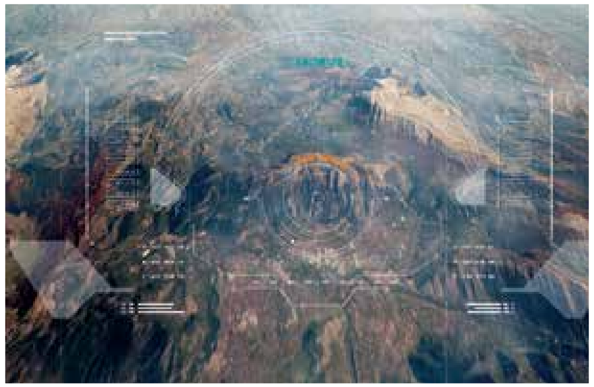

Artist rendering of a futuristic heads-up display during an aerial surveillance flyover. Photo by Pakpoom Makpan/Shutterstock

The artificial intelligence (AI) arms race is well under way with great powers, secondary powers, and even non-state actors actively pursuing the weaponization of this technology in a variety of ways. The purpose of this edited volume is to demystify the capabilities and limitations of AI-based military solutions. With a conversational tone and progressive learning trajectory across the chapters, *Big Data for Generals … and Everyone Else over 40* provides an accessible but comprehensive overview of the concepts and considerations for making emerging technology a true force multiplier for the Special Operations Forces enterprise.

United States Special Operations Command
ATTN: JSOU Press
7701 Tampa Point Boulevard
MacDill AFB, FL 33621-5323

https://jsou.libguides.com/jsoupublications

Comments?

JOINT SPECIAL OPERATIONS
JSOU
UNIVERSITY
PRESS

ISBN 978-1-941715-60-4

JSOU Report 21-9

Big Data for Generals … and Everyone Else over 40

Ellis/Grzegorzewski eds.

JSOU

JOINT SPECIAL OPERATIONS UNIVERSITY

"Let's shrink Big Data into Small Data …
and hope it magically becomes Great Data."

Big Data for Generals …
and Everyone Else over 40

Edited by Dr. David C. Ellis and Dr. Mark Grzegorzewski

JSOU Report 21-9

Joint Special Operations University

The Joint Special Operations University (JSOU) generates, incubates, and propagates (delivers and communicates) ideas, education, and training for expanding and advancing the body of knowledge on joint and combined special operations. JSOU is a 'hybrid organization' that performs a hybrid mission—we are a 'corporate university:' an academic institution serving a professional service enterprise, 'by, with, and through,' the United States Special Operations Command (USSOCOM). As such, we are both a direct reporting unit to the Commander, USSOCOM, on all Combined Joint Special Operations Forces education and leader development matters, as well as the educational and leader development component of the Command.

The JSOU Mission is to prepare Special Operations Forces (SOF) professionals to address strategic and operational challenges, arming them with the ability to think through problems with knowledge, insight, and foresight. **Our Vision** is to constantly strive to be(come) USSOCOM's "Think-Do Tank" Center of Special Operations Thinking. We pursue this mission and vision through our best-practice teaching & learning, research & analysis (R&A), and engagement & service-outreach operations, activities, and initiatives. We achieve these outcomes-based goals by providing specialized joint professional military education, developing SOF-specific and unique undergraduate, graduate, and post-graduate-level equivalent curriculum, and by fostering special operations-focused R&A and outreach, in support of USSOCOM objectives and United States national and global strategic goals.

JSOU's R&A efforts are guided and informed by the U.S. National Security, Defense, and Military Strategies, as well as the **USSOCOM Mission:** *USSOCOM develops and employs fully capable Special Operations Forces to conduct global special operations and activities as part of the Joint Force to support persistent, networked, and distributed global Combatant Commands operations and campaigns against state and non-state actors, to protect and advance U.S. policies and objectives.*

Joint Special Operations University

JSOU Press publications are available for download at
https://jsou.libguides.com/jsoupublications.

Print copies available upon request by writing
jsou_research@socom.mil.

JOINT SPECIAL OPERATIONS
JSOU
UNIVERSITY
PRESS

Big Data for Generals … and Everyone Else over 40

Edited by Dr. David C. Ellis
and Dr. Mark Grzegorzewski

JSOU Report 21-9
The JSOU Press
MacDill Air Force Base, Florida
2021

Recent Publications of the JSOU Press

Trained to Win? Evaluating Battlefield Effectiveness and Sociopolitical Factors among Partnered Forces, JSOU Report 21-8, Matthew Cancian

Barriers to Special Operations Forces-Led Counterterrorism Effectiveness, JSOU Report 21-7, Barnett Koven and Katy Lindquist

Muqtada al Sadr and Neo-Iraqi Nationalism: Implications and Opportunities, JSOU Report 21-6, Carole O'Leary and Nicholas Heras

On Competition: Adapting to the Contemporary Strategic Environment, JSOU Report 21-5, Edited by Aaron Bazin

Cross Domain IW Threats to SOF Maritime Missions: Implications for U.S. SOF, JSOU Report 21-4, Gary Kessler and Diane Zorri

Cyber Supply Chain Risk Management: Implications for the SOF Future Operating Environment, JSOU Report 21-3, J. Philip Craiger, Laurie Lindamood-Craiger, and Diane Zorri

On the cover. Business cartoon shows three businesspeople in a meeting. Businesswoman says, "Let's shrink Big Data into Small Data … and hope it magically becomes Great Data." Photo by iStock

Back cover. Artist rendering of a futuristic heads-up display during an aerial surveillance flyover. Photo by Pakpoom Makpan/Shutterstock

This work was cleared for public release; distribution is unlimited.

December 2021.

ISBN 978-1-941715-60-4

The views expressed in this publication are entirely those of the authors and do not necessarily reflect the views, policy, or position of the United States Government, Department of Defense, United States Special Operations Command, or the Joint Special Operations University.

****

JSOU is currently accepting written works relevant to special operations for potential publication. For more information, please email jsou_research@socom.mil. Thank you for your interest in the JSOU Press.

****

Contents

Foreword

The artificial intelligence (AI) arms race is well under way with great powers, secondary powers, and even non-state actors actively pursuing the weaponization of this technology in a variety of ways. The rate of scientific advancement in the various forms of military-oriented AI has increased markedly in recent years, and it appears now that many military professionals presume that AI applications constitute a necessary precondition for military success in future high-end conflict. Special Operations Forces (SOF) have similarly attempted to harness the power of AI, machine learning, natural language processing, and deep learning for their unique mission sets. While sound in principle, employing AI-based solutions efficiently and effectively first requires clear knowledge of (a) how they work, (b) the conditions for which are they are and are not appropriate, (c) the challenges of employing them in the field, and (d) how to employ them within ethical boundaries.

