WAR ON THE ROCKS

THE ABCS OF AI-ENABLED INTELLIGENCE ANALYSIS

IAIN J. CRUICKSHANK

FEBRUARY 14, 2020

SPECIAL SERIES - AI AND NATIONAL SECURITY

Editor's Note: This article was submitted in response to the call for ideas issued by the co-chairs of the National Security Commission on Artificial Intelligence, Eric Schmidt and Robert Work. It addresses the third question (part a.) on the types of artificial intelligence research the national security community should be doing.

From July 2014 to April 2015, a period of about 10 months, experts estimate there were 23 million tweets involved in the self-proclaimed Islamic State's online marshalling of support and influence operations. These tweets contained critical information about the group's leadership, information narratives, and even indications of tactical activities. While the Islamic State didn't tweet its way into Mosul, this open-source data was of significant intelligence value. But it's impossible for any given analyst to sort and understand 23 million tweets manually. This illustrates the dilemma that recent advances in technology pose for traditional methods of intelligence analysis: The digitization of human society has made huge amounts of information available for analysis. This information comes from an ever-increasing number of sources, like online social networks, digital sensors, or ubiquitous surveillance, and has been increasingly useful for intelligence. Too much information is being produced too quickly for an intelligence analyst to even comprehend it using current analysis techniques and software, much less derive meaningful intelligence from it or verify its veracity.

The changing information environment will force the conduct of military intelligence analysis to change too. This change cannot simply be the acquisition of some new analysis software or implementation of a new policy, but rather must be more comprehensive changes across all military intelligence organizations. To meet the new realities of the information environment, and by corollary the new realities of intelligence analysis, the whole of military intelligence needs to modernize in three areas. First, military intelligence organizations like the Army G-2, the J-2, and Futures Command should continue modernizing the tools and infrastructure supporting intelligence analysis and make these changes more broadly available to the force. Second, the military

intelligence schoolhouse ought to update how it trains and develops intelligence analysts. Third, military intelligence research organizations — like Intelligence Advanced Research Projects Activity and elements within U.S. Army Intelligence and Security Command — need research into potential disruptive technologies to maintain the integrity of intelligence analysis.

Adopting Data-Centric Systems

The first, and arguably easiest, area for modernization of intelligence analysis is to move towards data-centric systems for analysis for the entire force. At first glance it would seem the military already has and is pursuing data-centric technologies. And it is indeed the case that some organizations in the Department of Defense are developing cutting-edge data-centric systems for some types of intelligence analysis. In particular, groups like the Joint Artificial Intelligence Center, Special Operations Command, and the Army G-2 have been working on implementations of machine learning systems to be used in some intelligence analysis tasks. There are also several contract opportunities posted by the Department of Defense for developing AI solutions to military problems, including problems for military intelligence.

However, much of the work remains confined to specific intelligence problems like object identification in imagery, or are only available to certain organizations such as the special operations community. Additionally, many of these AI solutions are not configurable by the users. They are "black-box" software applications and, as such, the machine learning algorithms cannot be retrained on new data by a user.

Why does this matter? Imagine having a computer program that does a great job at detecting military vehicles from satellite imagery. Then the organization's mission changes to counter-insurgency, and it now needs a program to detect individuals carrying weapons on foot. If the computer program is closed and not configurable, you would have to contract out or otherwise build a brand-new

program. Whereas if the program were configurable you could just change the machine learning algorithms, or even just retrain the algorithms on new data from the new scenario. Further, many of the intelligence analysis tools and infrastructure available to the whole force are simply not data-centric by any of the principles of data-centric technologies (e.g. none of the algorithms or computational tools can be modified for different data scenarios). It is time for all operational military intelligence units to pursue data-centric technologies for their analysis systems, and move beyond applying AI to specific intelligence problems in an ad-hoc and one-off manner.

There are two key concepts to any data-centric system: First, analysis tools and applications should change with the data, and second, data should be easily accessible. Analysts must be able to configure the tools and algorithms of the systems to meet the realities of the battlefield, and data access should be as seamless as possible.

