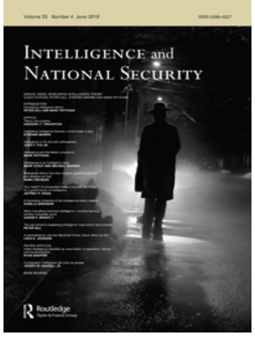

Intelligence and National Security

ISSN: 0268-4527 (Print) 1743-9019 (Online) Journal homepage: http://www.tandfonline.com/loi/fint20

# When everything becomes intelligence: machine learning and the connected world

Aaron F. Brantly

**To cite this article:** Aaron F. Brantly (2018) When everything becomes intelligence: machine learning and the connected world, Intelligence and National Security, 33:4, 562-573, DOI: 10.1080/02684527.2018.1452555

**To link to this article:** https://doi.org/10.1080/02684527.2018.1452555

Published online: 01 Jun 2018.

INTELLIGENCE AND NATIONAL SECURITY, 2018
VOL. 33, NO. 4, 562–573
https://doi.org/10.1080/02684527.2018.1452555

ARTICLE

# When everything becomes intelligence: machine learning and the connected world

Aaron F. Brantly

**ABSTRACT**
By 2020, the number of IOT devices will surpass 20.1 billion, and these devices combined with user interactions will generate enormous data streams that will challenge analytic capabilities constrained by human faculties, legal, regulatory, and policy frameworks designed for bygone eras. This work examines the impact of machine learning and artificial intelligence on legal, regulatory, policy and technical aspects of intelligence to provide insights into state, sub-state, and human behavior. This work develops an adaptive theory of Cyber Intelligence that will play an increasingly central role within the Intelligence Community in the decade(s) to come.

The human brain is able to process information at 1 exaflop – the equivalent of a billion billion calculations per second.[1] In comparison the Sunway-TaihuLight, the fastest super computer in the world as of 1 January 2017 has a processing speed of 125.436 petaflops – equivalent to more than one thousand trillion computations per second.[2] Michael Warner defines intelligence as 'a secret, state activity to understand or influence foreign entities'.[3] As information becomes increasingly prevalent, the boundary between intelligence and information is challenged insofar as it remains a secret and a state activity to understand or influence foreign entities. This paper argues that the rise of cyberspace and the connection of billions of Internet Protocol (IP)-enabled devices necessitates that future of intelligence will incorporate more and more machine learning to increasingly maintain a strategic and tactical advantage. Although the human brain still functions orders of magnitude faster than the fastest super computer, the ability to synthesize and process information and turn it into intelligence useful in understanding foreign entities remains a significant challenge.

Intelligence has always been a tug of war between art and science. While the balance has historically favored the art of human manipulation, post-World War II intelligence has increasingly privileged science. Nowhere is this truer than in the United States. This paper takes no position on the proper balance between art and science and instead acknowledges the intrinsic value of both. HUMINT collection remains relatively static in aggregate volume over time. Although there are increases and decreases in the number and quality of human assets they do not experience linear or even exponential growth patterns. The same is not true for virtually all other forms of data sources collected for intelligence from an ever-increasing array of novel sources. SIGINT, GEOINT, MASINT, OSINT, SOCINT, and CYBINT will be and are being challenged by the volume and velocity of the data being collected and analyzed.

Debate on a unified theory of intelligence remains as unsettled today as it did in 2009 when Peter Gill, Stephen Marrin, and Mark Phythian edited a robust volume on the key questions and debates in intelligence theory.[4] Despite efforts by scholars such as Johnson[5], Marrin[6], Scott and Jackson[7], Honig[8]

**CONTACT** Aaron F. Brantly ✉ abrantly@vt.edu

and others, the formalization of a unified theory of intelligence for national security remains elusive. The absence of a single theory is likely due to the diversity of issues which national security intelligence addresses. However, the absence of a single unified theory does not obviate the applicability of a theoretical approach to intelligence analysis. Rather than being constrained by a single theory, intelligence studies is uniquely positioned to leverage multiple theories from a plurality of disciplines to best address any given question relevant to a definition of intelligence such as the one by Michael Warner. At its most basic, theory serves as a roadmap or framework within which individuals can assess phenomena and or make predictions. The present work acknowledges the fluidity of the debate, yet focuses on the component aspects that lead to intelligence, the information or data that constitute the building blocks of subsequent intelligence.

At the time of Gill, Marrin, and Phythian's initial edited volume, the Internet and its attendant functions were growing in importance both within the intelligence community and within the field of intelligence studies as a subdiscipline of international relations. The volume published the same year as the establishment of the United States Cyber Command in June 2009 could not have anticipated the exponential increase in impact that the expansion of cyberspace would have on intelligence. In 2009 there were an estimated 1.77 billion people online in 2009, by the end of 2016 this number had nearly doubled to more than 3.5 billion.[9] Yet what is even more remarkable is not the number of users going online, but rather the increase in the number of internet enabled devices that received Internet Protocol (IP) addresses. Cisco estimated that between 2008 and 2009 the Internet of Things (IoT) first exceeded the number of human beings on the planet.[10] Cisco continues to estimate that by 2020 the number of internet enabled devices will reach approximately 50 billion.[11] The numbers estimated by CISCO are not without challengers, yet even these challengers admit that there are will likely be between 28 and 50 billion Internet connected devices.[12]

The expansion of the Internet is not simply the addition of computers and smartphones which have intrinsic intelligence value on individual or institutional targets, but large volumes of connected sensor systems, home, car, personal, corporate, and governmental systems that have been attached to the Internet. Documents leaked by Edward Snowden hinted at many of the devices being targeted by the NSA, GCHQ, and others in an effort to collect on new sources. Yet, even these sources are increasingly being outstripped by new products utilized by ranges of actors and industries.[13] The volume and velocity of data being generated by the expanding Internet is immense and incorporates collection methods from across all the collection types. The question facing the intelligence community is: in a world where data and information are increasing in every measurable capacity, where the hardest targets to track, the most difficult phenomena to measure and the most minute details are now increasingly available to analysts, what does the community need to do to serve the needs of its clients and can theory guide the community in its incorporation of new analytic methods and techniques to help it maintain relevance and still produce products that fit within the definition provided by Michael Warner?