The purpose of this edited volume is to demystify the capabilities and limitations of AI-based military solutions. The chapters are written with the assumption that readers have a limited background with the underlying scientific, modeling, and data science principles that make AI-based solutions viable. The contributors to the volume are scholars and expert practitioners who work closely with the Joint Special Operations University to write for the specific needs of the SOF community. Nevertheless, this volume has applicability across the U.S. Government since the SOF community operates under nearly the same conditions as the rest of the government sector. With a conversational tone and progressive learning trajectory across the chapters, *Big Data for Generals … and Everyone Else over 40* provides an accessible but comprehensive overview of the concepts and considerations for making emerging technology a true force multiplier for the SOF enterprise.

David C. Ellis, PhD
Research Professor
Center for Adaptive and Innovative Statecraft
Joint Special Operations University

About the Authors

Dr. Karl Aspelund is the department chair and an associate professor in the Department of Textiles, Fashion Merchandising, and Design at the University of Rhode Island. He is also a visiting assistant professor at the Department of Folkloristics/Ethnology and Museum Studies at the University of Iceland. He completed a PhD in 2011 in anthropology and material culture at Boston University, where his dissertation was awarded the university professor Edmonds Prize as the best dissertation of that academic year. Dr. Aspelund is the author of three textbooks: *The Design Process, Fashioning Society,* and *Designing: An Introduction*. His most recent book, co-edited with Dr. Terry Gunnell of the University of Iceland in late 2017, was nominated for the Icelandic Publisher Association's Literary Prize in the field of Academic Works. Dr. Aspelund is currently working on his sixth volume. Its working title is *What is Good Design? Seven Meditations from a House on Fire.*

Dr. Justin Brunelle is a principal researcher at the MITRE Corporation. Dr. Brunelle holds a PhD in computer science from Old Dominion University where his research focused on web science and information retrieval. At MITRE, he conducts research to help government sponsors more effectively adopt emerging technologies, including topics in web science, cloud computing, and big data. Dr. Brunelle also helps guide the internal research and development for MITRE's Software Engineering Innovation Center.

Mr. Dave Bryson is a principal engineer at the MITRE Corporation. His current focus is on decentralized technology such as blockchain. He's also an active contributor to several open-source projects.

Dr. David C. Ellis is a professor at Joint Special Operations University (JSOU). He holds a doctorate in international relations and comparative politics from the University of Florida. Dr. Ellis's research on democratization and development in identity conflict spans over two decades. His interests in peacekeeping, conflict resolution, development, and atrocity in ethnic conflict focused his doctoral research on identity, social movements, organization and social learning theory, and economic growth theory. Dr.

Ellis served as an intelligence analyst in the U.S. Special Operations Command J2, deployed to Afghanistan in support of Special Operations Forces from 2010–2011, and joined JSOU in 2016. His current research focuses on the intersection of complexity, organizational learning within the special operations community, and integrated campaigning.

Lieutenant Colonel Andrew J. Geyer, U.S. Air Force, PhD, is an associate professor of statistics and deputy head of the Department of Mathematics and Statistics at the Graduate School of Engineering and Management, Air Force Institute of Technology, Wright-Patterson Air Force Base, Ohio. His research areas include the military applications of data science, artificial intelligence, optimization, and design of experiments. Prior to academia, Lt. Col. Geyer served as a staff weather officer in the 82nd Airborne Division, 75th Ranger Regiment and Combined Joint Special Operations Task Force–Arabian Peninsula.

Dr. Mark Grzegorzewski is a professor in the Department of Strategic Studies at Joint Special Operations University (JSOU) where he is currently focused on researching cyberspace operations and artificial intelligence. He has recently published in the *Special Operations Journal* on "Demystifying Artificial Intelligence through DoD Education" and contributed a chapter in an edited volume titled *Russian Cyber Operations: The Relationship between the State and Cyber Criminals*. Moreover, he created JSOU's Quick Look series with publications titled *Artificial Intelligence* and *Cryptocurrency.* Dr. Grzegorzewski holds a PhD, MA, and BA in political science from the University of South Florida along with a graduate certificate in globalization studies.

Dr. Bohyun Kim is the chief technology officer and an associate professor at the University of Rhode Island libraries. She is the author of three books: *Moving Forward with Digital Disruption, Understanding Gamification,* and *Library Mobile Experience: Practices and User Expectations*. She published many articles and gave numerous peer-reviewed presentations on topics related to emerging technologies and their impact on libraries at international and national conferences. She was the president of the Library and Information Technology Association from 2018-2019, which became a part of Core in Fall 2020, and served on many advisory boards and committees including those of the American Library Association, San Jose State

University's School of Information, and Rhode Island Library Association. She holds a MA in philosophy from Harvard University and a MS of library science from Simmons College.

Dr. Paul Lieber is COLSA Corporation's chief data and social scientist, as well as science and technology advisor to the Information Professionals Association. He previously served as the command writer for two U.S. Special Operations Command commanders and strategic communication advisor to Special Operations Command-Australia. Within academic environs, Dr. Lieber was full-time graduate faculty at Joint Special Operations University, Emerson College, University of South Carolina, and the University of Canberra, respectively. He holds a PhD in mass communication and public affairs, a master's in mass communication from Louisiana State University, and a BS in broadcast journalism from Syracuse University.

Mr. Pedro Cesar Lopes Gerum is an assistant professor in the Department of Operation and Supply Chain at Cleveland State University. His research focuses on transportation optimization, stochastic modeling, data science, and machine learning. Originally from Brazil, he obtained his PhD in industrial and systems engineering from Rutgers University. He is the author of multiple publications in transportation reliability and has experience developing machine-learning applications for finance problems at American Express. He has also helped create deep-learning models for the classification of exoplanets at NASA's Ames Research Center in California.

Dr. Joan Peckham is professor emerita of computer science at the University of Rhode Island (URI). She served as a program director at the National Science Foundation in 2008-2011. At URI, she served as chair of Computer Science and Statistics and led several initiatives including campus-wide coordinator of big data and data science and acting director of Collaborative for Explorations in Mathematics and Science. She was co-principal investigator (PI) of the URI ADVANCE Institutional Transformation award in 2004-2008 to promote the careers of tenure-track women in science, technology, engineering, and math disciplines and PI of an interdisciplinary National Science Foundation (NSF) research experience for undergraduates in art and computing. Her research is focused on data modeling in several domains including transportation and bioinformatics and computing education. Currently, she is working with colleagues from other NSF established programs to stimulate competitive research institutions to formulate and articulate the concept of no-boundary (beyond interdisciplinary) research and education and its importance to data and artificial intelligence ethics.

