The Importance of Analytic Superiority in a World of Big Data and AI

Dr. Robert L. Grossman
Dr. Emily O. Goldman

ABSTRACT

Rapid advances in machine learning, deep learning, artificial intelligence (AI), large language models, and generative AI have accelerated efforts to leverage these technologies for military advantage. We refer to these and related technologies as analytics. We present a framework as a guide to achieving "analytic superiority," which is the operational advantage obtained through the ability to collect data required for analytics; build useful, performant, and robust analytic models; and deploy analytic models in operational systems to achieve objectives, while exploiting or denying an adversary's ability to do the same. Analytic superiority is best understood in the context of the analytic capabilities of one's adversaries, who also collect data, build models, and deploy them to achieve their own objectives and defeat the analytics of their adversaries. This framework emphasizes developing an analytic strategy, collecting the data required, developing analytic infrastructure for managing and analyzing the data, building analytic models, and deploying analytics into operational systems to meet the objectives required by the analytic strategy. Although analytic competition is not new, it is an under-appreciated dimension of military and strategic competition, and it is advancing at a faster pace than any previous technological competition. We discuss how U.S. cyberspace superiority, which is foundational to military advantage in the physical domains, now depends on prevailing in analytic competition with adversaries and thus requires adopting a strategy and processes to achieve analytic superiority.

© 2024 Dr. Robert L. Grossman and Dr. Emily O. Goldman

Dr. Robert L. Grossman is the Frederick H. Rawson Distinguished Service Professor in Medicine and Computer Science and the Director of the Center for Translational Data Science at the University of Chicago. He is also a Partner at Analytic Strategy Partners, which helps organizations develop and execute AI strategies. He is the principal investigator for the National Cancer Institute Genomic Data Commons (GDC), a platform for the cancer research community that manages, analyzes, integrates, and shares large-scale genomic datasets in support of precision medicine. He founded Open Data Group in 2002 and was its Managing Partner from 2002-2015. Open Data Group provided analytic services to help companies build and deploy machine learning models over big data and to operationalize AI.

ARTIFICIAL INTELLIGENCE AND MILITARY ADVANTAGE

Rapid advances in machine learning (ML), artificial intelligence (AI), and generative AI (GAI), which we will refer to simply as ML/AI, will play an ever-increasing role in commercial, defense, and intelligence systems. There has been continual progress in ML/AI over the past forty plus years, punctuated by several inflection points. One important inflection point occurred in the early 2000's with deep learning that leveraged GPUs (the graphics processors used by gaming applications). This allowed neural networks with many dozens of layers and millions of parameters to be used, leading to significant advances in the processing of text and images.[1] Increasingly larger deep learning neural networks with billions of parameters were built over larger and larger GPU clusters and then specialized to particular applications in various ways, leading to the emergence of large language models (LLM) and generative AI. Another important inflection point occurred on November 30, 2022 with the release of ChatGPT by OpenAI. ChatGPT provided Large Language Models as a Service, setting a new record by reaching over 100 million users by January 2023.

These two inflection points have fundamentally changed the power and capabilities of analytics. The rapid rise in ML/AI-focused strategies, policies, regulations, and guidance documents by States, international organizations, and supra-national organizations like the European Union reflects widespread recognition of ML/AI's potential and risks.[2] The impact on military systems cannot be overemphasized. Over time, ML/AI models and services will be ubiquitous, and enduring military advantage will depend on the strategy and processes for both employing ML/AI towards strategic objectives and countering adversaries' use of ML/AI toward the same objectives.

Dr. Emily Goldman serves as a strategist at U.S. Cyber Command and a thought leader on cyber policy. She was cyber advisor to the Director of Policy Planning at the Department of State, 2018–19. From 2014 to 2018 she directed the U.S. Cyber Command / National Security Agency Combined Action Group, leading a team that wrote the 2018 U.S. Cyber Command vision, *Achieve and Maintain Cyberspace Superiority*. She was a professor of Political Science at the University of California, Davis, for two decades and has published and lectured widely on strategy, cybersecurity, and military innovation. *Cyber Persistence Theory: Redefining National Security in Cyberspace*, with Michael Fischerkeller and Richard Harknett, was published by Oxford University Press in 2022.

In this article, we use the term *analytics* to refer to statistical models, machine learning, deep learning, AI, GAI, and LLM. More expansively, by analytics we also include more general algorithms for processing and analyzing data, including embedded algorithms, such as those used in electronic warfare and electronic countermeasures. From the perspective of analytic superiority, the differences between algorithms,[3] models,[4] rules, machine learning, deep learning, AI, LLM and GAI[5] are not important. There is no standard name to refer to all of these and we will use the term analytics. More narrowly, we will use the term ML/AI if we want to emphasize some of the recent work in machine learning, deep learning, large language models, and generative AI.

Although analytic competition is not new, the ubiquity of models in digital systems and services has made analytic competition an increasingly important, but underappreciated dimension of military and strategic competition between states. This is also the case for commercial competition, but better grasped by the private sector. For defense and intelligence systems, in particular, the United States is competing with adversaries using their own analytics while trying to defeat, disadvantage, interfere, or trick ours. The race to deploy analytics in operations to achieve mission effects and degrade and defeat adversary analytics is as critical as, and occurring at a faster pace, than any arms race we have ever confronted. The United States must set up better analytic infrastructure, employ better analytic models, update them more frequently, build better analytic systems, and train better data scientists, system operators, and end users than our adversaries in order to achieve and maintain analytic superiority.

ANALYTIC SUPERIORITY AND SOME COMMON MISCONCEPTIONS

We define analytic superiority as the operational advantage obtained through the ability to collect data

required for analytics; build useful, performant, and robust analytic models; and deploy analytic models in operational systems (from core to edge) to achieve objectives, while exploiting or denying an adversary's ability to do the same. Note that analytic superiority is not the same as information superiority. The latter is a much broader concept, while analytic superiority is specifically focused on analytics (including ML/AI, models, rules, and related concepts) and their role in warfare and other adversarial situations.

