need to supplement available measured data by systematically labeling current conflict indicators, as well as historical conflict triggers and tipping points. Automating routine, low-value bureaucratic tasks and obtaining available measured data (e.g. geospatial data, mobile phone data, and signals emissions) on the conditions of each conflict of interest will be important to eventually free the analyst's time for sense-making.[40] Leveraging OSINT to create time maps of unit movements could help analysts more closely investigate actual doctrine.[41]

2. **Quantifiable non-traditional data.** This could include a wide gamut of non-traditional data sources (e.g. hacker data like the Discord Leak, where allegedly leaked classified documents were released on the Discord platform[42]); humanitarian health, famine,

[38] Author interview with industry participant (November 2024); Author interview with industry participant (October 2024).
[39] Author interview with industry participant (October 2024).
[40] Author interview with academic participant (October 2024).
[41] Author interview with academic participant (October 2024).
[42] Shane Harris & Samuel Oakford, This Discord Leaks, Explained, The Washington Post (2023).

and economic data (which could build a picture of swelling grievances) lifestyle apps (e.g. dating app data, which can be used to map troop movements).[43]

3. **Mental model data.** This includes data on key decision makers' and groups' beliefs, biases, emotional responses, and reasoning patterns, especially from multilingual sources within the Axis of Disruptors.[44] This can also include more systematic capture of anthropological or biographical knowledge for the explicit purpose of developing AI training data sets.

Most repositories currently contain only a few dozen data points in their catalogs, with just dozens or hundreds of samples at best. Ideally, collecting comprehensive data on tens of thousands of conflict and instability events[45] —including the parameters mentioned—would help establish a standard conflict model, outlining assumptions, rules, and strategic choices for relevant actors and enabling predictive insights:

- Conditions: the individual characteristics necessary to describe a region, people, group, leader, or a set of key relationships.

- Trends: how those conditions change dynamically over time.

- System state: a summary combination of conditions and trends sufficient to describe the state of the region, people, group, leader, or set of relationships.

- Triggers: events that may result in system state changes, from stability to instability or into conflict (e.g. a protest).

- Tipping Points: when conditions and trends favor significant destabilization (e.g. a protest that leads to the Arab Spring).

- Scenarios: forecasts of what might happen after a trigger, whether it results in a tipping point, and what impacts various policies may have on the future.

As part of this, there is a need to continuously identify novel trends that challenge assumptions on the conflict risk indicators.[46]

A critical technical challenge to developing solid data foundations will be the ability to validate data and identify intentionally hidden, poisoned, or corrupted data.[47] This upfront investment will help address output and system validation challenges later down the line, including

[43] Simon Newton, Tinder Trap: Ukraine and Russia Using Fake Profiles to Trick Soldiers Into Revealing Intel, BFBS Forces News (2024).
[44] Author interview (2) with industry participant (October 2024); Author interview with industry participant (October 2024).
[45] Author interview with academic participant (October 2024); Author interview with academic participant (October 2024).
[46] Author interview with government participant (October 2024).
[47] Author interview with academic participant (October 2024).

whether the dataset supports the development of a standard model of conflict. A standard model of conflict would explain the social science of conflict in the same way we can explain why atoms and molecules in physics work with high confidence.

Part of this phase also involves overcoming data-sharing challenges across intra and inter-governmental agencies, as well as across the public and private sectors. Data classifications, proprietary information or data privacy regulations are no trivial challenges and will continue to inhibit the IC from exploiting data-driven capabilities if they are not considered from the onset of any AI-based capability development effort.[48]

Phase 1 will enable the IC to start to build muscle memory on how to build the data infrastructure and the analyst familiarity with these systems[49] needed for the successive phases.

Phase 2 – Nurture a Suite of Best-in-Class Models that Assist Analysts

This phase describes a suite of large and small, controlled, fast, and frugal models that assist analysts with forecasting conflict risks and stabilization opportunities.[50] The outputs of several models ingesting different types of data, which show a different understanding of how the world works, could be crowdsourced to triangulate perspectives and contribute towards mitigating bias.[51] The third-party or government developer of each model could develop models using different methodologies and applying different types of AI to augment each stage of the intelligence collection and analysis cycle however they know works.[52] For example, an ML-based model could model the economic forces that affect conflicts globally, and a different DL-based model could look at links between geopolitical events on a global scale.[53] Both government and third-party developers would develop system cards outlining the limitations, assumptions, and assurance cases for each model. In this Phase, the models do not have to be accurate; they simply have to be thought-provoking for the analyst.[54]

Mega models leveraging open-source internet data over large geographic areas could be built by industry suppliers.[55] For example, different vendors could develop separate models mining open-source data and covering slow-building dissent and grievances in China, Russia, Iran, or the Middle East and North Africa, while another vendor could develop a model mining open-source conflict risk data in Southeast Asia. Models that track intergroup conflicts and