Dr. D.J. Shyy is the principal communications engineer at the MITRE Corporation. Dr. Shyy has 29 years of industry experience in wireless communications including 19 years at MITRE. He has significant experiences in 3G (code division multiple access), 4G (LTE), 5G, 6G including system radio frequency design and optimization, protocols security, architect 5G labs and security testbed, lab/field performance testing, and spectrum sharing/ interference mitigation. He is the chair of a 5G secure profile working group and a 5G technical advisor for government sponsors. Dr. Shyy received his MS and PhD degrees in electrical engineering from Georgia Institute of Technology.

Mr. Guarav Tanwar attended Rutgers University where he earned his BA in international political economy—the only person to do so that year and, to his understanding, ever. After graduating in 2010, he enrolled at Georgetown for his master's in public policy, graduating in 2012. After several assignments throughout the defense contracting world, his last assignment was the Combatting Terrorism Technical Support Office–Irregular Warfare Team, where he was able to carve out a portfolio focused on the fourth industrial revolution's impact on national security and international stability. Though he specialized in deepfakes, he also led projects on artificial intelligence (AI), the encrypted digital ledger, social media exploitation, social network analysis, countering disinformation, and several other technical projects. He is currently a defense contractor at the Joint AI Center serving as the Joint Information Warfare Mission Initiative's senior analyst on information operations.

Ms. Nishka Uberoi is a technical solution specialist and open labs manager in the Internet of Things (IoT) business at Vodafone Group. She regularly works with low power, wide area network technologies to power cellular IoT devices. Originally from India, Ms. Uberoi obtained her MS in electrical and computer engineering from Rutgers University. She is interested in new IoT, artificial intelligence, and blockchain technologies and their possibilities.

Dr. Yaakov S. Weinstein is the leader of the MITRE Quantum Technologies Group and the editor-in-chief of the Nature-Springer journal *Quantum Information Processing*, which has undergone significant growth in number of submissions and number of published papers since his appointment in 2014. Dr. Weinstein received his PhD in nuclear sciences and engineering in 2003 from the Massachusetts Institute of Technology. After completing his PhD, Dr. Weinstein was awarded a National Research Council research associateship at the Naval Research Laboratory. Dr. Weinstein joined MITRE in May of 2005 and has become a vital resource in quantum for MITRE and its sponsors. Dr. Weinstein continues to contribute scientifically to all areas of quantum and is consistently expanding MITRE's scope and areas of expertise to a variety of physical science areas of importance to MITRE sponsors.

Colonel Mark Zais, U.S. Army, PhD is the chief data scientist at U.S. Special Operations Command and has over 24 years of military service as an Army aviator and operations research systems analyst (ORSA). He spent the first part of his career in multiple assignments as a Longbow Apache (AH-64D) attack helicopter pilot before transitioning to the operations research military specialty. Col. Zais's previous Army ORSA assignments include Office of the Secretary of Defense (cost assessment and program evaluation), Department of the Army Staff, U.S. Special Operations Command, and Carnegie Mellon University. His academic and professional research focuses on data science, machine learning, metaheuristics, and simulation optimization. Col. Zais has a degree in operations research from the United States Military Academy (West Point), master's degrees in operations research and industrial engineering from the Georgia Institute of Technology, and a PhD in operations research from the University of Colorado at Boulder.

Executive Summary

The purpose of this monograph is to help leaders and managers in the U.S. Special Operations Forces (SOF) enterprise become comfortable and conversant with the vocabulary and concepts associated with Big Data. It is not designed to make the reader a data scientist. Rather, it enables the reader to make better use of, provide the appropriate support and environments for, and more richly receive the advice of personnel who are trained in data science. A main finding of the research is that there is a substantial disconnect between the popular imagination of predictive analytics and what cutting edge science and technology can actually deliver. The magic of find, fix, finish, exploit, analyze and disseminate; social media trend analysis; the potential power of metadata analysis; and other powerful computer assisted analytic tools, such as Project Maven, seem to suggest that the military is on the cusp of an extraordinary era where an enemy's behavior can be predicted with a high degree of probability. Perhaps this is true, but it is more likely not. Certainly, insight can be gleaned from trend analysis and correlations, but it is essential to remember that the human behaviors underlying the predictive analytics most everyone experiences—through Google, Amazon, and other providers—are very different than the human behaviors with which SOF contend.

The chapters reveal important insights for military and civilian leaders across the SOF enterprise. These include:

- Successful incorporation of big data will require the deliberate creation of a data management culture instead of merely adapting current practices.
- Big data is not necessarily a labor-saving activity because proper data science requires a teaming approach among content experts, front line users, and data scientists; novel challenges require specialized teams.
- Dashboards and preset interfaces rely on assumptions and algorithms that often fail to match the social dynamics underlying novel challenges such as conflict environments.
- Artificial intelligence (AI), machine learning (ML), and data science more generally all rely on algorithms, which in turn rely on conscious and unconscious assumptions, biases, and frames, and there

are inevitable limitations on data available to train algorithms and to process for predictive analysis.

- Algorithms are statistical models that, by nature, can only approximate a slice of social reality, and it is consequently problematic to rely upon them for decision-making tools independent of human reasoning.
- Statistical models reflect the past with the presumption that the patterns detected will continue into the future, which is scientifically valid for closed laboratory systems, but open social systems are more prone to novel, often unpredictable pattern changes.
- Silicon Valley is a poor model for SOF as its AI-based systems rely on high degrees of self-reported behavior and interests that cannot often be replicated in SOF operating environments or processed by existing knowledge management practices and information technology platforms.
- Accruing personnel with data science skills requires a mixture of incentives and non-traditional military opportunities across military, civilian, and contractor options, but, more importantly, leaders should creatively engage the larger social system of data science capabilities outside the military instead of owning them outright.
- While AI and ML seem essential to warfare, the ethical considerations of employing such capabilities require deep review and leader education before employing them as kinetic options or as decision-making tools.
- Disruptive technology (e.g., quantum computing), communications logistics (e.g., internet access in austere environments), and electronic signatures all present significant future challenges to forward-deployed SOF and information security, which reinforces the first SOF Truth that people are ultimately more important than hardware and that technological capability should be critically evaluated instead of presumed to be an essential element of deployed forces.

Introduction

Dr. David C. Ellis

Thinking About Big Data for the Special Operations Forces Enterprise

We are in the midst of one of the most consequential technological and geopolitical shifts in recent history. Just like the U.S. leveraged the industrial revolution in its rise to Great Power status in the nineteenth century and to preeminence in the twentieth, the ongoing revolution in technology and data science could represent the opportunity for a competitor to achieve a similar ascent in the twenty-first. But, if harnessed correctly, this technological revolution could also prolong the influence, values, and order the U.S. has shepherded since World War II, irrespective of the growth and advancement of others.