Analytic superiority is best understood in the context of adversarial analytics and thus, measured against the analytic capabilities of one's adversaries. Competitors also collect data, build models, and deploy models to achieve their own objectives and deny, degrade, or defeat the analytics of their adversaries. Those who build better models; build them over larger volumes, higher velocities, and more varieties of data; and build agile and scalable infrastructures to deploy them are more likely to benefit from the operational advantages conferred by analytic superiority.

The importance of analytic competition for military competition is not new. A good example is provided by electronic warfare in World War II. As radar provided a military advantage to detecting and shooting down planes, electronic countermeasures were developed to trick or deceive radar. These included both physical countermeasures, like deploying chaff, and electronic countermeasures, like jamming or spoofing radar systems. The development of models for detecting planes and other moving objects using radar, and electronic countermeasures for deceiving these models, is an example of adversarial analytics and shows the importance of analytic superiority for achieving military objectives—in this instance destroying targets with bombers. Clearly, the concept of analytic superiority is not unique to ML/AI; it applies to all analytic models. This dynamic may enable the basic cycle of weapons development and counter measures, but these are not synonymous because many military innovations and counter innovations do not depend on the operational deployment of analytic models.

There are several common misconceptions about analytic superiority. The first is that analytic superiority is achieved solely by building more accurate and precise analytic models. One may have much better models with much higher precision and recall, but if the force cannot deploy the models into operational systems (for military and/or intelligence purposes), or not deploy them in time, while the adversary can, even if the adversary's models are worse, one will not achieve analytic superiority. Moreover, even with better models, if one lacks an analytic infrastructure that collects the data required and efficiently builds analytic models while an adversary has these capabilities, then one will not achieve analytic superiority. In general, the higher quality and the timelier the data, the simpler the models can be and the more effective and robust they will be when deployed. It is nearly impossible to achieve analytic superiority without a sufficiently powerful computing infrastructure to manage and analyze data and to build and deploy models.

A second common misconception is that today's ML/AI models require a fundamentally new approach to analytics. The importance of analytics is not new; nor is the viewpoint, as we argue below, that analytics in competitive environments require an end-to-end process for getting and managing data, building models, deploying models, and extracting value from models (as captured in Figure 1 by the framework of the analytic diamond, which is described below). It is also not new that novel models emerge from time to time and provide disruptive changes in what is possible in analytics.[6] Moreover, it has always been a challenge for large organizations to develop complex software systems, especially those involving large-scale or complex data.[7] That being said, over the past forty years, the exponentially growing amount of data and computing infrastructure, enabled by Moore's Law, has created year by year ever more powerful analytic models built on ever larger amounts of data. Even experts today are surprised at the speed at which new AI/GAI algorithms, models, systems, services, and applications are being developed, and the unexpected power and capabilities that these models, systems, and services have. This is widely recognized across commercial and military sectors, and by national governments globally.

A third common misconception is that analytics is mainly about improving the workflows of intelligence analysts. Of course, AI-powered analytics will undoubtedly improve intelligence analyst workflow, but they will do much more. They will impact any military system (e.g. weapons and navigational platforms) that employs analytic models, which means virtually all military platforms.

The U.S. Department of Defense's 2023 *Data, Analytics, and Artificial Intelligence Adoption Strategy* calls for a "systematic, agile approach to data, analytics, and AI adoption that is repeatable by all DoD Components" for enduring decision advantage. U.S. Cyber Command stood up an AI task force to move from "opportunistic AI application to systematic adoption."[8] Its focus is delivering capabilities to the force, posturing the Command to enable AI adoption, and countering AI threats. As the Department of Defense invests in new technologies, makes associated infrastructure, organizational, and force design decisions, it should do so informed by a framework for achieving and sustaining analytic superiority to provide both decision advantage for commanders and operational advantage for warfighters.

ANALYTIC SUPERIORITY IN THE COMMERCIAL WORLD

The importance of analytics for business competitiveness in general has been long recognized.[9] More specifically, in some commercial applications, the specific role of analytic superiority has also been long recognized. We discuss two examples in some detail—payments fraud and high frequency trading—because it is easier to talk at the granular level about commercial applications than it is about sensitive military applications.

A payment network enables a group of financial institutions to transfer funds digitally between individuals, businesses, or financial institutions. The most well-known examples are the payment networks between banks and merchants that enable individuals and businesses to use credit cards to buy merchandise and pay later when billed by their banks. Payment networks also support ATM withdrawals, gift cards, mobile payments, and other types of transactions. The development of analytic models by payments networks to stop fraud and the corresponding adjustment of tactics, techniques, and procedures (TTPs) by criminals trying to perpetrate fraud, is an example of competitive or adversarial analytics.

A wide array of bad actors, including individual criminals, criminal gangs, and states conduct payments fraud. In 2022, payments fraud totaled $11.64 billion in the U.S. and $32.34 billion worldwide.[10] Given the scale of payments fraud, a whole ecosystem has grown up around this industry. It includes the individuals using fraudulent cards or engaging in fraudulent transactions, software engineers developing software supporting fraudulent transactions, vendors selling stolen personal information that can be used to obtain fraudulent credit cards, lines of credit, etc., and electronic marketplaces where suppliers of data, software, and other services offer their goods and services to those engaging in fraud.

A payments system needs the right data at the right time to develop analytic models to quickly detect and stop fraudulent transactions. This requires a sufficiently powerful analytic infrastructure, analytic modeling capability, and, importantly, fraud systems (an example of analytic operations) to detect and stop fraudulent payments *as they are occurring* before losses are significant. At the same time, actions to stop fraudulent transactions should only minimally impact normal operations and the customer experience. By fraud systems we mean systems that not only detect putative fraud but also contact users to verify whether questionable transactions are in fact theirs, stop future use of compromised cards, and issue new credit cards to replace compromised cards. In this example, analytic operations include embedding the fraud models into operational systems for quickly stopping questionable transactions, quickly determining whether the transaction is valid or not, while minimizing impact on users. In the context of systems, there is always the balance of simpler models that work with the data immediately available versus more complex models that can leverage richer data that might not be immediately available.