Wihelm Agrell in a poignant article established a core concept that this article attempts to address: 'when intelligence is everything – nothing is intelligence'.[14] At the outset of this paper, it is important to clearly define that raw data collected by an ever-increasing number of sensors and devices globally is, in, and of itself not 'intelligence'. Rather, it is the informed utilization of this data to fit within Warner's definition above that constitutes intelligence. Data, absent rigorous and theoretically based analysis, remain simply data or information.

This article proceeds in three sections. First, how has the intelligence community dealt with disruptive changes in sources and methods in the past and does this provide insight for present and future challenges? Second, what role do analysts play in a world of ubiquitous information of enormous scale, velocity, and complexity? What should or could intelligence look like in the future with 50 billion or 100 billion devices all feeding back information to be analyzed? How do we ensure analytic integrity while increasing efficiency? Third, do our laws and policies within the US match the reality of an increasingly connected world? What are the current constraints imposed by laws and policies that hamper the incorporation of the variety and volume of new sources.

## The computers

The history of intelligence and computational mathematics are not commonly linked in social scientific literature. The Intelligence Studies subfield of International Relations focuses a large portion of its efforts on analytic methods.[15] Analytic methods are examined in great detail in an effort to minimize many of the cognitive, social, or rational pitfalls that occur when trying to assess events and provide intelligence to decision-makers. A sampling of these pitfalls includes clientitis, mirror-imaging, mindset, groupthink, polythink, linear analysis or any number of other failings that can and often do arise.[16] Analysts are taught to leverage a variety of techniques to avoid falling into these pitfalls. Although, as noted by Stephen Coulthart, a large percentage of analysts never use or find structured analysis unhelpful.[17] Despite limited application across various intelligence agencies, scholars and practitioners often regard the logic of structured analysis helpful. Moreover, the inclusion of standardized rigor into intelligence analysis and product development is not a novel concept and has been around in one form or another since the formal establishment of the Intelligence Community under the National Security Act of 1947. Some well-known examples of the evolving nature of intelligence analysis are robustly analyzed by Jim Marchio and indicate both the long-term importance and ongoing dialogue on issues of standardized analytic rigor and the consistent and recurring need to emphasize theoretical and methodological approaches.[18] The purpose of this section is not to delve into the details of analytic tradecraft, but rather to highlight the fact that it is a core attribute of intelligence analysis.

Increasingly, the collection and production of intelligence is being automated. The automation and integration of information collected from various non-human sources and platforms has been occurring since before the advent of the telegraph and continued to develop following the creation of the wireless. As communications became increasingly robust during the build-up to the Second World War ever larger volumes of data necessitated increasingly larger numbers of analysts to make sense of the information coming in. The complex structure of both encrypted and unencrypted data coming in required large numbers of computers. Although these were not computers in the modern sense. The term computer dates back to seventeenth century and typically referred to individuals who computed information.[19] The collection and accumulation of data requiring robust computation for war time needs grew extensively during the Second World War. Computers were responsible for intelligence functions such as codebreaking and for the computation of information related to the delivery of munitions. According to official records, the British codebreaking enterprise at one point had on staff more than 10,000 codebreakers each working to decipher encrypted message traffic.[20] It is also important to note that of these 10,000 codebreakers approximately 75% were female. Once Enigma was penetrated the combined decrypts amounted to more than 30,000 a month at the beginning of 1943 and nearly 90,000 a month by the end of that same year.[21] In the midst of a World War enormous human resources were allocated to read and process each of these messages, the combined organizational apparatus was immense. The scale and complexity of communications globally at the end of the war were significant. Yet, the coming expansion of communications mediums and the diversity of their use and applications were set to grow substantially in the decades following the war.

The development of a post war ordering of the world between the Soviet Union and the United States cemented the need for a robust national security community within the US. Truman, upon signing the National Security Act of 1947, enacted legislation that set in motion the creation and realization of an immense and growing security capacity unparalleled in US history. Signals intelligence was to be a central portion of the newly established security and defense community; however, it was not until the Brownell Committee's recommendation in NSCIB No. 9 on 24 October 1952 that the NSA was formally established.[22]

The first electric programmable computers Colossus Mark 1 and later Mark 2 demonstrated high levels of efficiency well in excess of their human computational counterparts. The creation of the NSA, the rise of electric programmable computers, nuclear weapons, expanding telegraph, telephone, and wireless communications resulted in a convergence of increasing volume and programmable analytic capacity. The use of computers to process large volumes of data was of critical importance in the early days of the cold war. Computers allowed for scientists to predict how various builds of nuclear weapons would perform. For instance, John Von Neumann's basic model for computer development was translated into the ENIAC and used in computing measurements associated with hydrogen bombs.[23] These measurements provided the foundational basis for intelligence assessments on yields of various classes of atomic weapons. Early programmable computers also continued be developed evolving from work begun in Bletchley Park and other codebreaking centers in the United States. Computers were uniquely suited to solving very specific computational problems within precise parameters. The evolution from human to machine computers is ongoing. The von Neumann architecture remains the basic computing architecture of most modern computers.