Unfortunately, the language and management concepts associated with this new technological revolution diverge in significant ways from those of the industrial revolution and the way computers augmented it through the first decade of the twenty-first century. Though called generically and approachably *big data*, as though it is just a scaled-up version of what has come before, this new era actually requires leaders and managers to think in fundamentally new ways. The good news is that the new ways of thinking are not really that hard to grasp. The bad news is that those ways of thinking have not often been taught to most of the people now assuming leadership and managerial roles in the implementation of big data solutions.

To complicate matters, those responsible for implementing big data solutions in the U.S. military face two additional challenges. First, though big data requires new ways of thinking, legacy architectures, and data systems, mental models pervade the U.S. Government's information technology (IT) infrastructure. In other words, it might be far easier and less costly to take full advantage of big data if a system can be built entirely from scratch. The military's existing systems and processes stand to be among the most difficult obstacles in the efficient implementation of big data solutions, so it falls to leaders and managers to navigate—or choose to radically disrupt—the status quo.

Second, though existing computing power is resulting in previously unthinkable commercial and military capability, the coming decade could very realistically see the development of quantum computing, a disruptive technology so consequential that it is impossible to forecast its true impact on future sociopolitical and military organization. The leader in the development of quantum computing will accrue extraordinary advantages in the twenty-first century. The ability to leverage future quantum computing power depends critically on how leaders and managers conceive of and implement big data capabilities today.

The purpose of this text is to help leaders and managers in the U.S. Special Operations Forces (SOF) enterprise become comfortable and conversant with the vocabulary and concepts associated with big data. It is not designed to make the reader a data scientist. Rather, it will enable the reader to make better use of, provide the appropriate support and environments for, and more richly receive the advice of personnel who are trained in data science. Think of this text as useful for digital immigrants—people born well before the year 2000 who are proficient with a range of software applications and technologies though not programmers—and even digital tourists—people who can use computers, software, and email yet struggle to master the word processing, spreadsheet, and presentation tools available to them. To frame the conversation in the pages that follow, a few important points should be emphasized from the start.

Distinguishing Between the Enterprise and the Mission

The first major point is that there seems to be some confusion about how big data capabilities fit with the SOF enterprise. Some confuse supporting the mission through big data analytics with serving the enterprise's data and information needs. Oftentimes this confusion is the result of applying current data storage and networking concepts to a scaled, operationally integrated data flow consistent with modern concepts of the find, fix, finish, exploit, analyze, disseminate (F3EAD) process. Certainly there are a growing number of terabyte and petabyte data feeds along with analytical requirements flowing from them, and while they can be decisive from an operational and largely tactical counterterrorism and counter threat network perspective, they represent only a subset of the concerns with which SOF enterprise leaders and managers must contend. Figure 1 illustrates how the current mission-oriented aspects of big data relate to the other perhaps more mundane but more critical elements.

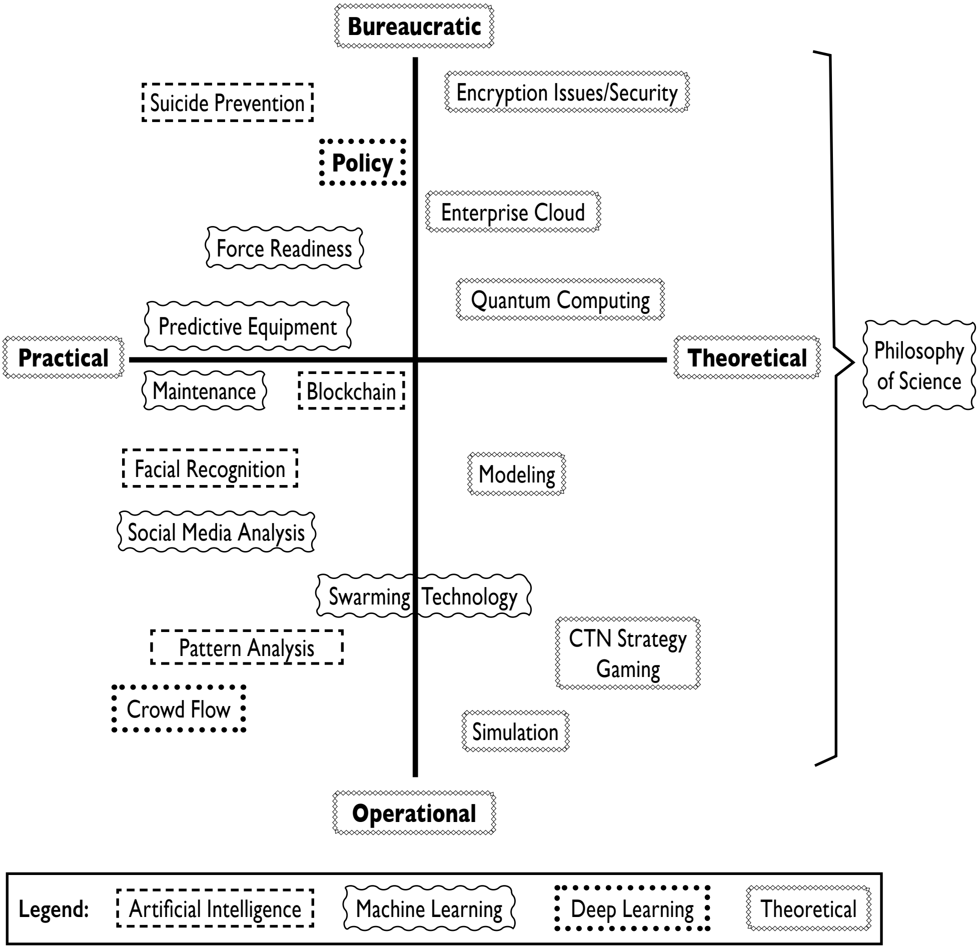

Figure 1. Dimensions of big data applications across the SOF enterprise.
Source: Dr. David C. Ellis

While the preponderance of the Force's experience will likely be in the lower left quadrant with a mixture of practical and operational applications, the coming technological revolution will require the enterprise's leaders and managers to focus their attention on the bureaucratic half of figure 1. The operational half of figure 1 first requires the right systems architecture, data capture process, and personnel talent management and discovery structures to make big data solutions viable on the operational side. Clearly delineating between the enterprise's big data requirements and the mission's analytical solutions is the first point the reader must recognize.