Meanwhile, bad actors will adjust their tactics, techniques, and procedures to elude analytic models trying to detect and block their fraudulent transactions. The payments system is a dynamic space with sophisticated bad actors not only developing sophisticated attacks, but also quickly identifying simple new vulnerabilities that arise as systems are updated, new devices are installed, and new products are introduced. Companies tolerate a certain amount of fraud as the "cost of doing business." As the amount of payments fraud increases and begins to impact reputation or customer experience, companies will increase their efforts to detect and stop the fraud. Building the perfect analytic model is not the goal, but

rather building a good enough model that can be deployed and updated quickly enough to keep fraud at an acceptable level.

Analytic superiority has also been a critical aspect of high frequency trading since its inception. The key to high frequency trading is not just having the best models, but rather a trading system that can beat other trading systems for the small opportunities available with each trade. This requires having the right data, the right models, and an underlying computing infrastructure that can get the latest trades a bit faster than others in the market.

A good example is provided by the efforts of different traders to reduce the latency between moving data between Chicago and the New York / New Jersey region. As described in the book *Flash Boys* by Michael Lewis,[11] moving data between the Chicago Mercantile Exchange, where futures and options are traded, and Carteret, New Jersey, where the Nasdaq data center is located, took about 14.5 milliseconds (ms) using commercial Internet Service Providers (ISPs). To generate a competitive competive advantage to high frequency traders, Spread Networks developed a dark fiber network that took a more direct route between Chicago and the Nasdaq data center in New Jersey that was able to bring the latency of sending a message down to about 13.1 ms, providing a slight advantage to traders using their networks versus commercial ISPs. Finally, since light travels slower in glass fibers than in the air, Windy Apple Technologies set up a microwave network between Chicago and the New Jersey data centers that brought down the latency of sending a message to about 9.0 ms. Of course, traders had to also develop the appropriate analytic models, trading strategies, and computing infrastructure to take advantage of these millisecond improvements in networking infrastructure.

A FRAMEWORK TO ACHIEVE ANALYTIC SUPERIORITY

Achieving analytic superiority is much easier if one adopts a framework for analytics. The analytic diamond is one such framework that has proven useful,[12] although other frameworks could be used. The analytic diamond distinguishes processes for developing an analytic strategy, for building analytic models, for deploying analytic models, and for managing all the data required to build and deploy analytic models. It further defines an associated analytic governance model as well as an analytic maturity model to measure progress towards developing the processes, infrastructure, and talent required to achieving analytic superiority.

More specifically, an analytic strategy examines analytic opportunities, decides which to pursue, ensures that efforts are most likely to result in value to the organization, and measures the value. Analytic infrastructure collects and manages the data required for analytic operations and modeling. Analytic modeling analyzes data and builds the analytic models required by the organization. Analytic operations deploy and operationalize analytic models into products, services, or internal systems to extract value. These four features, when arranged in a diamond as in Figure 1, reveal six processes required for effective analytics, which in the context of analytic superiority means effective against an adversary.

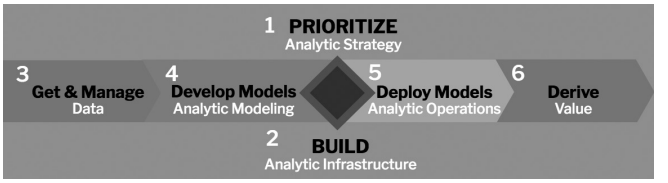

Figure 1. The Analytic Diamond

Organizations need to (1) develop an analytic strategy that prioritizes analytic projects and efforts, and (2) build the analytic infrastructure to train models over large amounts of data. Processes for (3) collecting and managing data so that it is available for modeling and operations, (4) developing appropriate analytic models, and (5) deploying the analytic models into operational systems must all be executed in the time frame required to achieve the objectives and value identified in the analytic strategy. An organization must also develop the operational systems, and their associated TTPs, to achieve the objectives and value identified in the analytic strategy. By operational systems here we mean any system that integrates, embeds, or interoperates with analytics, such as: systems for cyber defense and cyber operations; analysts' tools and support systems; electronic warfare and electronic countermeasures; systems supporting information superiority; fire systems for weapons platforms; systems supporting the targeting cycle, etc. The next step is to (6) accurately measure the value and impact achieved through the operational systems, and update the entire process accordingly. Enabling these processes is analytic governance, analytic security and compliance, and recruiting, training, and retaining the technical staff required. The analytic diamond can be applied to a wide variety of military operational priorities, as can be seen from the list of operational systems above that leverage analytics.

The analytic diamond shows why analytic superiority is not synonymous with a data strategy (although a data strategy is a subset of analytic strategy, as are the required components of an IT strategy to support analytics). Nor is analytic superiority reducible to building models and algorithms. In practice, analytic models have little value until they are turned into a functional system through an engineering process (AI engineering) and then the system is integrated into operational processes and/or operational weapon systems to bring value to the organization (AI operations or AIOps). Analytic superiority is sometimes misleadingly viewed as an OODA loop, which is a tactical activity rather than an enterprise level construct that integrates four analytic functions—strategy, infrastructure, modeling, and operations—to achieve competitive advantage over an adversary.