[48] Author interview with academic participant (October 2024).
[49] Author interview with academic participant (October 2024).
[50] Author interview with (2) government participants, December 2024; Daniel M. Benjamin, et al., Hybrid Forecasting of Geopolitical Events, AI Magazine at 112-128 (2023).
[51] Author interview with academic participant (October 2024).
[52] Author interview with academic participant (October 2024).
[53] Mayank Kejriwal, Link Prediction between Structured Geopolitical Events: Models and Experiments, Frontiers in Big Data at 860-896 (2021).
[54] Author interview with academic participant (October 2024).
[55] Author interview with academic participant (October 2024).

opinion dynamics[56] (e.g. between government forces, foreign fighters, and private security in Ukraine) could also yield new insight. Government agencies could develop in-house models mining official-sensitive data on smaller areas and routes, secret data on specific groups, and top secret data on specific individuals.[57]

Appropriately cleared human analysts could then synthesize the crowd-sourced outputs in addition to traditional intelligence artifacts, developing a better understanding of the conflict picture. This would enable the analyst to do what humans are best at—human reasoning—while allowing the analyst to take more data into account with the help of lots of different AI. AI is proficient at understanding patterns, but the state of the art in AI can overlook important nuance and contextual elements. For example, food subsidies in less developed countries are essential instability risk factors, but it is not equally important in Nigeria as it is in Egypt.[58] In Egypt, government subsidies on pita bread are more relevant to instability risk, whereas oil is more relevant in Nigeria. Maintaining human-machine teaming is also necessary because current AI-based systems have a tendency to fight the last war that it trained on, and human expertise is needed to balance some indicators and warnings and detect potential overfitting of data.[59]

Identifying ways that the different model outputs could be validated is also essential. One way of doing this would be to see if the model could have predicted the progress of 9/11, Pearl Harbor, or the Arab Spring given the same information as the analysts had before the outbreak of each crisis.[60] Operators could also compare the results of the model with those of an analyst who has applied structured analytic techniques.[61]

The family of large and small models would be continuously reviewed, retrained, reweighted, and annotated overnight with a human in the loop during several cycles until the final models are better than the original models.[62] The cascading models should also feed both bottom-up and top-down feedback signals to continuously refine weights in each model.[63]

Over time, Phase 2 could help the IC develop a standard model of conflict. The suite of models would instrument existing theories and, over time, eliminate theories that are incorrect.

Experts discussed that middle decision-makers in the defense and security sector could be very old-fashioned and resistant to the adoption of AI,[64] so having a suite of models that are continuously improved could also contribute to changing this risk-averse culture. In addition,

[56] Mayank Kejriwal, Link Prediction between Structured Geopolitical Events: Models and Experiments, Frontiers in Big Data at 860-896 (2021).
[57] Author interview with academic participant (October 2024).
[58] Author interview with academic participant (October 2024).
[59] Author interview with academic participant (October 2024).
[60] Author interview with government participant (December 2024); Author interview with government participant (December 2024).
[61] Author interview with academic participant (October 2024).
[62] Author interview with academic participant (October 2024).
[63] Author interview with academic participant (October 2024).
[64] Author interview with academic participant (October 2024).

focusing on a multitude of capabilities could stem from decision paralysis regarding how to prioritize a specific capability. Maintenance and progress reports on each model could increase confidence in AI over time and help more users understand its strengths and limitations. For example, some models may be better at predicting positive rather than negative changes.[65] Performance metrics for each model[66] could further increase confidence.

In Phase 2, data collection should focus on "filling in dark spots on the map" left over by Phase 1[67] and will be driven by the need to test theories for inclusion or rejection in a standard model of conflict. This may require fusing data streams collected in Phase 1, combining available measured data, quantifiable non-traditional data, and mental model data, as well as enabling the collection of new data suggested by novel hypotheses that have emerged.

Another key breakthrough needed in this Phase, is in developing metrics to compare different models and dataset provenance to be able to grade and communicate dataset quality.[68]

EXAMPLES OF MODELS			
Third-party mega-model mining open-source internet data on a large geographic region	Government model mining official-sensitive data on specific smaller areas and routes	Government model mining secret data on specific groups	Government model mining top secret data on specific individuals

EXAMPLES OF MODELS USED FOR SPECIFIC STRATEGIC WARNING COMPONENTS			
Intelligence Collection	Data Processing	Batch Analysis	Analysis of Selected Data
Graph Neural Networks Foundation Models Few Shot Learning	Large Language Models Machine Learning	Large Language Models Deep Reinforcement Learning Chain of Thought Reasoning	Deep Reinforcement Learning

[65] Paolo Vesco, et al., United They Stand: Findings from an Escalation Prediction Competition, International Interactions at 860-896 (2022).
[66] Paolo Vesco, et al., United They Stand: Findings from an Escalation Prediction Competition, International Interactions 860-896 (2022).
[67] Author interview with academic participant (October 2024).
[68] Insights from roundtable discussion at Centre for Emerging Technology and Security workshop (September 2024).