The Boundaries of Predictive Analytics

The second main point to emphasize is that there is a substantial disconnect between the popular imagination of predictive analytics and what cutting-edge science and technology can actually deliver. The magic of F3EAD, social media trend analysis, the potential power of metadata analysis, and other powerful computer assisted analytic tools, such as much discussed Project Maven of the Department of Defense (DOD), seem to suggest that the military is on the cusp of an extraordinary era where an enemy's behavior can be predicted with a high degree of probability. Perhaps this is true, but it is more likely not. For a host of reasons that will be discussed throughout this text, leaders should warily engage technology industry executives and vendors who promise to transform data into prediction. This is not to say that insight cannot be gleaned from trend analysis; rather, it is simply to warn that the human behaviors underlying the predictive analytics most everyone experiences—through Google, Amazon, and other providers—are very different than the human behaviors with which SOF contend.

Asking the Right Questions

Big data solutions for certain SOF challenges are feasible in the near term if leaders and managers ask the right questions and focus on reforming the enterprise's data capture processes. Most of the immediately actionable big data opportunities currently lie on the bureaucratic-practical quadrant of figure 1 due to reasonably effective data reporting structures already in place. Tweaks to the system need to be made to optimize the opportunities, but much can be learned about what it takes for the enterprise to generate big data policy and processes while still gaining much needed efficiency through high-powered analytics.

Though most of the emphasis to date has been on the operational-practical quadrant, asking the right questions on the bureaucratic side will actually set the conditions for holistic enterprise solutions. Otherwise, history illustrates that leaders and managers are prone to apply "Band-Aid" solutions to seemingly urgent—but temporary—challenges resulting in costly, poorly used interfaces. Competing for advantage in the era of strategic competition demands a comprehensive, enterprise solution to data management, analysis, and exploitation. More explicitly, the questions that the SOF enterprise's leaders and managers need to ask are not about tools and dashboards for tactical and operational missions but about how to cultivate a data culture orientation across diverse units and with unique reporting cultures and needs. This text provides the vocabulary and conceptual foundation for the current and future leaders of SOF to ask the right questions.

Key Insights

A few key takeaways course throughout the text and deserve highlighting. While the chapters are written by a diverse group of military and civilian academics, the following themes consistently appear.

It's the Question, Not the Interface

Data science and making the most of big data are first and foremost about assembling diverse teams whose capabilities and skills combine to address a specific question. No matter how elegant the technology or powerful the supercomputer, poorly constructed questions and teams will render them inert. Many in the SOF enterprise expect that artificial intelligence (AI)/machine learning (ML) can provide the easy button solution for analytics and decision-making or can cull the Internet to tell them what they need to know. This is exactly backward. There is no dashboard or data feed that can remain useful for long without teams of personnel capable of adapting big data analytics for ever-changing and emergent needs. Contrary to the belief that AI/ML reduces the burden on limited personnel, data science is often labor intensive because questions change but interfaces cannot.

The Data Management Culture

As a result, big data is really a covering term for a data management culture. Companies like Google, Amazon, and Apple can exploit their data in seemingly effortless, predictive ways precisely because they place data at the center of their operations and build their business processes around self-profiling human behaviors. Their organizational cultures consequently allow data to flow with evolving customer preferences and patterns and attract engineers and designers with a talent for exploiting them. In contrast, the SOF enterprise sits atop a legacy IT infrastructure that is spread across multiple components and further subdivided by directorates headed by executives of roughly equal rank. Data, information, and IT are packaged as a coequal directorate function, not at the center of operations. Nor are the personnel broadly trained or educated in data management or IT. To make the most of the petabytes of data the SOF enterprise will accrue over the coming years, it must transform to adopt a data management culture. If it is successful in developing the culture, then the big data infrastructure that the enterprise needs has a better chance of emerging.

Creativity in Personnel and Talent Management

Of course, Silicon Valley can attract the talented engineers and designers because the tech industry can pay the limited supply of data scientists the salaries they command given the demand for their specialty. The SOF enterprise is severely constrained in this respect and the working conditions tend to discourage data scientists from seeking employment in the government. It will take a mix of active duty, contractor, and government civilian personnel to fill the billets, but there will need to be flexibility in how they flow in and out of government service given the stereotypical data scientist personality. Fortunately, there are non-financial ways to attract and motivate data scientists to the SOF enterprise, but they have to be at the center of the system, not *ad hoc* solutions.

Developing a data management culture therefore requires a multifaceted and creative approach to talent management and cultivation. Moreover, it will require leaders and managers to move beyond the hierarchical and directorate-based units currently in place. A question-based, data management culture compels the enterprise to move toward multi-directorate, cross-functional teams, divorcing rank from perceived knowledge and wisdom and, on occasion, providing advice directly to leaders outside the normal chain of command. Nothing about this will be comfortable, but incorporating big data into the SOF enterprise will likely proceed slowly otherwise.

Resilience for Disruptive Technology

As amazing as technology now appears, the SOF enterprise could in a decade be faced with truly disruptive advances in computing technology. Irrespective of what leaders and managers put in place over the next five years, they need to be sensitive and open to radical transformation shortly thereafter.

Balancing the needs for near- to medium-term improvement while keeping an eye to long-term developments will require constantly updated appreciations of the technology context. However, with a data management culture in place, there is a greater chance that the SOF enterprise will manage the risk well and avoid becoming overleveraged in legacy technology or boutique solutions.

Organization of the Volume

The chapters that follow are designed to build upon one another. They represent the key insights that emerged from a Joint Special Operations University symposium entitled "Thinking about Big Data for the SOF Enterprise," held in February 2018 and validated through interactions with related military, academic, and industry partners. The text proceeds in three parts to give the reader logical stopping points depending on interest. Part I provides an overview of vocabulary, philosophy, and scientific principles at the heart of big data. The reader will conclude this section with the basics necessary to interpret a conversation about big data and AI/ML options. Part II dives directly into the leadership and management concerns associated with developing a data management culture. This section is essential reading for digital immigrants and digital tourists put in a leadership role with data management or analytics as part of the portfolio. Part III explores the advanced concepts associated with big data and is designed to help leaders think about future ethical concerns with AI/ML and interpret the technology and cost implications associated with a big data-centric force structure.

Part I: The Foundations of Big Data Analysis

Chapter 1 provides the reader with the basic vocabulary associated with big data. It covers the differences between automation, AI, ML, natural language processing, and deep learning. It also discusses the basic principles of statistics upon which big data relies. In particular, the chapter distinguishes between closed and open systems to explain why big data is effective in the former but less so in the latter. Chapter 1 concludes by providing the reader with a solid foundation on the limits of predictive analytics, a review of the standard tools SOF are likely to employ, and therefore, more reasonable expectations about big data solutions.

Chapter 2 takes the next step by explaining the basics of statistical modeling. Big data is driven by algorithms, and there are inherent limitations with algorithmic analysis, especially in open systems. The chapter begins with a description of how models are derived and how AI/ML adapts models as new data is obtained. It then lists a number of ways bias can impact computer models so that leaders and managers can critically evaluate big data solutions and analyses. Chapter 2 ends with a conversation on why the emphasis should be on the question instead of the dashboard (given what was revealed by the discussions on modeling and bias).