Although there is no single way to achieve analytic superiority, just as there is no one way to defeat an adversary in a particular engagement, in general achieving it includes some combination of the following: analytic task advantage (for example, having better analytic models than your adversary and updating them more frequently); asymmetric advantage (for example, asymmetrically attacking your adversary's models or infrastructure); and autonomy advantage (leveraging autonomy and semi-autonomy more effectively). Autonomy and counter-autonomy are familiar and well understood in the context of autonomous weapons

systems,[13] but are perhaps less familiar and less well understood in the context of adversarial analytics. Autonomy and semi-autonomy provide scale and speed for analytics, both for defensive and offensive operations. Countering these operations can be done by identifying suitable weak points in adversaries' analytic workflows and then disrupting and defeating those weak points. Figure 2 shows notionally how Red and Blue attack each others' analytic infrastructure, analytic models, analytic operations, and analytic strategy. For example, Red may try to poison Blue's data and embed in Blue's analytic infrastructure, while Blue may try to weaken or poison Red's analytic models and blunt the operational impact of Red's models.

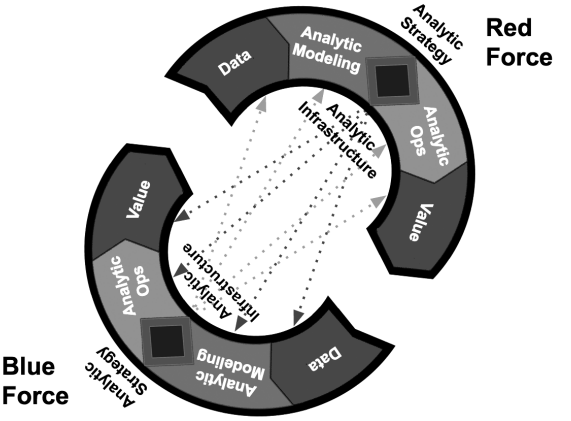

Figure 2. Achieving Analytic Superiority.

Achieving analytic superiority requires that all the analytic models needed to achieve the goals of an analytic strategy have the data they need to function, are updated as required, and are integrated with the required operational systems to achieve the desired outcomes. This is not easy to achieve. Here are two simplified examples to illustrate this point.

First, consider the role of sensors in cybersecurity. Sensors can produce a lot of data and unless there is an adequate analytic infrastructure (the bottom of the diamond) to manage the data, critical patterns that may span many days of data can be missed. Detecting behavioral patterns in log and sensor data require not only analytic models (the left-hand side of the diamond), but keeping these models up to date as TTPs change. Once models detect potential intrusions, analytic operations (the right-hand side of the diamond) determine the appropriate actions and counter measures to quickly pinpoint, contain, and mitigate the intrusions. Finally, there are always limited resources, and analytic strategy (the top of the diamond) prioritizes which problems are tackled and in what order. These factors have always been present in the management of data for analytics, the building of models for analytics, and the deployment of models to take appropriate actions. On the other hand, recent advances in ML/AI create new opportunities for both the U.S. and our adversaries, and managing and coordinating the components of the analytic diamond across the enterprise to achieve analytic superiority is critical.

As a second example, protecting a Navy ship requires object identification and tracking models to identify potential threats and track them so that appropriate fires and countermeasures can be assigned to different threats. Each of these systems requires appropriate analytics to protect the ship, and the different analytic models must work together. More precisely, the outputs of some models are the inputs to other models creating an *analytic workflow*. To achieve the

desired outcome each model must get the data it needs. This interdependence creates fragility; thus, various techniques are used to create more robust models and more robust analytic workflows. For example, models and workflows can impute data when it is missing,[14] detect and remove outliers,[15] and reduce the fidelity of models in order to make the models and workflows more robust and more likely to provide correct results when data is missing or delayed, or when sensors are down, for example.

The concept of a system of systems (SOS) is familiar in the context of complex weapon systems.[16] In the context of analytic superiority, it is helpful to extend this to what one might call a system of systems *and models* (SOSAM). This includes not only the system of systems, but also all the analytic models and workflows needed for the system of systems to achieve the desired outcomes for the analytic strategy. This almost always requires multiple analytic models and systems. In the simplified example, an adversary need only identify and defeat the weakest link in an analytic workflow to attack the ship successfully. To achieve analytic superiority, your SOSAM must be more capable than your adversary's SOSAM in all operating conditions, including in crisis and conflict. This is a concept familiar to cyber defenders trying to keep an adversary out of their system. No matter how good the algorithms, signatures, and models in your firewalls, an adversary need only exploit any weak link in the system, whether from weak passwords, default passwords, phishing, stolen credentials, cross site scripting attacks, etc. More generally, as we discuss below in the section on analytic maturity models, it is also critical to build analytics over two or more analytic systems or workflows ("hierarchical" analytics) to detect patterns and threats that may not be apparent otherwise. Analytic superiority, it turns out, is rarely about one's analytic model simply being more accurate (measured as the percentage of correct predictions) than the adversary's model, but rather about the robustness and effectiveness of entire analytic workflows or hierarchies of analytic workflows.

MEASURING THE EFFECTIVENESS OF ANALYTIC SYSTEMS

Defense and intelligence analysts are familiar with the application of analytic models to digest intelligence and defense data, organize it, draw conclusions with confidence levels, write reports, and disseminate the reports to decision makers and policy makers. Although multiple analytic models may be used to collect and analyze the data, the output is the report; actions, if any, are the responsibility of some other part of the organization. Analytic superiority, as applied to military activities, requires that the outputs of analytic models be integrated into operational systems to achieve specific mission related objectives. The focus, therefore, must be on the effectiveness of the analytic system as a whole, not just the output of analytic models.

One way to understand this difference is to think of the output of analytic models as scores (say from 1 to 100), which are used to take *actions*, and the operational significance of these actions are evaluated with *measures*. This sequence is summarized with the acronym *SAM* for scores-actions-measures.[17] Typically, the effectiveness of a single analytic model is measured

with detection and false positive rates. In contrast, from the SAM perspective, these model-specific metrics influence how often the appropriate actions are taken for the appropriate situations. The effectiveness of the actions determines the measures, which summarize the effectiveness of the analytic system as a whole.