The history of intelligence and the history of computers in this section is unduly brief and fails to account for the robust evolution and increasing inclusion of computers in intelligence, yet it serves to illustrate that the progression from human to machine is not a new concept, but rather one that has been critical to the profession of intelligence since before the Second World War. The number of computers and their capacity today far exceeds even those of a couple decades ago. A modern iPhone is approximately 120 million times more powerful than the Apollo Mission computers and costs $3.5 million less per unit.[24] The increasing computational power of modern computers is opening a wide array of possibilities for the intelligence community to exploit. To put the evolution in context, the 10,000 codebreakers at Bletchley analyzed message traffic primarily from one adversary. At their height they were analyzing 90,000 messages a month. By comparison the number of billed minutes of telephone service between the US and Poland in 1980 was less than 1 million, less than 4 years later in 1984 that number was 10 million and by the time the Soviet Union collapsed that number of had surpassed 90 million.[25] Poland was representative of one of many Eastern Bloc intelligence targets in the 1980s. The growth in volume was not in the thousands or tens of thousands as encountered by Bletchley, but in the millions and tens of millions. A more recent example in arguably one of the most difficult countries to collect social media data in the world, Afghanistan, where there are more than 3.1 million active Facebook users posting millions of posts daily. To collect and analyze meaningful data on ever increasing streams of communications is beyond the capacity of human computers and even the most robust analysts. A tailored approach might be taken, by focusing on only specific persons or groups of interest, yet even within narrowed target sets the number of individuals or connections is likely to grow untenable. A 2014–15 collection on two Islamic State media outlets' Twitter accounts resulted in more than 42 thousand posts and more than 35 million potential viewers of those posts.[26]

The (human) computers of the Second World War would be overwhelmed, just as present day intelligence and law enforcement officials would likely be overwhelmed in the absence of computational and analytic assistance. A professional intelligence analyst can only realistically view, read, process and synthesize into a product a limited volume of information at any given time. Moreover, the vectors of information have diversified as have the number targets against which to collect. The next section addresses the role of analysts in an increasingly complex information environment and begins to identify the tools that are and will become available to assist in the development of intelligence products. The section also addresses the challenges these tools pose to analysts and begins assessing the critical role analysts play in providing a check against what some have termed weapons of math destruction.[27]

## The analysts and the machines

National security intelligence is comprised of multiple sources of overlapping information combined through analytic processes into products designed to 'eliminate or reduce uncertainty for decision-makers'.[28] As was alluded to above, the primary role of the analyst has been to take in large volumes of data and develop products that provide foreknowledge to decision-makers. JP 2–0 identifies eight different types of intelligence products with the context of the Department of Defense: warning intelligence, current intelligence, general military intelligence, target intelligence, scientific and technical intelligence, counterintelligence, estimative intelligence, and identity intelligence.[29] These products contain information that is compared, analyzed, and weighted often within and across agencies and departments.

Since the early days of the NSA, computers have played an increasingly important role in the development of intelligence products. Declassified document 3,575,750 provides some insight into the importance of digital computers to the evolution of the nation's preeminent signals intelligence agency.[30] Most of these computers were focused on cryptanalysis efforts.[31] As early as 1948, the precursors to the NSA and other COMINT agencies recognized the problems posed by labor intensive plain language analysis.[32] Despite the recognition of these problems, the implementation of complex learning mechanisms on digital computers was still largely in its nascent stages. The development of computer-based systems to facilitate intelligence analysis has been a long and arduous process, one that has increased in velocity in recent years. One of the first major steps towards contemporary machine learning was the creation of the Stochastic Neural-Analog Reinforcement Computer (SNARC) developed by Marvin Minsky in 1951 and later examined in more detail in his 1961 paper 'Steps Towards Artificial Intelligence'.[33] Minsky recognized that despite all the talk about computers being 'smart' they really only did what they were told to do. His research led to the creation of a computer and a method of analysis that fostered the combination of pattern recognition and learning. It was not until 1965 when Alexey Ivaknenkoa a Soviet mathematician and computer scientist wrote a pioneering work on Cybernetic Predicting Devices that the race to leverage computers for analysis beyond specific predictions on weapons yield or cryptanalysis began to emerge.[34] As the volume and complexity of sources increased intelligence agencies and their analysts were increasingly challenged by information overload.[35]

Margaret McDonald in writing on the challenge posed to intelligence agencies by the increasing volumes of information notes: 'the intelligence community no longer suffers from information scarcity but from information overload'.[36] She continues: 'Analysis must cover enormous quantities of data, in which valuable information may at best be implicit'.[37] Alan Dupont writes that information overload is a significant problem for intelligence analysts and managers.[38] The sources of information overload are not singular in nature. Instead they comprise collections from across all major intelligence agencies. The volume of every form of intelligence increased markedly in the post war era and was not confined to SIGINT. Virtually every form of Technical Intelligence from SIGINT, MASINT, and IMINT (now GEOINT) to include the emerging fields of CYBINT and SOCINT (Social Media Intelligence) are expanding at near exponential rates. The signal to noise ratio within this data is very low, and vast collections of data make analysis extremely difficult. Yet, despite criticisms from individuals such as William Binney, a former NSA Official,[39] the National Academy of Sciences in a 2015 report on bulk collection found that, while the capacity to do real-time intelligence analysis of large data streams was lacking, this is likely to change in the future, according to their conclusions.[40]

When theorizing on the future of intelligence it is impossible to ignore the vast quantities of data being generated. Moreover, the current state of data generation is but a fraction of the data that will be generated in the coming decades. John Pannerselvam, Lu Liu, and Richard Hill contend that agencies faced with an ever-increasing variety, volume, and velocity of data will struggle to see the big picture.[41] Despite robust structured analytic techniques and other methods, intelligence analysis will increasingly rely on assistive technologies.