Within a data-centric context, the use of machine learning algorithms has led to breakthroughs in nearly every analysis endeavor, from fraud detection to image identification. To take advantage of these advances, intelligence analysts need systems that allow them to use computational tools and to constantly adjust, or retrain, their algorithms to a changing battlefield. Unfortunately, nearly all analysis software products in use today — including advanced systems like Palantir or Analyst Notebook — are closed systems that do not allow analysts to code custom algorithms, use the latest machine-learning algorithms, use the latest research in "explainable AI," or even allow analysts to provide feedback to the software's algorithms.

The inability to adjust analysis tools to the operational environment is a prodigious problem. Every battlefield scenario is unique in some respect and will therefore produce different information. What is more, the battlefield is a dynamic place; an operation can begin with a conventional tank battle and then quickly transition to an urban infantry fight. Furthermore, military intelligence

analysts are called upon to analyze and provide intelligence estimates for everything from tactical missions and operations planning to long-term strategic plans. Given these realities, there will never be a particular algorithm or set of data that will always work to produce the best battlefield intelligence. (In fact, this a direct result of the foundational "No Free Lunch theorem.") So, intelligence analysis cannot solely rely on an analysis system that treats algorithms and data as a black box. Nor can it solely rely on contracting for all the AI solutions to intelligence problems, or rely on software that requires constant tuning from contractors, and hope to keep pace with realities on the battlefield.

There is no one right way to accomplish the goal of giving analysts data-centric systems that can keep pace with changes on the battlefield. There are ongoing efforts and proposals to address this issue. At a minimum, however, any analysis system meant for all military intelligence units ought to allow analysts to code in the common languages used for data analysis and machine learning (i.e. Python, R, etc.). The analysis system can have special contractor-specific algorithms or interfaces to solve specific intelligence analysis problems, but it should allow for coding. While the ability to write code will entail some additional risk to both the computational resources and data — through things that can occur with data programming like program bugs or data corruption — careful oversight of the environments and use of virtualization can mitigate a lot of this risk. What's more, having analysts that can configure their tools and algorithms to the battlefield environment will decrease operational risk. Configuring an algorithm to a set of data requires you to understand how well that algorithm is performing on that data. And, if you have a good understanding of how well the algorithm is performing on a particular set of data, it will make it that much easier for you to have confidence (or uncertainty) in the predictions being output by the program. Ultimately, military intelligence analysts and organizations need computational tools that are both flexible and powerful enough and this can only come from data-centric systems that support programming in appropriate languages.

Information Storage

To support the modernization of analysis tools, the military intelligence data-storage infrastructure is also in sore need of modernization. The fuel of all machine learning algorithms, and really intelligence analysis itself, is data. Digital data is most useful when it is stored in a way that maintains some kind of consistent format (i.e., all data entries have the same fields of information: date, time, location, for example) and can be easily accessed (i.e., can be queried from the same programming environment where the analysis is taking place: the analyst doesn't have to launch a new program or window to search the database). However, many current databases of information used by military intelligence analysts make query of data difficult and they have no standardized data formats or documentation. Often, an intelligence analyst will have to navigate to a half-dozen information sources that only allow for manual click-through menus to search for information on them. Then they will have to download any information as a bunch of .csv files, manually fix the formatting errors between the files such as differences in date formats — just to get usable data. This severely inhibits an analyst from using advanced, computational tools as much of their time will necessarily be spent manually downloading and formatting data.

Intelligence information storage systems need to be more accessible for analysis. One way to do this is to provide policies so that certain standards are observed by all intelligence databases. For example, any given repository of information should have clear documentation on what resides within it. That way, an analyst can quickly understand how to craft queries to get the information they are after. Furthermore, databases should also endeavor to have programmatic access to information repositories.. This allows an analyst to quickly download large volumes of information right into their analysis environment rather than fumbling around with manually downloading, formatting, and then uploading the same information. It would also be worthwhile to create and maintain a directory of all of the databases of information that exist. While these steps are not a panacea, they are simple, concrete first steps to addressing current data issues within military intelligence organizations. Ultimately, without

modernizing how the community stores and accesses information, many advanced tools for intelligence analysis will be hard to use.