Consider a cyber defense model that identifies users whose credentials may have been compromised. Assume that the model produces a score in the range 1-100, that credentials are revoked for scores above 90 and then manually investigated (the action), and credentials are manually investigated for scores between 80 and 90 (another action). The model producing these scores has a detection rate and false positive rate, but measures of effectiveness that incorporate action might be how quickly compromised credentials are detected (seconds, minutes, etc.), how much damage is done before they are found (say measured with dollars), and how much lost productivity (measured with person-hours) arises from improperly revoked credentials. Measures can be created at various levels, such as the analytic task level, at the analytic system level, and at the mission level.

As another example of SAM, a network or system intrusion might set off multiple alerts from multiple models, including alerts from firewalls, behavioral alerts from endpoint systems, alerts from identity and access management systems, etc. Many of these are false positives, producing alert fatigue. One solution is to build models that process the outputs of all these systems and produce alerts at the incident level. The associated actions might include automatically revoking access credentials or isolating network segments when the alert scores are high enough. Measures might reflect the amount that alerts are reduced, say from 100 alerts from the original systems that alert to 5 alerts of candidate incidents, and the percentage of true incidents detected.

Other types of analytic models produce different outputs. For example, large language models and generative AI models produce text, images and other outputs, giving rise to text-action-measures, or *TAM*. It is not the text that is evaluated, but text associated with some action, such as producing SQL code for querying a database of cyber threat data or log data, and the measure might be the productivity increase for junior and senior software developers. Measures can be devised to assess the cumulative impact of actions taken against ransomware groups or an advanced persistent threat actor such as reducing the success of intrusions or improving the targeting process. Other types of measures can be devised to assess decision advantage, such as efficiency in organizing and prioritizing information so that a commander can make decisions faster and still leverage more of the available information.

ANALYTIC GOVERNANCE

Analytic superiority requires a governance framework to ensure that the right people, processes, and frameworks are in place to identify and achieve the organization's priority analytic objectives. There are broad similarities between analytic governance and IT governance

processes designed to control and oversee risk. IT governance ensures that IT investments generate business value; that risks associated with IT are mitigated; and that the organization makes sound, long-term decisions with accountability and traceability to those funding, developing, supporting, and using IT resources.[18]

With this definition and the analytic diamond in mind, we argue that an analytic governance framework should accomplish the following: (1) ensure sound long-term decisions about analytics are reached and that investments in analytics generate value; (2) operate in such a way that data, derived data, and data and analytic products are protected and managed in a secure and compliant fashion; (3) ensure accountability, transparency, and traceability to those funding, supporting, and using analytic resources; and (4) ensure analytic workflow: that data is available to modelers, that analytic models can be deployed, and that impact and value of analytic models is quantified and tracked.[19]

Analytic governance is Commander's business and cannot be delegated because it crosses the entire enterprise. By contrast, data and sensor governance can be delegated further down in the organization. Typically, data governance is delegated to the CIO. Analytic governance, however, requires an influential champion to ensure the integration of functions--managing data, building models, exploiting models for advantage by analysts and operators--at scale. These processes must be driven by operational priorities, and by the organizational element that sets those priorities.

ANALYTIC MATURITY

Just as software maturity can be measured with a Capability Maturity Model that quantifies the level at which an organization's business processes develop software and complete a software project,[20] processes for developing and deploying analytics can be developed to measure an organization's analytic maturity. Depending upon the desired outcome, analytic models must be integrated across various hierarchies and across different units within an organization. For this reason, the analytic maturity of an organization is assessed along three dimensions.[21] First, how repeatable are the processes for managing data, building models, deploying models into operations, and quantifying the effectiveness and impact of the models? Second, how widespread are these processes, both horizontally and vertically throughout the organizational structure? Third, do these processes support the organization's analytic strategy?

Each organizational unit that supports analytics can be evaluated along a maturity dimension. For example, can the unit 1) build reports based upon data; 2) build and deploy analytic models based upon data; 3) build and deploy analytic models with a repeatable process; 4) build and deploy the appropriate analytic models, as prioritized by the analytic strategy. Similarly, the entire enterprise can be assessed along a maturity dimension: 1) does the organization provide the enterprise services so that a unit can get the data, build the models,

deploy the models, and extract the value required; 2) how widespread across the organization are units that build and deploy high quality and effective analytics; 3) is there analytic governance at the enterprise level; 4) are analytic models built by different units integrated together in such a way to achieve the organization's overall analytic objectives or do they create unanticipated challenges for some analytic models?

As a simple example from the financial services world, one division of credit card issuers is responsible for acquiring new customers with acquisition models and expanding the company's "share of their wallet" with cross sell models. Another division (credit risk) is responsible for determining how likely customers are to default on the credit extended to them with credit default models. The enterprise dimension of an analytic maturity model manages whether the right customers are being acquired so that the acquisition division does not meet its yearly targets by acquiring customers who are likely to cause future problems by defaulting on their loans.

ANALYTIC SUPERIORITY AND PERSISTENT ENGAGEMENT IN CYBERSPACE

Superiority in cyberspace has come to be seen as foundational to military advantage in the physical domains of land, air, sea, and space. The United States is not alone in this assessment. The People's Republic of China (PRC) also sees superiority in cyberspace as core to its theories of victory.[22] For this reason, we apply the concept of analytic superiority to cyberspace, which has itself become a major battleground in strategic competition among states.