The intelligence community and in particular the Office of the Director of National Intelligence is aware of the challenges faced. The Intelligence Advanced Research Projects Activity (IARPA), formed in 2006 and modeled on the Advanced Research Projects Agency, now known as the Defense Advanced Research Project Agency (DARPA) serves as a vehicle to fund high-risk, high payoff innovative technologies for 'future overwhelming intelligence advantage'.[42] IARPA funds a variety of research initiatives, some of the initiatives of relevance to large volumes of computational information include the ACE, ForeST, FUSE, and OSI programs. ACE or Aggregate Contingent Estimation was designed to focus on probabilistic assessments for contingent events, the aggregation of events by multiple human analysts and the representation of these forecasts and their distributions.[43] Forecasting Science and Technology (ForeST) was designed to fund projects that could accurately forecast significant advances or milestones in science and technology.[44] Foresight and Understanding from Scientific Exposition (FUSE) funded research to create a system to process, generate, and prioritize technical terms and areas and provide evidence of technical developments and advancements as they are emerging.[45] The Open

Source Indicators (OSI) program focused on developing methods for continuous, automated analysis of publicly available data to anticipate significant events.[46]

The programs listed above are only a few of the many programs within the intelligence community focusing on innovative solutions to large volumes of data. Outside of IARPA, In-Q-Tel serves as a market-based accelerator for cutting-edge technologies to facilitate national security. Combined these and other programs are focused keeping up with the increasing challenges faced by the intelligence community. At the core of each of these programs is the need to assess novel volumes and varieties of information and develop robust intelligence products.

At the root of the challenges facing intelligence analysis in the future is data. Data, bits of information, as Robert Kitchin writes,

> are commonly understood to be the raw material produced by abstracting the world in to categories, measures and other representational forms – numbers, characters, symbols, images, sounds electromagnetic waves, bits – that constitute the building blocks from which information and knowledge are created.[47]

The programs of IARPA and others are focused on providing technical solutions to an increasingly complex information environment.

The volume and diversity of data can increase the number of potential questions analysts can attempt to address. Although traditional methods dissuade exploratory data analysis (the analysis of data absent theory and hypotheses), big data makes such analysis more profitable and often informative towards advancing knowledge. Big data offers significant strengths associated with the diversity and specification of data. Inclusive data collection can provide varying degrees of nominal, ordinal, and interval measures. Because the volume, specificity, variety, and velocity (both speed and timeliness) of it are increased, often big data can provide more value in its exhaust (data captured as a residual) than in its deliberate data collections. This exhaust can prove useful to future analysis efforts. In particular, in addressing questions not yet formulated by a given client.

For example, Burt Monroe et al. highlight the fact that large data collection can offer insights into sub-populations.[48] Many of the complex questions fielded by intelligence agencies often reside within sub-populations and outside of easily accessible information sources. It is extremely difficult to capture historical data in the present, whether by survey on what an individual's perceptions of an event, particular policy, law, practice, or decision were, yet big data is increasingly building a historical repository of data on these perceptions in the form of social media posts, news archives, recorded emails, web-traffic logs and much more. Information captured in stream and stored for later use such as email communications, forums, chat logs, browser histories, and many other types of data can help to inform future analysis when novel questions and their subsequent informed hypotheses are developed.

Although many phenomena will still present significant data challenges, the ability to operationalize or capture information into robust data stores is likely to improve in the coming decades. Issues will likely still persist as the data available for collection is unbalanced between connected and unconnected nations, communities, and individuals. As the Internet and its associated technologies expand globally, the areas where data collection is currently lacking will be minimized. What should be recognized is that data availability is a strength that is likely to be a powerful driver of intelligence analysis.

It is convenient to use big data as a stand in for good intelligence analysis methods. Within the social scientific literature commentators have remarked that data obviates the need for theory and hypotheses.[49] The notion that large volumes of data in some way presupposes omniscience is false. Correlation is not a substitute for rigorous scientific method and robust causal models predicated on thoughtfully informed hypotheses. While human analysts make use of structured analytic techniques to minimize biases, the underlying attributes of data analysis for intelligence is likely to reside in robust theory.

Theory is fundamental to the scientific study of the world. Stephen Van Evera writes: 'theories are general statements that describe and explain the causes or effects of classes of phenomena. They are composed of causal laws or hypotheses, explanations, and antecedent conditions'.[50] Scientific study absent theoretical foundations is prone to systemic error. No error is likely to give more pause when encountering large data-sets than spurious relationships. Spurious relationships derived from data sources and assisted analysis are likely to result in weak intelligence products. Theory helps to establish the relationship between variables and their effect on a potential outcome. Theory guides all aspects of scientific study from the operationalization of concepts, to their use within quantitative or qualitative analysis. Just as structured analytic techniques facilitate the development of robust intelligence products, theory facilitates the collection and automation of analysis of large volumes of data. Both the human analyst and automated computer analysis stages of the intelligence process must be focused on developing accurate predictions about phenomena.

Analysis of data collected and processed through machine learning predicated on bad theoretical foundations is likely to lead to poor quality products that might hurt rather than help answer questions relevant to national security. Machine learning refers to the automated detection of meaningful patterns in data.[51] Machine learning is most commonly adaptive, meaning that it uses a priori information to inform future output. The parameters of this learning mechanism are as vital to automating intelligence analysis for computers as structured analytic techniques are to a human analyst mitigating biases.