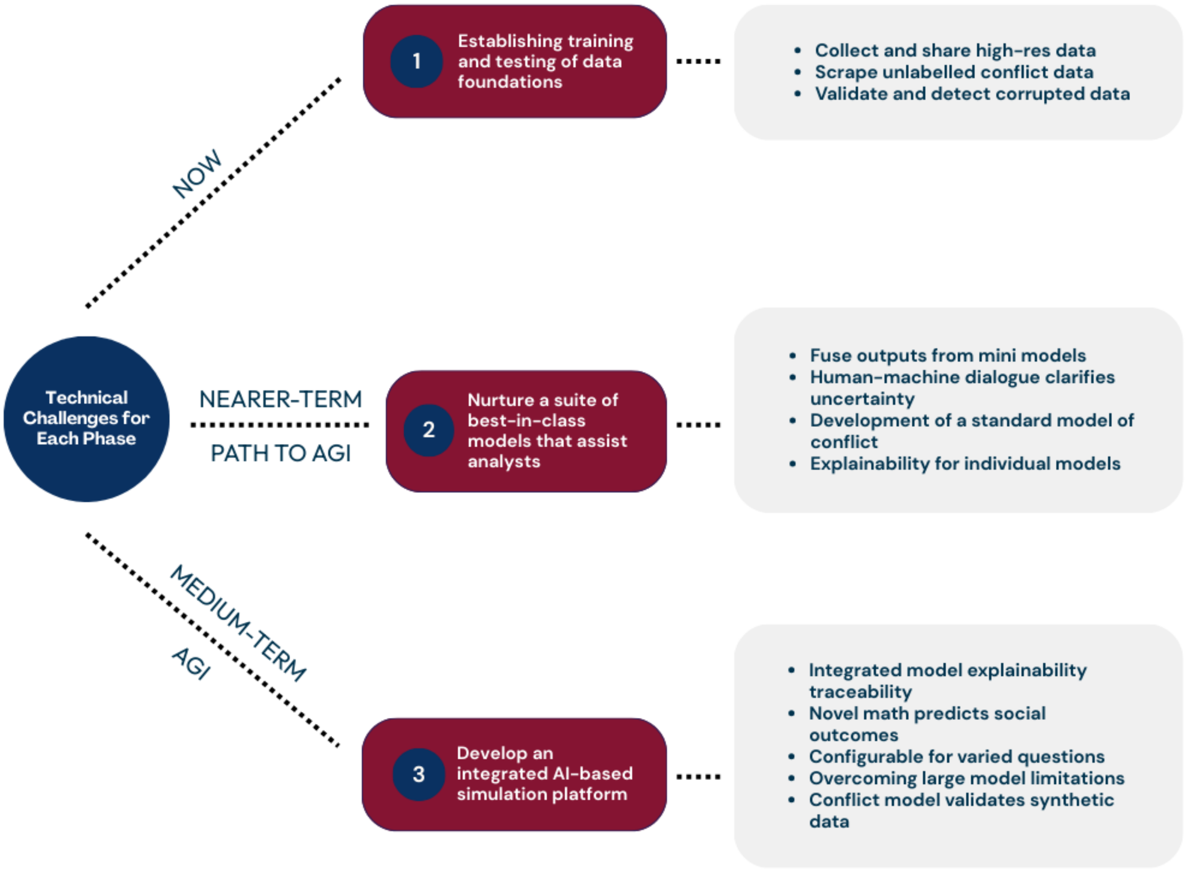
A simplified visualization of Phase 2 is presented in the tables above. Across these models and component capabilities, a human analyst triangulates and analyzes the results of a suite of models in addition to traditional manual intelligence analysis.

It is possible that adversaries may waste time and resources chasing the unachievable, and a pragmatic approach that enables the U.S. and UK ICs to spend valuable resources on more proven methods could generate an advantage. Phase 2 would allow the U.S. and UK ICs to test models, theories, and datasets against one another, both to identify bad ones and to improve confidence in analysis.

Phase 3 – Develop an Integrated AI-based Platform

This Phase describes an integrated AI for strategic warning simulation platform built by a single system integrator, which could be used to run and test different hypotheses about conflict.[69] This Phase will only be possible after significant developments in agentic AI to pursue complex tasks and after a standard model of conflict is developed and will likely require considerable investment in interdisciplinary technical and social science research and development to overcome the fundamental challenges.