Chapter 3 makes the point that big data is not so much about the computer or the technology but how teams of personnel make use of them. Despite the general sense that big data solutions can replace human labor, there is a labor-intensive component to big data analytics. Chapter 3 expresses the reasons why leaders and managers need to focus on personnel and a data management culture instead of technology solutions. It covers how to think about team composition and offers ideas about how to accrue necessary talent and systems over time. Getting the personnel aspect of big data right is essential because having correct vocabulary and knowledge of modeling cannot compensate for poorly designed questions and teams.

Part II: Management Issues with a Big Data Capability

Chapter 4 transitions the text to content directed at the SOF enterprise's leaders and managers responsible for either implementing big data solutions or overseeing analysts who utilize them. It begins with three values and four attitudes that should guide the development of a SOF enterprise big data capability. It then critically discusses the main deficiencies and cultural disconnects between what a big data capability requires and how the military behaves. Chapter 4 concludes with different models that different elements of the DOD have adopted in their attempt to develop big data capabilities. This provides leaders with options for cultivating a data management culture across the SOF enterprise.

Chapter 5 next surveys ideas for cultivating and attracting the talent necessary for the SOF enterprise's big data needs. It covers the options and challenges for generating data scientists through the military, through recruiting civilian talent from the private sector, and through the contracting industry. There are benefits and risks to each approach, but the chapter identifies some creative options that have been tried in the U.S. Government to attract and retain much needed talent.

While hinted at in chapter 1, chapter 6 directly explores the differences between Silicon Valley organizational culture and the military's. Silicon Valley is often associated as a model for SOF, but in reality, they work in truly divergent social realities. Chapters 1 through 5 provide the launch point for a deeper discussion about the mismatch. In particular, chapter 6 explores why modeling and predictive analytics seem to work so well for Silicon Valley but then illustrates why the operating environment of SOF violates the assumptions that make Silicon Valley's algorithms so effective. In the end, Silicon Valley's challenges relate more readily to patterned, replicated behavior, whereas the challenges of SOF are rooted in open systems where new, adaptive behavior is the norm thereby undermining the effectiveness of algorithmic modeling.

Together, parts I and II provide the foundational knowledge that leaders and managers should know as they engage in conversations about big data. However, readers have the option of moving to part III, which contains important conceptual content for evaluating big data solutions and challenging tech executives and vendors in their presentation of AI/ML solutions.

Part III: Advanced Concepts with Big Data in the Social Sciences

Chapter 7 turns to ethical concerns that have arisen in the era of big data. Human beings can accomplish unprecedented feats thanks to big data, but there are growing ethical concerns—both domestically and internationally—about which leaders and managers in the SOF enterprise should be aware. This chapter provides an overview of those concerns.

Chapter 8, the final chapter in the volume, takes a future-oriented look at disruptive technologies that could impact the SOF enterprise in a relatively short period of time. While many in the enterprise envision a hyper-enabled operator powered by AI and other emergent technologies, there are significant challenges to making the concept practical. The chapter first discusses the basics of current technologies, including cloud computing, bandwidth issues both within the U.S. and abroad, and blockchain encryption. Chapter 8 then explains the foundations of quantum computing and what it could mean for encryption and bandwidth issues down the line. It concludes with ways to think about building a big data capability for the SOF enterprise while avoiding investments in big data infrastructure that could become obsolete in the event that quantum computing becomes a viable technology.

Big data is a vast field with a wide range of issues to discuss. This volume certainly cannot touch upon all of them but endeavors to provide the reader with sufficient background to ask good questions when confronted with big data opportunities and challenges. If the design works, the reader will take a qualitative leap forward in adopting the concepts and traits that will make the SOF enterprise effective in the transition to AI/ML-based technology and techniques.

Part I: The Foundations of Big Data Analysis

Chapter 1. The Basics of Big Data Terminology

Dr. Karl Aspelund, Mr. Pedro Cesar Lopes Gerum, Ms. Nishka Uberoi, Dr. Paul Lieber

Big data is a vague umbrella term referring to the immense amounts of digital data that are being produced in a wide variety of fields in the present day. Data of this kind can, for example, be retrieved from a number of digital content social activities ranging from social media, such as instant messaging and posted images, to participation in genomics projects, to entries in search engines. Alternatively, one might find massive amounts of data in intelligence gathering activities, political demographic analysis, weather monitoring, or medical research. The term itself is now (in 2020) spoken of as the new normal, and the hype around it may have passed its peak.[1]

Whether in the private sector or military context, the common feature of big data is the emergence of large-scale, continuously expanding databases allowing computer-assisted analytics to generate insight into a variety of areas in the social and physical sciences that would otherwise be hidden by the sheer volume of data. Corporate big data practices are above all aimed at generating profit by selling or employing data analytics to provide customized user experiences. Military, security, and intelligence agencies, on the other hand, collect big data in various contexts for a variety of applications. For example, drones collect vast amounts of video for U.S. military and counterterrorism operations. Similarly, the public health and science communities gather vast amounts of data to study epidemics, climate change, and other natural phenomena.[2] Of course, for different reasons, both the public and private sectors have interests in big data's ability to capture Internet users' physical conditions, the frequency and qualities of their social contacts, their search preferences and patterns, and their geographic mobility.[3]

Big data differs from conventional, large-scale datasets by virtue of what is known as the Three V's: volume, velocity, and variety.[4] These three elements, described below, constitute the baseline characteristics of big data.

1. **Volume.** Big data means working with extremely large amounts (terabytes, petabytes) of unstructured, low-density information such as message feeds, click-data from webpages or apps, or livestreams from monitoring equipment. Very often, people misperceive large Microsoft Excel spreadsheets, PDFs, or presentation slide decks as constituting big data, but this is in error. In general, if a standard software program can effectively manage data, then one is not dealing with big data. Think of it this way: a common metric for big data is one petabyte, which is 100,000,000 times larger than the 10 megabyte file rejected by many email programs.
2. **Velocity.** The rate at which large amounts of data arrive and the time sensitivity with which data can or must be acted upon are also factors to consider with big data. How will the data be received, stored, and retrieved? How rapid must the response to incoming data be? Must user experience algorithms respond in the moment? While not all big data requirements need to be acted upon in real time, there are applications across the SOF enterprise that would certainly benefit from this capability as part of the enterprise infrastructure, especially as it confronts global threat networks and near peer competitors.
3. **Variety.** Data now arrives in a multitude of types and formats as compared to the number and text-based, structured, often statistical datasets typically encountered before the onset of big data. Nowadays, systems and analysts must contend with streams of social media data (audio, video, and written messages), and sensor-based data requires pretreating and analysis to be placed in context and to be positioned correctly into the larger analytical framework. For digital immigrants and tourists, it is often difficult to accurately conceptualize the challenges associated with such massive, unstructured data when their mental models are anchored to mostly structured, spreadsheet-oriented versions of data.