Managing Analysts

Modernization also requires new training and management of intelligence analysts. Analysts need to know how to handle massive amounts of digital data, which requires some programming skills and basic data science skills. Some would argue that there simply isn't enough time to train these kinds of skills in addition to everything else a military intelligence analyst needs to know. However, if one considers how much time intelligence analysts spend manually handling data, military organizations can't afford *not* teach these skills to their intelligence analysts. Imagine being an intelligence analyst tasked with analyzing threats and information warfare on social media for your area of operations, which could be as large as a country or region of the world. And imagine you have no idea how to use regular expressions or how to parse data files, meaning you could only use rough keyword searches and manually scan every result. All of your time would be spent in just manually looking through search results — most of which may be completely useless — and not thinking about the adversary or developing actionable intelligence. There is, and will continue to be, too much information for analysts to parse and sort manually, and analysts must be equipped with skills in data programming to handle this information deluge.

Talent management for intelligence analysts should also include their ability to handle and leverage digital data in their analyses. As analysts improve their ability to handle digital information, they will need advanced training and schooling in data visualization and, for more experienced analysts, machine learning. This need may necessitate changes to the skills trained at advanced analysis schools. Furthermore, the varying level of digital data skills across analysts will also require some means of tracking for appropriate career management. Functional Area 49 in the U.S. Army has already implemented a "data scientist" skill identifier. The CIA has full data science career tracks for

intelligence analysts. Something similar will be needed for intelligence analysts within the military. As the skills for analysis shift, so too must the development and management of intelligence analysts shift.

Research Technologies Relevant to Intelligence

Research and development arms of military intelligence and organizations like Army Futures Command ought to conduct practical research into emerging technologies and trends that will likely significantly impact intelligence analysis. Nearly all machine learning research today, while often funded by the Department of Defense, has been driven by commercial, medical, and academic problems. It's unclear how much of this research will translate into military applications, especially since military applications have unique and critical ethical considerations. Many commercial machine learning algorithms are designed to work by training on large volumes of data without any regard for things like the nationality of the individuals producing the data. Most commercially used machine learning algorithms are not tied to life-or-death decisions, so the assumptions underpinning these algorithms may not meet standards for ethical military use. Thus, there is a need for research, within the military intelligence community, into what machine learning and other AI technologies work for specific military applications and their specific ethical employment.

Military intelligence analysis would also greatly benefit from research into new, disruptive technologies. Two disruptive technologies in particular are adversarial machine learning and the production of believable, artificial data. Adversarial machine learning, or learning how to fool machine learning algorithms into making wrong predictions, is a fast-growing field of research. If one considers the parallel growth of surveillance and machine learning for intelligence purposes in many other nations, adversarial machine learning could be hugely disruptive (and likely a great boon for things like special forces operations). Adversarial machine learning will likely be critical to all things intelligence, concealment, and

deception and so should be researched by those that it will impact. Similarly, machine learning algorithms like Generative Adversarial Networks have shown great ability to generate fake data, including video and audio, which looks entirely real to a human observer. Since data is the fuel of intelligence analysis, it follows that there is a need to also research these new technologies in order to preserve the integrity of any analysis. Military intelligence research organizations should conduct their own research into how to counter these threats and understand their impact upon intelligence analysis, in order to preserve its effectiveness.

The operational environment confronting intelligence analysts is undergoing an accelerating digitalization. As such, military intelligence analysts and organizations are confronting new problems of data volume, velocity, and veracity which necessitate a comprehensive modernization of military intelligence organizations. In particular, the military needs force-wide, data-centric tools and infrastructure for intelligence units, training of intelligence analysts in digital data handling skills, and research by military intelligence research organizations on the impact and mitigation of adversarial machine learning and digital fake data generation.

Capt. Iain J. Cruickshank is a Ph.D. candidate in societal computing at Carnegie Mellon University as a National Science Foundation graduate research fellow. His previous assignments include company commander for D Company, 781st Military Intelligence Battalion (Cyber), and sub-element lead for planning and analysis and production on a national mission team in the Cyber National Mission Force. This article was produced in conjunction with the Defense Entrepreneur Forum's Gutenberg Writer's Collaborative.

The views expressed in this article are those of the author and do not represent those of the U.S. Army, the U.S. military, or the U.S. government.

Image: U.S. Army Combat Capabilities Development Command (Photo by Edric Thompson)

SPECIAL SERIES, AI AND NATIONAL SECURITY