[69] Author interview with academic participant (October 2024).


Technical Challenges for Each Phase
NOW
NEARER-TERM
PATH TO AGI
MEDIUM-TERM
AGI
1
Establishing training and testing of data foundations
Collect and share high-res data
Scrape unlabelled conflict data
Validate and detect corrupted data
2
Nurture a suite of best-in-class models that assist analysts
Fuse outputs from mini models
Human-machine dialogue clarifies uncertainty
Development of a standard model of conflict
Explainability for individual models
3
Develop an integrated AI-based simulation platform
Integrated model explainability traceability
Novel math predicts social outcomes
Configurable for varied questions
Overcoming large model limitations
Conflict model validates synthetic data

In Phase 2, narrow and siloed models would be generating outputs, but if we expect AI to be able to fuse all the information we know about the conflict, then we not only require better data but better math to establish causality between actors' decisions, actions, and their outcomes.[70] The rewards would likely be considerable, and an integrated model may be able to elucidate the inter-relationships between conflict indicators.

Phase 3 will likely require more globally and sectorally comprehensive and longitudinal data and would need to run the data tens of thousands of times in an emulation where we can hypothesize potential outcomes based on known patterns.[71] Phase 3 will also require more hardware and data sharing agreements between industry and government, as well as enormous compute resources.[72]

A method is also needed to validate the Phase 3 system's methodology because, in addition to validating individual models' outputs, integrating diverse data—each with its own assumptions, limitations, and conceptual frameworks—complicates traceability and explainability.[73] One method of achieving this would be to test models for conflicts that

[70] Author interview with academic participant (October 2024).

[71] Author interview with academic participant (October 2024); Author interview with industry participant (October 2024).

[72] Author interview with academic participant (October 2024); Author interview with academic participant (October 2024).

[73] Author interview (2) with government participants (December 2024).

ultimately did occur or did not by drawing on data up until that point. For this reason, smaller, more focused models are architecturally the better choice. There are few human interactions more complex than military aggressions, and in such simulations, models typically face a trade-off between the complexity of the parameters and the performance of the AI.

Developers must understand whether the corpus of human knowledge already possesses the correct theories on what is happening in the AI system and what is happening in reality.[74] At the tool development and testing stages, it will be essential to ensure it is clear if the model is focusing on the correct relationships and that the multiple streams of information with different pedigrees have been weighed appropriately.[75] To mitigate bias, developers must also assess if the single, integrated model is trained on data that is representative of all the demographic cohorts in the world.[76]

It remains to be seen if a single integrated system can be as flexible to several types of possible questions analysts may wish to answer or patterns they may want to focus on.[77] Temporally, models should operate at different time horizons towards the past and the future at 6-month, 1-year, 10-year, and 20-year time horizons. Strategic warning tools can develop a valid methodology, but analysts and senior decision-makers need to be able to answer different questions as the strategic landscape develops.

With a standard conflict model in hand, the dramatic development in Phase 3 is the application of synthetic data to broaden the training set beyond the historical data available.

Modeling techniques in Phase 3 are more similar to frontier model development in less complex areas. Large models of tens or hundreds of billions of parameters trained off the optimized datasets built throughout Phase 1-2 and synthetically generated in Phase 3 will radically advance AI for strategic warning.

Chapter 3: The Price and Payoff of Applying AI to Strategic Warning

At the beginning of the study, the team set out to develop a cost-benefit analysis of either ultimately building an AI-based strategic warning capability within government from the ground up, procuring an industry solution, or maintaining the human-led process. As the study progressed, it became apparent that it would take a moonshot effort carried out in phases to develop the data infrastructure that would bring significant performance enhancement in strategic warning and that a mixture of these elements could appear in each phase.

[74] Author interview with academic participant (October 2024).
[75] Author interview with government participant (October 2024); Giuseppe Nebbione, Deep Neural Ranking for Crowdsourced Geopolitical Event Forecasting, Computer Science and Engineering (2019).
[76] Author interview with academic participant (October 2024).
[77] Author interview with government participant (October 2024).

Traditional cost-benefit analysis in a field where the technology is not yet developed is challenging, given the difficulty of quantifying several unknown variables and limited publicly available information on analogous AI-based capability costs. For example, investments in data infrastructure would benefit not just AI for strategic warning systems but many other analytic priorities as well. Yet, it is possible to identify quantifiable rough order of magnitude (ROM) costs and non-quantifiable costs that should be considered.