An additional two V's can be added to the list: value and veracity. Data has value, but it is often not obvious and must be discovered before the data is seen to be of any use. Much of the attractiveness of artificial intelligence (AI)/ machine learning (ML) is that computers can process massive volumes of data to discern correlations and patterns that would be nearly impossible to discover through human labor alone. Big data is popular as a concept precisely because it can find value in data that is obscured by the Three V's. Also, it is not always clear how truthful or reliable the data might be, so the veracity of data must also be assessed. Processes, therefore, need to be in place to evaluate the data in order to decide whether it is of value and to determine its veracity.[5]

Terms of Reference

Much of the confusion with big data stems from the way the terms associated with it tend to be used almost interchangeably. Fortunately, there is a hierarchy to the terms that can distinguish one from another if memorized in order. They are, in order of increasing technical difficulty: automation, AI, ML, natural language processing (NLP), and deep learning (DL).

Automation

An important aspect of processing big data is the need to move away from using human labor to perform tasks dedicated to the routine capturing, categorizing, and storage of huge amounts of historical data (snapshots of an enterprise over time) into so-called warehouse databases. Automation occurs when computers accomplish these predictable, consistent, and highly repeatable tasks, thereby freeing up limited human resources to focus on more valuable tasks. The rapid processing of data via automation helps ensure that it does not languish unused but is rather delivered to the right place at the right time where it can be viewed, considered, and acted upon.

Big data automation should therefore touch on each of the five major steps in a data warehouse process:

1. Extracting data from applications into temporary data structures
2. Improving the quality of the data by making corrections, fixing errors, and removing irregularities (e.g., cleansing/scrubbing data) and transforming it into a required uniform/normalized format (e.g., all birth-gender information transformed into either M, F, etc.)
3. Loading the transformed data into the warehouse
4. Distributing the data into subsets (data-marts) devoted to the precise needs of specific teams

5. Translating the data into valuable information and distributing it to the right places at the right times[6]

Artificial Intelligence

Whereas automation is essentially about translation and sorting, AI is designed to approximate human decision-making. AI is a general term that implies the use of a computer to model and/or replicate intelligent behavior. The academic discipline was established in the 1950s, and the term *artificial intelligence* was first used in 1956 by Professor John McCarthy of the Massachusetts Institute of Technology for a conference defining the major goals of AI. The basic properties of AI are reasoning, learning, and problem solving.

In the present day, AI systems capable of human levels of decision-making (called expert systems) are found in a variety of sectors: finance, healthcare, heavy industries, aviation, communications, military, and more. Research has broadened into several sub-disciplines such as robotics, NLP, computer vision, computational biology, and e-commerce.[7] The research field of AI is generally and currently considered to be "a variety of research areas concerned with extending the ability of the computer to do tasks that resemble those performed by human beings."[8]

Machine Learning

ML takes an evolutionary step in computer-based decision-making. ML is the process of teaching a computer through repeated experience to analytically carry out a task rather than programming the computer to carry that task out step-by-step as with AI. ML utilizes mathematical and computational science approaches in the coding, which are typically split into (a) supervised learning, where the computer learns by being provided with labeled, training data which it eventually begins to recognize and (b) unsupervised learning, where the computer groups similar data and experiences in iterative experiments in order to pinpoint anomalies or discover patterns or correlations that result in failure. Much like humans, especially children, ML systems have the ability to learn what not to do through trial and error, which gives them a better chance of returning productive outputs once they recognize the decisions correlated with failed outputs.

Since 1983, ML has been applied to problems in agriculture, chemistry, computer programming, education, expert systems, game playing, image recognition, mathematics, medical diagnosis, music, NLP, physics, problem solving, robots, and speech recognition.[9] Since roughly 2000, there has been an increase in the number of successful applications, ranging from Internet searches to autonomous vehicles and from medical imaging and diagnosis to speech recognition. The growing range of ML applications has been driven by the increased availability of inexpensive computers, an increase in computing power, the development of improved ML algorithms, greater interest in the area from both the research community and the commercial sector, and most notably by the deluge of big data pouring in from an increasing number of sources.[10] ML systems usually work in the background of big data infrastructure and are used to analyze information to help understand patterns and plan responses and actions. For example, a program might determine whether a person in one photograph is the same as a person in another. Or it might, through NLP, identify a spam email by analyzing and categorizing the message's content.

ML systems usually work in the background of big data infrastructure and are used to analyze information to help understand patterns and plan responses and actions.

Natural Language Processing

NLP is the study of mathematical and computational models of the structure and function of language, its use, and its acquisition. It also deals with the design, development, and implementation of a wide range of systems such as speech recognition, language understanding, and language generation. On the theoretical side, the study involves mathematical and computational modeling of syntax, semantics, pragmatics, and discourse aspects of language. These may involve certain aspects of the relationship of the speaker and the hearer, or in the case of a NLP system, the user and the system. Investigations such as these are interdisciplinary and involve concepts in computer science including AI, linguistics, logic, and psychology.[11]

NLP is a subcategory of ML that centers on allowing computers to process languages at a human level of cognition—to develop an intuitive understanding of language. Today, NLP is used in text and social media analytics tools to analyze issues and opinions. A popular use case for NLP is analyzing posts or reviewing sites for feedback on products.

Although analyzing text for marketing is extremely important, another use of NLP is to enable systems to interface with humans by generating original conversational text or analyzing text written by people. This is found in interactive applications such as chatbots or other customer experience applications using sentiment analysis such as routing a customer to a certain agent based on status, what was said, and even how it was said (recognizing mood).

NLP is also necessary for search-driven analytics where users employ a natural language experience to search and analyze their data to find insights, such as search engine autofill functions. Some of these search engines learn about users from their analytics history and then provide them with search suggestions based on what might be most relevant to them. NLP, together with ML, is also used in other applications such as text summarizing and classification.[12] Additionally, NLP has its own subfield called natural language understanding (NLU). NLU goes beyond analyzing and replicating the structure of language to interpret intent and resolve context and word ambiguity so the intended meaning of spoken or written language may be understood with all the subtleties, context, and inferences that humans can instinctively grasp through years of habituation.