In 2018, U.S. Cyber Command defined cyberspace superiority as "the degree of dominance in cyberspace by one force that permits the secure, reliable conduct of operations by that force, and its related land, air, maritime, and space forces at a given time and place without prohibitive interference by an adversary."[23] Cyberspace is not only a domain of military operations, however, but also a strategic environment in and through which adversaries put U.S. national security at risk.[24] Adversaries operate continuously in and through cyberspace below the threshold of armed conflict, extending their influence and eroding U.S. military, economic, and political power without resort to physical aggression. General Timothy Haugh, Commander of U.S. Cyber Command and Director of the National Security Agency, recently remarked that the PRC is pursuing a policy of global dominance, but hopes to achieve that without a kinetic, real-world military fight. They are using cutting-edge technologies to achieve advantage.[25]

U.S. cyber forces adapted their operational approach to address the requirements of strategic competition outside armed conflict as well as those aimed at helping to deter armed conflict and prevail should it occur.[26] Recognizing that the United States was losing ground to adversaries operating in and through cyberspace below armed conflict, U.S. Cyber Command pivoted from a bias for "cyber response" to action through "cyber persistence." Empowered with new authorities, military cyber forces began proactively defending and contesting adversaries globally, continuously, and at scale. Reflecting upon his five years as commander, General Paul Nakasone

remarked, "I think we got persistent engagement completely right. ... If you're on the sidelines watching this, you're going to get hit. That's why I think it's so important for our forces worldwide to be able to be engaged, and being able to act and understand what our adversaries are doing ... Being able to continue to operate day in and day out, this is how you get really good. You operate in the domain."[27] The Russia-Ukraine war has shown how persistent engagement also sets conditions for successful contingency operations by curtailing an adversary's freedom of maneuver (e.g., precluding options, eroding confidence, generating organizational friction, and degrading capabilities), and generating options and opportunities for crisis and conflict.

Cyberspace is an interconnected, fluid space wherein those who continuously anticipate and act can set, reset, and maintain conditions in their favor. This is what it means to hold the initiative in cyberspace, where the scale and scope of change is so vast that anticipating exploitation and acting based on that insight is the key to security. A state that cedes the initiative can assume that its adversaries will set conditions to undermine its security while increasing their own.[28] Highly capable adversaries are conducting operations in cyberspace at a volume and pace that overtake the ability of defenders to discover and counter. Increasingly sophisticated techniques, such as "Living off the Land" (LOTL),[29] allow them to evade detection more easily and preposition in military and civilian critical infrastructure, setting conditions for use during crisis and conflict. Defeating complex obfuscation techniques requires timely processing of massive data and deploying analytics to detect this type of behavior into operational (in this instance defensive) systems at speed and scale to outpace the malicious actor's ability to shift across infrastructures and preclude interdiction. With innovations in big data and AI, the cyberspace force that achieves and sustains superiority in analytic competition will have the initiative in persistent cyber engagements.

Tactically, one can still think in terms of prevailing in defensive and offensive analytics, but cyberspace is so fluid and dynamic that offensive and defensive advantage is not meaningful at a strategic level. What is meaningful is whether or not one has the initiative in setting the conditions of cyberspace in one's favor, where "those conditions are measured as the relative balance between being cyber vulnerable to exploitation and being able to exploit the cyber vulnerabilities of others" by "anticipating the exploitation that *will come next* by either you as a defender or another State as an attacker."[30] Developments in AI have made analytic superiority necessary for achieving and sustaining cyberspace superiority.

The Joint Cyber Warfighting Architecture (JCWA) is U.S. Cyber Command's platform and associated capabilities that enable Cyber Operations Forces to conduct full-spectrum cyberspace operations, globally at-scale. The Unified Platform is the data hub of JCWA and provides the analytic infrastructure for deploying and managing data. The federated nature of this system for deploying and managing data is one reason U.S. Cyber Command was granted greater technical responsibility and authority to direct the development, integration and fielding of critical capabilities and infrastructure in the JCWA.

The framework of the analytic diamond specifies other requirements for achieving analytic superiority, foremost being a strategy that prioritizes analytic objectives and requirements. A subordinate data strategy should identify data collection, management, standards, and analytics for full-spectrum operations, which in turn should inform and drive specific computing capabilities, data management systems, data models, and analytic platforms. Designated entities should be made responsible for building and deploying analytic models. Some element of the organization should have the mission to attack all elements of the adversary's analytic diamond as a priority or objective, for example by poisoning data, countering models or infrastructure, or blunting the operational impact of the adversary's models. Just as there are well defined disciplines of counter intelligence, counter terrorism, and cyber defense to counter cyber-attacks, we need to develop a discipline of counter analytics to understand how to disrupt and defeat the analytics of our adversaries. Given that DoD is organized around weapons platforms, U.S. Cyber Command may want to think of its entire organization as a platform to acquire, manage, develop, and deploy data and analytics for full-spectrum operations at scale—and counter adversaries' ability to do the same.

CHALLENGES AND LESSONS LEARNED

Achieving analytic superiority is challenging. Many of the hurdles are familiar ones that DoD has faced when developing software, managing and analyzing data at scale, and interoperating defense systems. For example, DoD's transition to cloud computing, an important enabling technology for training and using ML/AI over large datasets, has been rocky, significantly lagging behind commercial adoption of the technology.[31] Many software projects fail for a variety of well understood reasons,[32] and software systems supporting intelligence and defense systems are particularly challenging.[33]

Many analytic/AI projects, like many data warehousing projects, fail because assembling and managing teams with multiple skills is hard. Building a data warehouse requires assembling and managing a team that knows about both software and data, which are two distinct skills. Many analytic/AI projects also fail because a disproportionate amount of time, ingenuity, and effort is spent focusing on the AI algorithm or analytic model of interest, say a convolutional neural network (CNN), or fine tuning an LLM. AI projects that succeed, however, spend significant time, effort, and ingenuity to get all the data required and to deploy the model in such a way to achieve an operational advantage.[34] It is helpful to think of this as a three-phase process. Phase one focuses on acquiring and curating data to be used as inputs to the AI and analytic algorithms ("collecting the data required for the models"). Phase three focuses on integrating the outputs of the AI and analytic algorithms into current operational systems or developing new systems around them ("deploying the analytic models"), and then extracting operational value from the systems as a whole, once the model has been deployed. Many successful projects spend less than 20% of project's total time and effort on phase two—

finding the right algorithms and processing the data using them ("developing the analytic models"). See Figure 1.