Assessing the Financial Costs

If a government customer is simply procuring, the cost of the system would be the price and the subscription to the system, but if the system is being developed from the bottom up, there are significantly more investment costs to consider. This includes direct, indirect, start-up, sustainment, procurement, salary, and benefits costs.[78] This includes IT infrastructure (e.g., computing hardware, data storage, servers, bandwidth and connectivity, cloud-based services, physical infrastructure), personnel, as well as data.[79] A very rough order of magnitude estimate of data acquisition can range from annual licenses of $5,000 - $500,000 per year.[80] This might involve some configuration of simulations, models, and data, as well as the costs of data rights or intellectual property.[81] The granularity of data labeling activities will also affect how cost-intensive this endeavor is. The IT infrastructure to train foundational models on acquired data is becoming more expensive,[82] and there are other tools to consider (e.g., digital engineering tools, workflow tools, translators, graphics emulators, etc.).[83]

As illustrated in the data capture requirements in the previous chapter, an ambitious effort to enable AI for strategic warning is not just about procuring an off-the-shelf narrow AI-based tool but addressing systemic data challenges. This effort would require research and development at a scale similar to that of frontier AI efforts. This means resourcing the effort with strategic-level funding is essential. Many leaders are not prepared for the cost and time it takes to acquire, structure, and explore their own organization's datasets and expect AI adoption to be possible in weeks instead of months.[84]

Industry leaders estimate the costs of foundational frontier models that are tens to hundreds of millions of dollars now[85] may be $10 billion by 2026 and $100 billion clusters in 2027, and

[78] U.S. Army Cost Benefit Analysis Guide, Office of the Deputy Assistant Secretary of the Army (2018).
[79] N. Peter Whitehead, et al., A Framework for Assessing the Costs and Benefits of Digital Engineering: A Systems Approach, RAND Corporation (2023).
[80] Eugenio Caterino, What Is AI Training Data? Examples, Datasets and Providers, Datatrade (last accessed 2025).
[81] N. Peter Whitehead, et al. A Framework for Assessing the Costs and Benefits of Digital Engineering: A Systems Approach, RAND Corporation (2023).
[82] Gaël Varoquaux, et al., Hype, Sustainability, and the Price of the Bigger-Is-Better Paradigm in AI, ArXiv (2024).
[83] N. Peter Whitehead, et al., A Framework for Assessing the Costs and Benefits of Digital Engineering: A Systems Approach, RAND Corporation (2023).
[84] N. Peter Whitehead, et al., A Framework for Assessing the Costs and Benefits of Digital Engineering: A Systems Approach, RAND Corporation (2023).
[85] Dylan Patel & Nathan Lambert, DeepSeek, China, OpenAI, NVIDIA, xAI, TSMC, Stargate and AI Megaclusters, interview by Lex Fridman, Lex Fridman Podcast (2025).

even those may not be sufficient for capabilities envisioned by Phase 3 Strategic Warning AI.[86] Task- specific fit-for-purpose models with less than 10 billion parameters as envisioned by Phase 2 can take up to 60,000 kWh to train and fine-tune[87] while multi-purpose frontier models with hundreds of billions of parameters, as envisioned in Phase 3, barring energy saving innovations may exceed 1,500-2,000 mWh.[88] Current plans for some data centers are even estimated to be as high as 2.8 gWh.[89] An AI-based capability is particularly costly because the high algorithm complexity and upfront capital requirements make them unlike a typical IT project.[90] While these costs are undeniably high, as much of the development is already happening within the private sector, there are significant opportunities for the ICs to leverage that development, especially as a Phase 3 Strategic Warning system is a component of a more substantial trajectory to AGI.

Once the models are built, inference energy costs and the energy used per inquiry will become more efficient. Moreover, given the far fewer likely inquiries in an AI-based strategic warning capability versus a commercial, public model, over time, the costs per inquiry will be incremental to the overall costs.[91] However, as inference complexity increases—from generative chat to reasoning to autonomous tasking, the inference energy costs increase by an order of magnitude at each step.[92]

The suite of digital tools necessary to manage the training, fine-tuning, and other efforts around the models are incremental to the models' costs but vital to get right. This is why a diversity of providers is important in Phases 1 and 2 rather than picking one system and forcing all providers to work within the same framework. As organizational muscle memory improves in Phases 1 and 2, the knowledge of what digital tools primary integrators need to have for success in Phase 3 should emerge and solidify as industry best practices. There may also be cost and commercial challenges in concurrently acquiring data from several models and suppliers. Furthermore, given how sensitive these models are, there will be costs associated with storing them in air-gapped systems to prevent nefarious actors from achieving access.

