How to Win with Data: The US SOF-Cyber Partnership Supporting Ukraine

Description

Winning the information war has been a strategic enabler for Ukraine since the Russian offensive began on February 24, 2022. Shortly after the invasion, various entities in Ukraine began systematically flooding Western news and social media with highlights of Ukrainian national resistance and tactical successes. These stories were sometimes false and often debunked, like the one about the â??Ghost of Kiev,â?• a mythical Ukrainian fighter pilot who ruled the skies over the capital city. True or false, the stories knit together a tale of resistance, leadership, and early military successes that galvanized Western support and aid. Today, many Western countries, including the United States, are leveraging all elements of national power (diplomatic, information, military, and economic) to help Ukraine repel Russian assaults. However, the massive outpouring of foreign aid and support may not have materialized if Ukraine had failed to win the information war in the West.

The war in Ukraine highlights informationâ??s important role in modern conflict; military and civilian leaders must understand the tactical, operational, and strategic implications of the information dimension in combat. For many reasons, the information dimension presents several complex problems that are difficult to answer. For example, which narratives will stick or go viral? How do narratives affect different audiences? How do adversaries adjust their narratives accordingly? What are the second- and third-order effects of releasing classified information? Adding to that complexity is the fact that most information operations are played out over several social and traditional media outlets and are largely dependent on the data that underlies those platforms â?? namely the commercial data, or data that is proprietary and commercialized by a company. Without access to commercial data research and analysis on the role of information in war is difficult and largely incomplete. To better understand how the information dimension can be leveraged successfully in modern warfare, the Department of Defense (DoD) should invest in data-centric lines of effort that exploit advances in machine learning, artificial intelligence, and computational social science to have an impact on the conventional and irregular battlefields and beyond.

Why Data Is Critical

Many pundits have referred to data as â??the new oil.â?• Dataâ??s importance to DoD was most recently highlighted in the National Defense Strategy and in official comments from senior leaders.

Secretary of the Army Christine Wormuth identified data-centric operations as a top objective for the US Army because the information dimension is vast and warfare is becoming increasingly data-driven. Modern warfare, therefore, demands that any assessment of the information dimension start with a data-driven approach to understanding it, supported with theory and applications from cyber, systems engineering, social sciences, marketing, psychology, and other disciplines. In recent years, social media and other data platforms have become the terrain through which nation-states and their proxies compete to control narratives. Whether an actor is launching a narrative (employing information â??firesâ?•) or manipulating a network (conducting information â??maneuverâ?•), third-party information technologies and cyber platforms record every action in exacting detail.

Data, whether within a social media post (text, image, or video), or a traditional website or blog, is created and stored in information technology systems external to the Department of Defense Information Network and typically commercially owned and operated. Analysis and assessment of information campaigns requires access to commercial data, and this access can occur through any of the following methods:

1. Browser-based access, where the data remains on the originating social media company servers.
2. Third-party tool access, where data is acquired by a third-party company and retained on its servers.
3. Government-acquired access to data, i.e., when data retention and analysis are conducted on government servers.

To assess the full breadth and scope of the information dimension, the third option is the most attractive to government analysts. When analysis is conducted on government servers, analysts do not reveal their priority information requirements to third parties, and the government can merge open-source data with other data sources. Additionally, government-acquired access to data is the only option that will enable government data-science experts to process and assess the data at speed and scale. Importantly, DoD data-science talent, including graduates of the Armyâ??s Artificial Intelligence Scholars Program, are unable to leverage their skillsets for information advantage if they do not have access to relevant datasets residing on government systems. Data science, by its very nature, requires data.

Data Powers SOF-Cyber Partnership

As Russia invaded Ukraine, US Army cyber mission forces and special operations forces (SOF) pooled talent and resources to assess the information dimension surrounding the conflict. The SOF-cyber partnership proved valuable, because it brought together the two largest formations in the Army operating in the information dimension: the information operations (IO) force structure largely found

within Army Cyber Command; and the psychological operations (PSYOP) force structure belonging to 1st Special Forces Command. Additionally, the pairing combined the technical systems and talent of Army Cyber Command (ARCYBER) with the irregular warfare mission and mindset of 1st Special Forces Command. It was a match made in heaven. While the characteristics of the IO-PSYOP partnership have evolved over the course of the Ukraine conflict, the United States has continued to make progress toward better data integration and employment across formations and communities. Ultimately, the SOF-cyber partnership has demonstrated the value of the larger â??Space, Cyber, and SOF Triadâ?• that is now under development and designed to enhance integrated deterrence of adversary malign activity during competition.

Together we (the three authors) led the SOF-Cyber data efforts in February 2022 â?? including the data science, data engineering, and cloud infrastructure portions of the project â?? and the following analysis relays our experience as practitioners on the project. As part of the SOF-cyber data efforts, we identified similar but disconnected efforts to acquire the same sorts of commercial data from the information dimension. There were even instances where SOF and cyber formations had separate contracts with a single vendor for the same data. The partnership helped to build efficiencies across the organization and to synchronize data acquisition and analysis efforts. Importantly, synchronization efforts created a single repository for all government-acquired data, allowing the data to be acquired once and then made available to the various commands as needed. For example, in one instance, we facilitated merging the, data into ARCYBERâ??s Big Data Platform where access was then provided to other DoD elements with an information advantage mission and authorities.