Achieving analytic superiority also requires that the right data from the right sensor get to the right analytic model in the right time. This is especially challenging when the sensors, models, and effects cross war fighting domains. Programs such as the Joint All-Domain Command and Control (JADC2) are designed to address these issues, but face challenges of their own.[35] Finally, achieving and maintaining analytic superiority requires a developer, DevOps, and DevSecOps mindset, environment, and acquisition process. For many DoD software projects this has been challenging to achieve because the defense acquisition process is not agile.

AI IN THE PRC

The PRC is investing heavily in AI, cloud computing, and related technologies.[36] It is engaged in a campaign to attain technological superiority and place U.S. critical and national infrastructure at risk.[37] In some ways, the Chinese have an edge with their smart cities initiatives and extensive industrial online-to-offline (O2O) industrial base.[38] With these efforts, the PRC's many data engineers are gaining deep experience developing and deploying systems at scale that use big data and AI every day. Furthermore, much of the technology developed by Chinese companies for smart cities is dual use and can also be applied to military systems. Those Chinese companies work both commercially and for the government, and competition among them is usually regarded as extremely intense, yielding rapid technological advancement.

The PRC is also acquiring a wide range of data sources, to include proprietary big data, machine learning, and AI technology stolen from U.S. companies through computer intrusions.[39] Another source of large scale data and analysis that is particular to the PRC is provided by its social credit system, which integrates data from multiple sources, including analysis of internet traffic, monitoring of social media, analysis of mobile telemetry data from phones and cars, and pervasive video surveillance. All of this data is analyzed at scale using AI and other techniques to produce a score for each individual.[40] The scale of data provided by Baidu, Alibaba and Tencent (BAT) provides smaller companies that are part of the Baidu, Alibaba or Tencent ecosystem the data required for AI and machine learning. These qualities of the PRC's analytic ecosystem make achieving analytic superiority over China challenging, but essential.

CONCLUSION

Statistical models, machine learning, AI models and related technologies (what we call analytics in this article) have been critical enabling technologies ever since military systems integrated digital technology. A decade ago deep learning substantially increased the performance of these models. More recently, the emergence of LLM and GAI models introduced significantly new capabilities to these models. Yet powerful models alone do not necessarily provide an advantage in competition, crisis, or conflict.

We focus on the importance of analytic superiority and offer a framework—the analytic diamond—as a guide to achieving analytic superiority. The framework stresses the importance of an analytic strategy for identifying and prioritizing analytic opportunities; obtaining the required data and building an analytic infrastructure that manages and analyzes the data; analytic modeling, which builds the models; analytic operations that deploy models into operational systems; and measures that capture the operational impact of analytics against the strategic goals identified. Analytics are usually integrated into analytic workflows and adversaries can be expected to attack the weakest point in these workflows to obtain an advantage.

It is standard in many organizations today to develop a sensor strategy, a data strategy, a cloud strategy, and an AI strategy. But without an organizing principle like analytic superiority to tie these together, they are not likely to provide the advantages needed in competition, crisis and conflict.

DISCLAIMER

The views expressed are those of the authors; in particular, they do not reflect the official position of any U.S. Government Agency.

NOTES

1. Yoshua Bengio, Yann Lecun, and Geoffrey Hinton, "Deep Learning for AI," (2021). *Communications of the ACM* 64, no. 7: 58-65.
2. World Travel & Tourism Council, *Artificial Intelligence: Global Strategies, Policies and Regulatory Landscape* (April 17, 2024). https://researchhub.wttc.org/product/responsible-artificial-intelligence-ai, 4.
3. Algorithms are instructions to solve problems, perform computations, or process data that are precise enough that they can be implemented with a computer.
4. We can think of machine learning or statistical modeling as a process that takes data as an input and produces a model as an output (this is called a training model"). There are several different types of models, but some common types are models that make predictions, models that summarize data, and models that correlate one variable with another. If you fix a model, inferencing, which is also called scoring, is the process that given input data produces an output. For example, in a large language model, the input is the prompt, and the output is the response. In a predictive model, the input is a feature vector, and the output is the prediction (yes/no, a numerical prediction, etc.) Rules are a very basic type of model that, in the simplest case, consists of multiple if...then...else statements, which may be nested.
5. A neural network is a type of predictive model that processes data in layers, with each layer a very simplified mathematical model of a neuron. Deep learning is a type of neural network that uses many sequential layers for processing data. Large language models are a particular type of deep learning models that take prompts asking questions as inputs and produce text responses as outputs. Generative AI are models that take prompts as inputs and can produce pictures, videos, and other objects as outputs.
6. Yash Sherry and Neil C. Thompson, "How Fast do Algorithms Improve?," (2021). Sherry, *MIT Initiative on the Digital Economy* Vol. 6. https://ide.mit.edu/wp-content/uploads/2021/09 /How_Fast_Do_Algorithms_Improve.pdf, 2-4.
7. Benjamin Jensen, Christopher Whyte, and Scott Cuomo, "The Future of Algorithmic Warfare: Fragmented Development," (July 20, 2023) *War on The Rocks*, https://warontherocks.com/2023/07/the-future-of-algorithmic-warfare-fragmented-development/.
8. Timothy D. Haugh, *Summit on Modern Conflict and Emerging Threats*, (April 17, 2024). Remarks, Vanderbilt University, Nashville, TN.
9. Thomas H. Davenport, "Competing on Analytics," (2006). *Harvard Business Review* 84, no. 1: 98. https://hbr.org/2006/01/competing-on-analytics.
10. "Global Network Card Losses Worldwide," (May 2024) in *The Nilson Report* No. 1264, Credit Card Fraud, https://nilsonreport.com/the-current-issue/, 5.
11. Michael Lewis, *Flash Boys: A Wall Street Revolt* (2014). New York: WW Norton & Company.
12. Robert L. Grossman, "A Framework for Evaluating the Analytic Maturity of an Organization," (2018) *International Journal of Information Management* 38, no. 1, https://www.sciencedirect.com/science/article/pii/S0268401217300026: 45-51.
13. Ruth A. David and Paul Nielsen, *Defense Science Board Summer Study on Autonomy* (Defense Science Board: 2016). Mark Maybury and James Carlini, *Defense Science Board Report on Counter Autonomy* (Defense Science Board: 2020), https://dsb.cto.mil/reports/2020s/CA_ExecutiveSummary.pdf. A. Etzioni, "Pros and Cons of Autonomous Weapons Systems," (with Oren Etzioni), in A. Etzioni, ed., *Happiness is the Wrong Metric: A Liberal Communitarian Response to Populism* (Springer International Publishing: 2018), pp. 253-263, https://doi.org/10.1007/978-3-319-69623-2_16.
14. W.C. Lin and C.F. Tsai, "Missing value imputation: A review and analysis of the literature (2006–2017)," (2020) Artificial *Intelligence Review*, 53:2, https://doi.org/10.1007/s10462-019-09709-4, 1487–1509.
15. V. Hodge and J. Austin, "A Survey of Outlier Detection Methodologies," (2024). *Artificial Intelligence Review* 22:2, https://doi.org/10.1023/B:AIRE.0000045502.10941.a9, 85-126.
16. J. Boardman and B. Sauser, "System of Systems—The Meaning of Of," (2006) *2006 IEEE/SMC International Conference on System of Systems Engineering*, 118-123. https://doi.org/10.1109/SYSOSE.2006.1652284, 6.
17. Robert L. Grossman, *Developing an AI Strategy: A Primer* (2020). https://analyticstrategy.com/wp-content/uploads/2020/03/Analytic_Strategy_Primer-TOC.pdf
18. Allen E. Brown and Gerald G. Grant, "Framing the frameworks: A Review of IT Governance Research," (2005). *Communications of the Association for Information Systems*, Volume 15, AIS eLibrary, https://aisel.aisnet.org/cgi/viewcontent.cgi?article=3160&context=cais, 696-712.