Whether using frequentist or Bayesian methodologies, machine learning through algorithms, reduction of data-sets across dispersed clusters or any host of new means of examining data, the potential to introduce error into analysis is as great with big data as it is with small. Whether the error enters through the poor operationalization of a phenomena or in the search for relationships between data where none actually exists, the fundamental challenges of deriving meaning from data remain. One problem commonly faced in the application of big data is the derivation of meaning (knowledge) from two phenomena based on operationalized concepts that are, in reality, are not related. Luke Keele contends that irrespective of the amount of data available, science must necessarily be rooted in a concise understanding of the assumptions upon which analysis occurs.[52]

For machine learning to facilitate intelligence analysis, the algorithms underpinning the learning structures must consider a variety potential pitfalls in much the same way that structure analytic techniques function for intelligence analysts. However, with the increasing incorporation of machine learning the avoidance of biases within algorithms will become increasingly important. Algorithms that are predicated on biases within learning structures are self-reinforcing and will produce progressively less accurate analysis. Whereas, a human analyst can correct for bias between analyses automated learning algorithms absent oversight and corrective adjustments might result in increasingly significant error. Often these errors can arise inadvertently during the training phase of machine learning due to biased data inputs, alternatively they can occur later in the utilization of a particular algorithm as data structures or exogenous conditions change. As more and more data are increasingly processed through machines and then provided to analysts, it is also vital that analysts have insight into how the algorithms function to provide them with information for further analysis. One of the most famous examples outside of the intelligence community was Google's 'Flu Trends' project.[53] The project suffered from systemic problems because the collection and interpretation of data inputs was dynamically based on a changing Google Search algorithm while the learning mechanism for predicting flu outbreaks was static. A similar MASINT collection system that altered the weighting of inputs from CBRN materials without altering the subsequent interpretation of those weights in the learning mechanism would likely result in either false positives or false negatives that might inadvertently lead to conflict. In a complex systems environment of intelligence understanding having insight into how systems collect and analyze data is valuable and reduces the likelihood of intelligence failures.

As the volume and velocity of data expands exponentially, the intelligence community will increasingly rely on machine learning. The trend towards reliance on machine learning has been evolving since 1947. As intelligence agencies move towards ever more complex uses of machine learning, expanding beyond basic pattern recognition and cryptanalysis the increasing complexity of the systems upon which human analysts will rely will pose problems for mitigating embedded biases. The next section builds on the progress made towards the inclusion of machine learning and focuses on where the law and policies associated with big data and machine learning for intelligence are and where they are likely headed.

## The Dutch boy and the dike: laws and policies on big data and machine learning

The previous sections of this paper have built a case for the ever-increasing importance of big data and in particular machine learning in intelligence analysis. The context provided above, albeit brief, establishes the framework for a world in which everything that can be collected can also be incorporated into intelligence analysis. The deluge of data is rapidly outpacing laws and policies on intelligence for national security purposes and the problem is likely to become worse in the coming decades.

The synthesis of multiple streams of data into a meaningful product for the purpose of foreknowledge of phenomena has been one of the core functions of the intelligence community. When this collection was conducted against signals broadcast by adversary nation states or images taken from satellites from air and orbital assets or collected through the measurement of signatures there remained significant numbers of ambiguities to cause frequent consternation within oversight entities. The significance of debates during the Church and Pike Committees on the extent of intelligence and potential violations including domestic spying, the opening of mail within the United States and a host of other allegations remain pertinent today, but the scale and complexity of problems has reached new heights.

The Snowden leaks revealed in *The Guardian* newspaper and by the *Washington Post* highlight the many challenges faced in a world of increasingly ubiquitous information assets available to intelligence agencies.[54] The details of these leaks make clear that the United States intelligence community and its allies were continuing the progression towards the incorporation of ever more copious amounts of data and leveraging novel machine learning methods to exploit that data for intelligence purposes.

External to the Intelligence Community the opportunities and challenges posed by Big Data warranted two separate public reports by the Obama Administration in 2014 and 2016.[55] Both these reports begin to address many of the legal and policy challenges that big data and machine learning pose to American's outside of their exploitation by the intelligence community. Among the various issues identified is the removal of privacy, the durability of data, the biasing of analyses and much more. Many of the problems faced external to the IC are found within the IC. These two reports serve as policy statements on the responsible use of data for non-intelligence purposes. As these reports note, the sources of data that can be collected for both public and private use are numerous.

Although the sources of data available for collection are growing the intelligence community must adhere to laws and policies governing the collection of intelligence. There are three basic categories of authorities defined within US Code. These roughly constitute Military (Title 10), Law Enforcement (Title 18) and Intelligence (Title 50). As Andru Wall notes, that in particular the debate between Title 10 and Title 50 constitutes the 'epitome of an ill-defined policy debate'.[56] Yet within cyberspace and the collection of data from networks the largest distinction is between the collection of intelligence under Title 10 and Title 50, and Title 18 constituting rules associated with crimes and criminal procedure. As the Snowden leaks reveal and the later NSA report demonstrate there is a fine line between legally allowable intelligence activities focused external to the United States and activities more commonly associated with criminal procedures that occur internal to the United States.[57] Moreover, the line and distinction is further confused when one considers the reports from the Obama administration discussing the reasonable use and development of big data within the US for governmental use.

Whereas during the cold war it was reasonable and responsible to highly differentiate the signals traffic external to the United States from that of traffic internal to the United States TCP/IP and other core protocols of the Internet make the origin of data sources increasingly difficult. Moreover, the problem of origin is further compounded by encryption or other obfuscation techniques that mask data. Constraints on the collection of intelligence are codified across multiple laws and policies, many of which offer confusing interpretations of what is and is not legal. Among the most pertinent laws and sections of those laws of relevance are the National Security Act of 1947, the Foreign Intelligence Surveillance Act (Section 702), The Patriot Act (Sections 215 and 216). The basic intent of these laws is to protect American's from undo surveillance by intelligence agencies and to ensure that activities focused against American adhere to due process rules and rights.