The workforce needed to staff an AI Strategic Warning capability goes beyond the analysts who use it. Given the effort it will take to reset the way the IC captures and prepares data for AI, as well as the variety of different roles that will be needed, the costs are more likely to be analogous to when the United States established Space Command or the National

[86] Dario Amodei, Anthropic CEO on Claude, AGI & the Future of AI & Humanity, interview by Lex Fridman, Lex Fridman Podcast (2024).
[87] Alexandra Sasha Luccioni, et al., Power Hungry Processing: Watts Driving the Cost of AI Deployment, ArXiv (2024).
[88] Alex De Vries, The Growing Energy Footprint of Artificial Intelligence, Joule at 2191–2194 (2023).
[89] Dylan Patel & Nathan Lambert, DeepSeek, China, OpenAI, NVIDIA, xAI, TSMC, Stargate and AI Megaclusters, interview by Lex Fridman, Lex Fridman Podcast (2025).
[90] James Ryseff, et al., The Root Causes of Failure for AI Projects and How They Can Succeed, RAND Corporation (2024).
[91] Alexandra Sasha Luccioni, et al., Power Hungry Processing: Watts Driving the Cost of AI Deployment, ArXiv (2024).
[92] Dylan Patel & Nathan Lambert, DeepSeek, China, OpenAI, NVIDIA, xAI, TSMC, Stargate and AI Megaclusters, interview by Lex Fridman, Lex Fridman Podcast (2025).

Counterterrorism Center requiring entire new career paths, management structures, and facilities.[93] A common concern is the possibility of relegating talented and skilled analysts to the role of data labelers. This could happen if the appropriate workforce requirements are not considered for data scientists who understand AI, suitably certified assurance experts who specialize in AI for strategic warning, and others. This workforce will also need appropriate workforce training and development.

Using estimates from a Congressional Budget Office (CBO) 2019 analysis in 2020 for Space Command new hires and facilities,[94] the costs of Phase 1 could be similar to a Policy Directorate in Phase 1 (40-300 people, <$10 million start-up funding and $10-60 million annually).

Evolving into Phase 2, with its increased focus on governance of managing a tiered ecosystem of models, costs may look like a new Development & Acquisition Agency in size and composition (1,200-1,300 people, $220-$560m startup and $240-$460m annually).

As Phase 3 capability matures into a single or few primary integrators with more advanced AI capabilities, personnel needs may decrease, and when combined with other AI capabilities, look more like a Combatant Command (400-600 people, $520-$1,060m startup & $80-$120m annual). If the total workforce that requires government funding, military and civilian, ends up being substantially larger, such as a Military Department (5,400 - 7,400, $1,400-$3,240M startup & $1,080-$1,540 annually) or workforce increases similar to those estimated to be required to support digital engineering of ~40,000 personnel can reach costs of $500-$800m startup and $500m annually.[95]

Non-Quantifiable Costs

Opportunity Costs of Not Adopting AI

It will not be possible to have good coverage on conflict risk indicators without AI, and the race for AI for strategic warning might set conditions for the next several decades of international affairs. At stake is not simply that adversaries might use AI capabilities to out-think us, but also out-imagining us and creating an environment in which the rules of engagement remain undefined and ever-changing.[96]

In January 2025, Chinese hedge fund High-Flyer AI announced the DeepSeek-V3 foundational training model followed shortly after by the DeepSeek-R1 reasoning model that

[93] Author interview with academic participant (October 2024); Author interview with academic participant (October 2024); Author interview with academic participant (October 2024); History, The National Counterterrorism Center (last accessed 2024).

[94] The Personnel Requirements and Costs of New Military Space Organizations, Congressional Budget Office at 2 (2019).

[95] N. Peter Whitehead, et al., A Framework for Assessing the Costs and Benefits of Digital Engineering: A Systems Approach, RAND Corporation (2023).

[96] Author interview with academic participant (October 2024).

stood "toe-to-toe with the best OpenAI, Google, and Anthropic."[97] Not only did this represent a large leap in performance, but DeepSeek was allegedly developed for a fraction of the cost ($5.5M) on export-controlled less powerful GPUs than are available to U.S. firms.[98] In contrast, in the same week, a coalition including OpenAI, Oracle, and the UAE MGX announced a $500B investment in additional U.S. data centers named "Stargate."[99] A further concern is that within a week of the DeepSeek moment, Alibaba announced its own frontier reasoning model, "Moonshot," which is competitive with OpenAI o1 released in September, behind the performance of o3 but still a remarkable feat. Behind these two are a host of other Chinese firms looking for similar gains.[100] There are some doubts if the claimed costs represent true costs, and it is unclear if High Flyer, having developed DeepSeek, has the compute power to implement it broadly,[101]. Still, regardless, the 'DeepSeek moment' highlighted that Chinese AI firms are closing the distance with U.S. firms in terms of frontier model development, training, and the ability to innovate creatively.