NOTES

19. Grossman (2018).
20. Watts S. Humphrey, *Managing the Software Process*, (1989). Reading, MA: Addison-Wesley, 89.
21. Grossman (2018).
22. U.S. Department of Defense, *Summary 2023 Cyber Strategy* (September 2023), p. 2, https://media.defense.gov/2023/Sep/12/2003299076/-1/-1/1/2023_DOD_Cyber_Strategy_Summary.PDF.
23. U.S. Cyber Command, *Achieve and Maintain Cyberspace Superiority: Command Vision for U.S. Cyber Command* (2018), https://www.cybercom.mil/Portals/56/Documents/USCYBERCOM%20Vision%20April%202018.pdf.
24. Michael P. Fischerkeller, Emily O. Goldman, and Richard J. Harknett, *Cyber Persistence Theory: Redefining National Security in Cyberspace (2022).* Oxford University Press.
25. Julian E. Barnes, "China Could Threaten Critical Infrastructure in a Conflict, NSA Chief Says," (17 April 2024). *The New York Times*, https://www.nytimes.com/2024/04/17/us/politics/china-cyber-us-infrastructure.html.
26. Emily O. Goldman and Michael Warner, "The Military Instrument in Cyber Strategy," (2021). *SAIS Review of International Affairs, vol, 41, no. 2.* Johns Hopkins University Press, https://muse.jhu.edu/article/852326, 51-60.
27. Martin Matishak, "Cyber Command, NSA usher in Haugh as new chief," (22 February 2024). *The Record,* https://therecord.media/cyber-command-nsa-usher-in-haugh-as-new-chief.
28. Fischerkeller, Goldman, and Harknett (2022).
29. Joint Cybersecurity Advisory (February 7, 2024), https://media.defense.gov/2024/Feb/07/2003389935/-1/-1/0/CSA-PRC-COMPROMISE-US-CRITICAL-INFRASTRUCTURE.PDF.
30. Joint Cybersecurity Advisory (February 7, 2024).
31. U.S. Department of Defense Defense Science Board, Cyber Security and Reliability in a Digital Cloud, January 2013 retrieve from https://dsb.cto.mil/reports/ on May 10, 2024. Odell, L., Wagner, R. and Weir, T., 2015. Department of Defense use of commercial cloud computing capabilities and services. IDA P-5287, Institute for Defense Analyses, Alexandria, Virginia.
32. R. N. Charette, "Why Software Fails [software failure]," (2005) *IEEE Spectrum*, 42(9): 42–49, https://doi.org/10.1109/MSPEC.2005.1502528
33. William LaPlante and Robert Wisnieff, *The Design and Acquisition of Software for Defense Systems* (2018), https://apps.dtic.mil/sti/citations/AD1048883.
34. Grossman (2020).
35. S. Lingel, J. Hagen, E. Hastings, M. Lee, M. Sargent, M. Walsh, L. A. Zhang, and D. Blancett, D. *Joint All-Domain Command and Control for Modern Warfare* (Santa Monica, RAND Corporation, 2020) https://community.apan.org/cfs-file/__key/docpreview-s/00-00-17-19-55/8156.RAND_5F00_RR4408z1.pdf.
36. Kai-Fu Lee, *AI Superpowers: China, Silicon Valley, and The New World Order*, (2018). Boston: Houghton Mifflin, 3.
37. Robert O. Work and Greg Grant. "Beating the Americans at their Own Game: An Offset Strategy with Chinese Characteristics" (2019). *Center for New American Security*, https://www.cnas.org/publications/reports/beating-the-americans-at-their-own-game: 195-260.
38. Lee (2018).
39. United States Department of Justice, "Two Chinese Hackers Associated with the Ministry of State Security Charged with Global Computer Intrusion Campaigns Targeting Intellectual Property and Confidential Business Information," (December 20, 2018), https://www.justice.gov/opa/pr/two-chinese-hackers-associated-ministry-state-security-charged-global-computer-intrusion.
40. Rogier Creemers, "China's Social Credit System: An Evolving Practice of Control," (May 9, 2018). Leiden Institute for Area Studies, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3175792, 3.