To adhere to these rules in an evolving collection environment, the systems which collect and analyze data through machine learning functions must necessarily be trained to avoid unauthorized collection and analysis. Herein law and policy, rightly or wrongly, introduces inefficiencies into the evolution of computing and analysis and makes legal and or ethical judgements on what should and should not be collected and analyzed. Amitai Etzioni and Oren Etzioni in discussing artificial intelligence, a macro field associated of computational analysis that often heavily relies on machine learning, write that it is necessary that as more and more systems are automated and they becoming increasingly 'smarter' these systems will require guardians entrusted with ensuring values, ethics, laws, and policies are adhered to.[58] Devdatt Dubhashi and Shalom Lappin implore conduct an 'intelligent discussion on the nature of change and the implementation of policies'.[59]

The progression from thousands of messages being analyzed at Bletchley through the creation of the NSA and the increasing need to seek out better and more efficient ways to both collect and analyze large scale data is less constrained by data and math than by law and policy in the near term. The laws and policies associated with the collection and analysis of data are almost exclusively focused on governmental use. At the same time that governments are seeking solutions to pressing national security issues and running into controversy on issues such as violations of law or privacy considerations private corporations such as Google, Facebook, Twitter and others are profiting on every bit of data excreted. Whole markets of data exploitation are emerging in which information is bought and sold to increase the sales of products or make better business decisions. As everything increasingly becomes of value to national security intelligence, the value is not in the individual data points, but in their ability to be aggregated, mined and analyzed. Intelligence theory remains as vital today as it has ever been. Its importance is in being able to shape and structure the algorithms that will inevitably play a larger and larger role in intelligence analysis. The extrapolation from small numbers of data points using human analysts, to the inclusion of machines into a hybrid human-machine forecasting mix and potentially one day to a machine dominated field requires intelligence studies scholars to increasingly incorporate math and science into the art of intelligence. As a field intelligence analysis is at the dawning of a new era, a digital era in which information collected from everything from toasters and coffeemakers can now be incorporated with human sources. The melding of multiple sources to form truly all-source intelligence should not happen in the absence of theory and oversight from existing practitioners.

## Notes

1. Markram, "The Human Brain Project".
2. "1st LD-Writethru-China Headlines".
3. Warner, "Wanted," 18–22.
4. Gill et al., *Intelligence Theory*.
5. Johnson, "Bricks and Mortar," 1–28.
6. Marrin, "Intelligence Analysis Theory," 821–46.
7. Scott and Jackson, "The Study of Intelligence in Theory and Practice," 139–69.
8. Honig, "A New Direction for Theory-Building," 699–716.
9. http://www.internetlivestats.com/internet-users/.
10. Evans, "The Internet of Things".
11. Tillman, "How Many Internet Connections".
12. Nordrum, "Popular Internet of Things Forecast".
13. "Edward Snowden".
14. Agrell, "When Everything Is Intelligence".
15. Clark, *Intelligence Analysis*; and Hedley, "Analysis for Strategic Intelligence".
16. Hedley, "Analysis for Strategic Intelligence," 221; and Mintz and Wayne, *The Polythink Syndrome*.
17. Coulthart, "Why Do Analysts Use Structured Analytic Techniques?," 1–16.
18. Marchio, "Analytic Tradecraft and the Intelligence Community".
19. "computer, n.". OED Online.
20. McKay, *The Secret Lives of Codebreakers*, 6.
21. Hinsley and Stripp, *Codebreakers*, 144.
22. Howe, *The Early History of NSA*.

23. Rhodes, *The Making of the Atomic Bomb*, 772.
24. Puiu, "Your Smartphone Is Millions of Times".
25. Blake and Lande, *Trends in the U.S. International Telecommunications Industry*.
26. Brantly, "Innovation and Adaptation in Jihadist Digital Security".
27. Oneil, *Weapons of Math Destruction*.
28. Clapper, "Remarks, Association of Former Intelligence Officers".
29. Joint Publication 2–0 Joint Intelligence.
30. "Before Super-Computers".
31. Burke, *It Wasn't All Magic*.
32. Ibid., 263.
33. Minsky, "Steps Toward Artificial Intelligence," 1–23.
34. Ivakhnenko and Lapa, *Cybernetic Predicting Devices*.
35. Betts, "Analysis, War, and Decision," 93.
36. MacDonald and Oettinger, "Information Overload," 44.
37. Ibid.
38. Dupont, "Intelligence for the Twenty-First Century".
39. Whittaker, "NSA Is So Overwhelmed with Data".
40. National Research Council, *Bulk Collection of Signals Intelligence*.
41. Pannerselvam, Liu, and Hill, "An Introduction to Big Data," 20.
42. https://www.iarpa.gov/index.php/working-with-iarpa/open-solicitations/incisive-analysis-office-wide-baa?highlight=WyJvdmVyd2hlbG1pbmciLCJpbnRlbGxpZ2VuY2UiLCJhZHZhbnRhZ2UiLCJvdmVyd2hlbG1pbmcgaW50ZWxsaWdlbmNlIiwib3ZlcndoZWxtaW5nIGludGVsbGlnZW5jZSBhZHZhbnRhZ2UiLCJpbnRlbGxpZ2VuY2UgYWR2YW50YWdlIl0=.
43. Aggregative Contingent Estimation (ACE), https://www.iarpa.gov/index.php/research-programs/ace.
44. Forecasting Science and Technology (ForeST), https://www.iarpa.gov/index.php/research-programs/forest.
45. Foresight and Understanding from Scientific Exposition (FUSE), https://www.iarpa.gov/index.php/research-programs/fuse.
46. Open Source Indicators (OSI), https://www.iarpa.gov/index.php/research-programs/osi.
47. Kitchin, *The Data Revolution*.
48. Monroe et al., "No! Formal Theory," 71–4.
49. Anderson, "The End of Theory; Grimmer"; and Grimmer, "We Are All Social Scientists Now," 80–3.
50. Van Evera, *Guide to Methods for Students of Political Science*.
51. Shalev-Shwartz and Ben-David, *Understanding Machine Learning*, XV.
52. Keele, "The Discipline of Identification," 102–6.
53. Lazer et al., "The Parable of Google Flu".
54. MacAskill and Dance, "NSA Files Decoded"; and Gellman and Poitras, "U.S., British Intelligence Mining Data".
55. "Big Data: Seizing Opportunities, Preserving Values"; and "Big Data: A Report".
56. Wall, "Demystifying the Title 10-Title," 86.
57. Morell et al., *The NSA Report*.
58. Etzioni, and Etzioni, "Designing AI Systems That Obey Our Laws," 29–31.
59. Dubhashi and Lappin, "AI Dangers," 43–5.