There may also be opportunity costs in terms of cost savings and avoidances if enhanced foresight prevents resource wastage, increases analyst productivity by automating some tasks, as well as reductions in error rates by helping analysts triangulate hypotheses.

There may be non-quantifiable opportunity costs when it comes to enhancing safety outcomes or saving lives, reducing uncertainty, increasing strategic choices, reducing redundancy, and achieving strategic objectives.

Finally, if the U.S. and UK ICs do not pursue AI for strategic warning, they could struggle to understand adversary capabilities and how to counter them. In addition to its commercial sector, China's intelligence apparatus is also rapidly building up its AI capabilities,[102] which means a Chinese AI-based strategic warning system could be on the horizon. Integrating AI tools for high-impact IC purposes is how we keep our competitive advantage, especially as AI becomes part of the threat landscape that necessitates strategic warnings.

Unilateral vs. Multilateral Development Costs

A single government pursuing this ambitious program can fund the dollar costs but may experience challenges harnessing the necessary partners if some technical breakthroughs

[97] Timothy Morgan, How Did DeepSeek Train Its AI Model On A Lot Less – And Crippled – Hardware?, TheNextPlatform (2025).
[98] Timothy Morgan, How Did DeepSeek Train Its AI Model On A Lot Less – And Crippled – Hardware?, TheNextPlatform (2025).
[99] Lucinda Shen, How Is Stargate's $500B Getting Funded?, Axios (2025).
[100] Scott Singer, DeepSeek and Other Chinese Firms Converge with Western Companies on AI Promises, Carnegie Endowment for International Peace (2025).
[101] Dylan Patel & Nathan Lambert, DeepSeek, China, OpenAI, NVIDIA, xAI, TSMC, Stargate and AI Megaclusters, interview by Lex Fridman, Lex Fridman Podcast (2025).
[102] Dylan Patel & Nathan Lambert, DeepSeek, China, OpenAI, NVIDIA, xAI, TSMC, Stargate and AI Megaclusters, interview by Lex Fridman, Lex Fridman Podcast, (2025).

emerge from outside a particular country. There may also be duplicate costs if a nation pursues AI for unilateral strategic warning.

Given that enormous opportunity cost, pooling resources amongst bilateral or Five Eyes allies could reduce the financial burden of the capability for each partner. Still, equally, each partner will have to consider and harmonize different legal and regulatory frameworks and policy requirements. Data access challenges could also be compounded. If the multilateral approach is to be pursued, data sharing and data sovereignty considerations must be addressed upfront.

Preserving Autonomy vs. Vendor Lock-In

It may be easier to maintain existing security hierarchies if most models are built in-house, but this would bring the up-front set-up costs to the government instead of the industry. Furthermore, governments are inhibited by data-sharing constraints and lack of infrastructure.[103] While there are pockets of excellence, government models are not likely to be the center of gravity for the future of AI applied to strategic warning.[104]

Instead, governments are turning to industry to lead research and development to reduce costs, but government involvement at the prototyping stage will remain necessary and essential to keep the work program classified.[105] If the development of technology does not need to be done by someone with a clearance, it is typically more economical to do it outside of government.[106]

It is tempting to 'lock-in' with first-mover companies, hoping to leverage efficiencies of scale by organizing a common platform now. However, the field of AI is evolving quickly, and the nature of the standard model of conflict and where the technical breakthroughs will come from are still too unknown to select a single provider now.

Maintaining a healthy and diverse ecosystem of suppliers could contribute towards preventing the same cost escalation challenges that defense has experienced with traditional defense suppliers for major platforms. It is well documented that reducing competition in the market reduces suppliers' incentives to bear down on costs.[107] Technology companies may charge more for their products if the government does not have alternatives.[108]

[103] Author interview with academic participant (October 2024).
[104] Author interview with government participant (October 2024).
[105] Author interview with academic participant (October 2024).
[106] Author interview (2) with industry participant (October 2024).
[107] Evidence Summary: The Drivers of Defence Cost Inflation, UK Ministry of Defence (2022).
[108] Author interview with academic participant (October 2024); Author interview with industry participant (October 2024).

Chapter 4: Recommendations

SCSP noted in a previous study on AI and intelligence analysis[109] that ICs in the Five Eyes must take a more proactive approach to integrating AI tools into their workflows to avoid losing competitive advantages. This study illustrates one high-impact use case, which can be deployed in several different ways. The IC could use these tools to enhance the monitoring of high-priority regions—like the Axis of Disrupters—in real-time, or it could choose to take a more global coverage perspective. These uses are distinct but not mutually exclusive and speak to the practical application of this system.