However, access to data is only one step in a data science effort. As the data began to accumulate, another key task became determining if we had the right data. Acquiring a big data set is useless if it is the wrong data, and while both cyber and SOF elements were acquiring large amounts of data, neither had the resources (or, at times, the authorities) to acquire *all* the data related to the information dimension surrounding the Ukraine conflict. In essence, each command was sampling a comparatively small amount of data from a very large data fire hose. For any researcher in DoD, it is important to know if ongoing data acquisition efforts support current and projected operational requirements. To ensure that they do, data sampling methodology needs to be evaluated and reevaluated regularly. In the case of the SOF-cyber partnership, we determined that the sampling for both cyber and SOF data acquisition methods required adjustments to support contingency operations in Europe. Ultimately, a large portion of the SOF-cyber coordination efforts focused on synchronizing and, in some cases unifying, our approach to aggregating, storing, and analyzing data in the information dimension.

Creating an Agile Data-Science Environment

In late 2021, ARCYBER developed an agile cyber data-science environment within their Big Data Platform ecosystem. The new environment is a containerized JupyterHub environment that provides data scientists with a scalable computing platform, loaded with their favorite tools, and enables access to all the varied data stores supported in Big Data Platformâ??s environment. At first, the new cyber data-science environment was used to support defensive cyber operations (including support for the Army response to the Solar Winds and Log4J compromises). Still, the emerging crisis in Europe provided its first large-scale use in support of information advantage assessment in an operational environment.

Many of the questions that senior Army and DoD leaders were asking about the conflict in Ukraine were not entirely answerable in the Big Data Platformâ??s preexisting third-party tools and dashboards. The new cyber data-science environment provided analysts with an agile platform that could pivot quickly to the right data and answer senior leader questions. Incredibly, our team assessed that 90% of the analysis and analytic products produced by the SOF-cyber information advantage team came from using the new cyber data-science environment. Thus, in a short period, the new environment proved critical to providing decision-makers with the information needed to better understand the Ukraine conflict, which could only be had from a data-science environment and not just a data dashboard.

Data Science Talent

The cyber data-science environment (and the agile DevSecOps process behind it) allowed the authors to fully leverage the data-science talent resident in the cyber and SOF formations. The environment was used by a wide variety of personnel, including those with civil affairs, military intelligence, psychological operations, special forces, and cyber backgrounds. But, behind the scenes, two career fields proved critical. The data science was largely conducted by Army Cyber Capabilities Development officers, or junior officers from the 17D career field, with some oversight and input from Operations Research and System Analysis officers, or field-grade officers from the functional area 49 career field. While most of the 17Ds support other cyber workflows, a few have distinguished themselves in cloud infrastructure and data science and have been consolidated at the enterprise level. These officers primarily acquired their machine learning and artificial intelligence expertise during their undergraduate education, while others have graduate school experience (both masters and PhD experiences are supported through scholarship and Army Civil Schooling programs). Access to relevant data and availability of the data-science environment allowed US forces to rapidly iterate on senior leader questions and provide relevant information at speed and scale.

Why Data Matters

Within three weeks of Russiaâ??s invasion of Ukraine, the combined SOF-cyber team developed seven new analytic approaches (including two new deep learning models and three new network science models) to support the unique requirements of the information dimension in Eastern Europe. In addition to the seven models, the team also relied upon existing models that the team had already developed and deployed, including machine learning, network science, natural language processing, and image-analysis models and visualizations. These models were deployed into production in custom machine-learning pipelines and atop the unified SOF-cyber data to feed a daily product automatically produced and distributed to relevant Army, Joint, and SOF commands in Europe.

Having the right data in the right environment enabled the daily delivery of relevant information to senior leaders and helped generate a big win for the Big Data Platform concept. Because the information dimension continues to be a critical aspect of Russiaâ??s aggressive actions in Ukraine, DoD needs to continue to build out its data-science capabilities to ensure we are ready to fight effectively and efficiently in future conflicts. Understanding the information dimension of conflict is more critical than ever and developing senior leadersâ?? understanding should begin with a data-centric approach. Meaningful operational questions can be answered, and operational insights gleaned when we have the right data, in the right environment, and in the hands of the right talent. If any of these ingredients are missing, the value proposition tends to fall apart. The SOF-cyber data partnership that began with Ukraine should continue and evolve to encompass new areas of research and data acquisition to develop new operational efficiencies and expand upon current successes.

Lt. Col. David Beskow, Ph.D. is an academy professor in the Department of Systems Engineering at West Point. He served as the Chief Data Scientist for Army Cyber Commandâ??s Technical Warfare Center from 2020-2022.

Maj. Daniel Hawthorne, Ph.D. is the lead for the Army Cyber Command agile data science environment and has led the ARCYBER Technical Warfare Center infrastructure team since February 2021.

Capt. Tommy Daniel is the data team lead for Task Force 40-25, 1st Special Forces Command (Airborne). His most recent assignments were with the 1st Special Forces Commandâ??s data office and as a cross-functional team leader and Special Operations Liaison Element to Moldova.

Main Image: Staff Sgt. Gregory Fretz, a cyber operations specialist with the 178th Cyber Protection Team, Mississippi Army National Guard, monitors cyber activity during Exercise Southern Strike at Camp Shelby, Miss., April 21, 2023. Southern Strike 2023 was a large-scale, joint multinational combat

exercise hosted by the Mississippi National Guard that provided tactical level training for the full spectrum of conflict. (Photo by Staff Sgt. Renee Seruntine)

Date Created
2023/07/27