## Acknowledgement

Special thanks to Peter Gill and Hamilton Bean for review and revision support.

## Disclosure statement

The views expressed are those of the author and do not reflect the official policy or position of West Point, the Department of the Army, the Department of Defense, or the US Government.

## Notes on contributor

***Aaron F. Brantly*** is an assistant professor of Political Science in the Department of Political Science at Virginia Polytechnic and State University, Cyber Policy Fellow at the Army Cyber Institute, and Non-Resident Cyber Fellow for the Combating Terrorism Center at West Point.

## ORCID

*Aaron F. Brantly* http://orcid.org/0000-0003-4193-3985

## Bibliography

"1st LD-Writethru-China Headlines: Sunway-TaihuLight Outperforms Tianhe-2 as World's Fastest Supercomputer." *Xinhua News Agency*, June 20, 2016. Infotrac Newsstand.

Agrell, Wilhelm. "When Everything Is Intelligence – Nothing Is Intelligence – Central Intelligence Agency." Central Intelligence Agency, October 2002. https://www.cia.gov/library/kent-center-occasional-papers/vol1no4.htm.

Anderson, C. "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete." June 23, 2008. http://www.wired.com/2008/06/pb-theory/.

"Before Super-Computers: NSA and Computer Development." National Security Agency, 2009. https://cryptome.org/0002/nsa-computers.pdf.

Betts, Richard K. "Analysis, War, and Decision: Why Intelligence Failures are Inevitable." In *Intelligence Theory: Key Questions and Debates*, edited by Peter Gill, Stephen Marrin, and Mark Phythian, 93. New York: Routledge, 2009.

"Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights." Executive Office of the President, 2016.

"Big Data: Seizing Opportunities, Preserving Values." Executive Office of the President, 2014. https://obamawhitehouse.archives.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf.

Blake, Linda, and Jim Lande. *Trends in the U.S. International Telecommunications Industry*. Washington, DC: Federal Communications Commission, 1998.

Brantly, Aaron. "Innovation and Adaptation in Jihadist Digital Security." *Survival* 59, no. 1 (2017): 79–102.

Burke, Colin B. *It Wasn't All Magic: The Early Struggle to Automate Cryptanalysis 1930s – 1960s*. Ft. Meade: Center for Cryptologic History at the National Security Agency, 2002.

Clapper, James R. "Luncheon Remarks, Association of Former Intelligence Officers." In The Intelligence, AFIO newsletter, McLean, VA (October 3, 1995) cited in Loch K. Johnson *The Oxford Handbook of National Security Intelligence*. Oxford: Oxford University Press, 2012.

Clark, Robert M. *Intelligence Analysis: A Target-Centric Approach*. Washington, DC: CQ Press, 2004.

"computer, n.". OED Online. Oxford University Press, December 2016. Accessed February 12, 2017. http://www.oed.com/view/Entry/37975?redirectedFrom=Computer

Coulthart, Stephen. "Why Do Analysts Use Structured Analytic Techniques? An in-Depth Study of an American Intelligence Agency." *Intelligence and National Security* 33, no. 2 (2016): 942.

Dubhashi, Devdatt, and Shalom Lappin. "AI Dangers: Imagined and Real." *Communications of the ACM* 60, no. 2 (2017): 43–45.

Dupont, Alan. "Intelligence for the Twenty-First Century." *Intelligence and National Security* 18, no. 4 (2003): 15–39.

"Edward Snowden: Leaks That Exposed US Spy Programme – BBC News." *BBC.com*, January 17, 2014. http://www.bbc.com/news/world-us-canada-23123964.

Evans, Dave. "The Internet of Things." CISCO Internet Business Solutions Group, 2011. http://www.cisco.com/c/dam/en_us/about/ac79/docs/innov/IoT_IBSG_0411FINAL.pdf.

Etzioni, Amitai, and Oren Etzioni. "Designing AI Systems that Obey Our Laws and Values." *Communications of the ACM* 59, no. 9 (2016): 29–31.

Gellman, Barton, and Laura Poitras. 2013. "U.S., British Intelligence Mining Data From Nine U.S. Internet Companies in Broad Secret Program." *The Washington Post*, June 7.

Gill, Peter, Stephen Marrin, and Mark Phythian. *Intelligence Theory: Key Questions and Debates*. London: Routledge, 2009.

Grimmer, Justin. "We Are All Social Scientists Now: How Big Data, Machine Learning, and Causal Inference Work Together." *PS Political Science and Politics* 48, no. 1 (2014): 80–83.

Hedley, Jason H. "Analysis for Strategic Intelligence." In *Handbook of Intelligence Studies*, edited by Loch K. Johnson. London: Routledge, 2010.