Moreover, there are two other clear advantages that the ICs should keep at the forefront. First, while investment into this system is undeniably high, there are positive externalities to be enjoyed. The benefits of this system will not be limited to just analytic units. A system that helps analysts give policymakers insights faster could also be deployed to help collectors give intelligence to analysts just as quickly. In addition, all components of the IC, from those in signals intelligence, imagery, human intelligence, and open-source intelligence, will find advantages to this tool.

Second, it is imperative to think about AI as a tool for strategic warning and as part of the technological landscape for which policymakers need strategic warnings. China's intelligence apparatus[110] is using every tool in its arsenal, including AI, to challenge our ICs, which means a Chinese AI-based strategic warning system could also be under development.

Time is of the essence, but one of the key challenges of moving from the status quo of a manual-based approach to Phase 3 is understanding where to begin. While the financial costs are undeniably high—covering everything from IT systems and data acquisition to specialized personnel and energy consumption—the investment promises far-reaching benefits by transforming raw data into actionable insights of the utmost importance. To propel existing momentum forward, we recommend that the U.S. and UK ICs undertake the following four actions:

1. **Action 1 - Launch a Comprehensive, Large-Scale Effort**
2. **Action 2 - Build Robust, Multinational Partnerships**
3. **Action 3 - Prioritize Human–Machine Teaming with Rigorous Safeguards**
4. **Action 4 - Encourage Policymakers to Leverage Enhanced Decision Advantage**

Action 1 - Launch a Comprehensive, Large-Scale Effort

Many aspects of a strategic warning system—data processing, data cleaning, predictive analytics, and scenario generation—constantly face improvement. Trying to select one

[109] The Future of Intelligence Analysis: U.S.-Australia Project on AI and Human Machine Teaming, Special Competitive Studies Project (2024).
[110] Edward Wong, et al, Chinese Spy Agency Rising to Challenge the C.I.A., The New York Times (2023).

component to focus on while eschewing the others could exacerbate the technological gap and existing vulnerabilities. **Therefore, a concerted, large-scale effort is critical to transitioning from Phase 1 to Phase 2,** ultimately pushing the IC to a possible Phase 3. This moonshot should be a component of larger moonshot efforts towards AGI.[111] Taking such a comprehensive approach comes at a financial cost but provides the necessary structure, resources, and most critically, momentum to accelerate innovation across multiple sectors. By establishing clear milestones and benchmarks within this broader framework, stakeholders can measure progress more effectively and maintain accountability. These explicit goals should reflect technical complexity, ethical considerations, and security requirements, ensuring that every step forward is deliberate, transparent, and aligned with shared values.

Action 2 - Build Robust, Multinational Partnerships

Given the immense scope of this challenge, no single government or organization can drive the entire process alone. While existing programs like the United States's Intelligence Advanced Research Projects Activity (IARPA) and the UK's Advanced Research and Invention Agency (ARIA) offer pathways for funding and collaboration, the burden is too large for any nation to bear. Consequently, a partner approach—either bilaterally between the United States and UK or even across the Five Eyes alliance—can serve as an optimal foundation. Such a coalition allows for pooling resources, expertise, and strategic vision. At the same time, it diminishes financial risks by distributing them among multiple stakeholders, increasing the likelihood of sustained progress.

In forging these partnerships, it is crucial to balance broadening the coalition through public-private partnerships, ensuring the national security community has access to the AI/ML skills needed to build this capability, and preserving a level of agility. Too many participants can convolute decision-making processes, dilute accountability, and slow the pace of innovation. Conversely, too few partners risk overlooking the diverse expertise needed for meaningful breakthroughs. Therefore, the optimal balance lies in assembling a robust yet manageable consortium capable of multiple parallel experiments, data-sharing, and consistently refining best practices.

Action 3 - Prioritize Human–Machine Teaming with Rigorous Safeguards

All these efforts must remain firmly rooted in the principles of human–machine teaming. The interplay between human creativity and machine efficiency provides the strongest foundation for transformative progress, offering advantages in everything from data analysis to operational execution. This cooperative dynamic also highlights the importance of rigorous safeguards to maintain the integrity and security of each partner's systems. Regulatory frameworks, ethical guidelines, and technical standards should evolve to minimize vulnerabilities and ensure the responsible development of emerging technologies.

[111] Memos to the President: Artificial General Intelligence (AGI), Special Competitive Studies Project at 4 (2025).

Action 4 - Encourage Policymakers to Leverage Enhanced Decision Advantage

With technological change accelerating, deliberate action and foresight are indispensable. However, if policymakers are given warnings even earlier, they must take full advantage of them. Receiving strategic warnings earlier always risks policymakers thinking that the issue does not need immediate attention. In doing so, they will position the project—and the broader international community—to move confidently through each phase, culminating in a lasting, transformative impact.