Hinsley, F. H., and Alan Stripp. *Codebreakers: The Inside Story of Bletchley Park*, 144. New York: Oxford University Press, 2011.

Honig, Or Arthur. "A New Direction for Theory-Building in Intelligence Studies." *International Journal of Intelligence and Counter Intelligence* 20, no. 4 (2007): 699–716.

Howe, George F. "The Early History of NSA." *Cryptologic Spectrum* 4, no. 2 (1974): 17. https://www.nsa.gov/public_info/_files/cryptologic_spectrum/early_history_nsa.pdf

Ivakhnenko, A. G., and Valentin Grigor'evich Lapa. *Cybernetic Predicting Devices*. New York: CCM Information Corp, 1973.

Johnson, Loch. "Bricks and Mortar for a Theory of Intelligence." *Comparative Strategy* 22, no. 1 (2013): 1–28.

Joint Publication 2-0 Joint Intelligence. Joint Chiefs of Staff. US Department of Defense, 2013. http://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp2_0.pdf

Keele, Luke. "The Discipline of Identification." *PS Political Science and Politics* 48, no. 1 (2014): 102–106.

Kitchin, R. *The Data Revolution*. Thousand Oaks, CA: SAGE, 2014.

Lazer, D., R. Kennedy, G. King, and A. Vespignani. "The Parable of Google Flu: Traps in Big Data Analysis." *Science* 343, no. 6176 (2014): 1203–1205.

MacAskill, Ewen, and Gabriel Dance. "NSA Files Decoded: Edward Snowden's Surveillance Revelations Explained." *The Guardian*, November 1, 2013.

MacDonald, Margaret S., and Anthony G. Oettinger. "Information Overload: Managing Intelligence Technologies." *Harvard International Review* 24, no. 3 (2002): 44–48. Harvard International Relations Council.

Marchio, Jim. 2014. "Analytic Tradecraft and the Intelligence Community: Enduring Value, Intermittent Emphasis." *Intelligence and National Security* 29, no. 2 (2014): 159–183.

Markram, Henry. "The Human Brain Project." *Scientific American* 306, no. 6 (2012): 50–55.

Marrin, Stephen. "Intelligence Analysis Theory: Explaining and Predicting Analytic Responsibilities." *Intelligence and National Security* 22, no. 6 (2008): 821–846.

McKay, Sinclair. *The Secret Lives of Codebreakers: The Men and Women Who Cracked the Enigma Code at Bletchley Park*, 6. New York: Plume, 2012.

Minsky, Marvin. "Steps Toward Artificial Intelligence." Proceedings of the IRE, January 1961.

Monroe, Burt L., Jennifer Pan, Margaret E. Roberts, Maya Sen, and Betsy Sinclair. "No! Formal Theory, Causal Inference, and Big Data Are Not Contradictory Trends in Political Science." *PS Political Science and Politics* 48, no. 1 (2014): 71–74.

Morell, Michael J., Richard A. Clarke, Geoffrey R. Stone, Peter P. Swire, and Cass R. Sunstein. *The NSA Report: Liberty and Security in a Changing World*. Princeton: Princeton University Press, 2014.

Mintz, Alex, and Carly Wayne. *The Polythink Syndrome: U.S. Foreign Policy Decisions on 9/11, Afghanistan, Iraq, Iran, Syria, and ISIS*. Stanford: Stanford University Press, 2016.

National Research Council. *Bulk Collection of Signals Intelligence: Technical Options*. Washington, DC: The National Academies Press, 2015.

Nordrum, Amy. "Popular Internet of Things Forecast of 50 Billion Devices by 2020 Is Outdated." *IEEE Spectrum*, August 2016.

Oneil, Cathy. *Weapons of Math Destruction How Big Data Increases Inequality and Threatens Democracy*. Oxford: Broadway Books, 2017.

Pannerselvam, John, Lu Liu, and Richard Hill. "An Introduction to Big Data." In *Application of Big Data for National Security*, edited by Babak Akhgar, Gregory B. Saathoff, Hamid R. Arabnia, Richard Hill, Andrew Staniforth, and Petra Saskia Bayerl, 20. Oxford: Butterworth-Heinemann, 2015.

Puiu, Tibi. "Your Smartphone Is Millions of Times More Powerful That All of NASA's Combined Computing in 1969." *Zmescience.com*, October 12, 2015. http://www.zmescience.com/research/technology/smartphone-power-compared-to-apollo-432/.

Rhodes, Richard. *The Making of the Atomic Bomb*, 772. London: Simon & Schuster, 2012.

Scott, Len, and Peter Jackson. "The Study of Intelligence in Theory and Practice." *Intelligence and National Security* 19, no. 2 (2004): 139–69.

Shalev-Shwartz, Shai, and Shai Ben-David. *Understanding Machine Learning: From Theory to Algorithms*. New York: Cambridge University Press, 2016, XV.

Tillman, Karen. "How Many Internet Connections Are in the World? Right. Now." *CISCO Blogs*. Accessed July 29, 2013. http://blogs.cisco.com/news/cisco-connections-counter.

Van Evera, Stephen. *Guide to Methods for Students of Political Science*. Ithaca, NY: Cornell University Press, 1997.

Wall, Andru E. "Demystifying the Title 10-Title." *Harvard National Security Journal* 3 (March 2014): 86.

Warner, Michael. "Wanted: A Definition of 'Intelligence'." *Studies in Intelligence* 46, no. 3 (2017): 18–22.

Whittaker, Zack. "NSA Is So Overwhelmed with Data, It's No Longer Effective, Says Whistleblower | ZDNet." *ZDNet*, April 27, 2016. http://www.zdnet.com/article/nsa-whistleblower-overwhelmed-with-data-ineffective/.