Deep Learning

All ML is focused on specific features of the data (e.g., source Internet protocol address or interarrival rate). Given enough prior data, a system will classify, predict, or cluster new observations, but ML can be divided into two schools of thought. One school does not attempt to model the physiology of the human brain or neural networks but rather focuses on mathematical algorithms. In the other, DL, systems are modeled on the physiology of the brain, specifically mimicking the roles of neurons and synapses.

The DL artificial neural networks, algorithms inspired by the connectivity of the human brain, learn from large amounts of data just as humans learn from experience, creating larger knowledge from smaller bits of information. A DL algorithm performs a task repeatedly, each time reflexively tweaking the formula a little to improve the outcome. Artificial neural networks thus break down complex problems into a multitude of tiny problems. Finding a face in a photograph is, for example, commonly broken down into deciding whether an eye, nose, or ear is present in the image and whether they are correctly located relative to each other. The connection between neurons is such that the output of the first neuron (e.g., there is an eye in the frame) is fed into the input of the next connected neuron by a multiplicative parameter that determines the weight of the connection. These parameters—or weights—are adjusted through algorithms to enable the system to learn to match the input pattern to the desired output classification or prediction.[13]

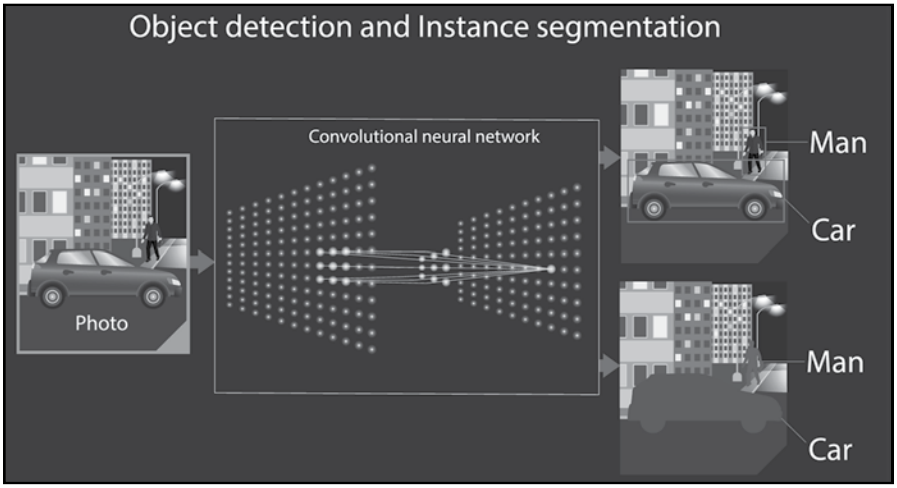
Object detection and Instance segmentation
Photo
Convolutional neural network
Man
Car
Man
Car

Figure 2. Convolutional neural network performing the task of object detection and instant segmentation. Source: Olga Salt/Shutterstock

This allows computers to solve a number of complex problems without human intervention. DL has enabled the writing of programs to allow text summarization, language translation, facial recognition, and vision for driverless vehicles and drones. In addition, virtual assistants like Alexa, Siri, and Cortana all use DL to understand speech and language in order to interact with humans. Recently, computers have mastered complex games like chess and Go through DL, outperforming the most talented humans.[14]

Correlation versus Causation and the Limits of Predictive Analytics

Advocates of big data reason that gathering, combining, and/or analyzing more information will lead to a superior operating picture and better decision-making. In this view, big data datasets can be crunched through ML/DL "gonculators" to produce results so informative that the end user can predict future events or phenomena. For SOF, the idea is that big data might yield increased strategic awareness, better long-term planning capability, and greater situational awareness when a long-term presence is required.

Assessing relationships among compiled data to predict future outcomes is referred to as predictive analytics. Predictive analytics operates under the assumption that if a dataset is truly indicative of a population and/or problem, identifying statistical relationships within the collected data can speak to what future events will hold. Being statistically bound, it can also—with a mathematically based level of confidence—place a numeric weight on the likelihood of these events occurring. Numerical weighting joined with a confidence level create the impression that the present is knowable and therefore sets in motion the expectation that the future can also be forecast. To the extent that sociocultural patterns repeat day after day, year after year, there is merit to this perspective. Unfortunately, the patterns of human interaction change—sometimes very slowly but other times very rapidly—especially in the age of the Internet, social media, and instant communication. While AI/ML/DL interfaces appear to be predictive to the average user such as the ubiquitous "You Might Also Like…" algorithms on shopping websites, two different dynamics interfere with this expectation for SOF. The first dynamic is between closed systems and open systems. The second dynamic is the difference between causation and correlation.

Closed versus Open Systems

ML algorithms can be implemented in either closed or open systems. Closed systems are systems that assume complete information and no gray areas. They assume that unavailable information is of no importance to the prediction. In other words, in closed systems, all the variables impacting an interaction can be known, controlled, and evaluated for effect, and all other external factors are prevented from impacting the interaction.[15] Common closed systems include cars and chemical production facilities—both have myriad pipes, tubes, temperature controls, sensors, and finely tuned instruments that have to work in perfect synchronization to produce the desire output. Closed systems are ideally suited to the scientific method—and therefore AI/ML/DL statistical modeling—because all the variables can be known and controlled.

Open systems, on the other hand, allow outside variables or additional information to change the direction of the current interaction. In fact, it is impossible to know all the potential interactions in the system precisely because the object of study cannot be isolated from the environment around it. As a result, there are some known variables impacting the outcome but uncountable other variables about which nothing is known—and consequently cannot even be factored into the algorithm.[16] Noteworthy open systems include financial markets, highway systems, and climate systems. Open systems are highly problematic for the scientific method (and AI/ML/DL statistical modeling by association) because environmental factors change the starting condition of each case, unknown variables impact the outcome, and known variables often have the capacity to change themselves by choice.[17] In short, the statistical error increases significantly in open systems because there are never two exactly alike experiments; all interactions are unique in some way. In each of the examples provided, AI/ML/DL capabilities can be applied, but they cannot predict outcomes. Instead, they can provide the probability that a sociocultural pattern will basically repeat given how the chosen variables interacted in the past, but this is different than prediction in closed systems. So what exactly is the difference between causation and correlation?

Causation

In closed systems, it is possible to talk about prediction because the variables can be controlled well enough to test the interactions hundreds or thousands of times to generate a statistically significant cause and effect relationship. Moreover, the tests can be reset anywhere on the globe and the same results will occur as long as the environmental conditions can be controlled. Predictive analysis in closed systems is possible, but even then, laboratory experiments must be undertaken to determine if the prediction actually holds true.

The allure of predictive analytics with big data occurs because there are closed system applications to big data. For instance, companies are using their data analytical insights to find correlations by focusing on stable patterns and filtering noise that falls outside of the model. Thus, the closed system reduces variability in the model, but in return, the model makes stronger, more consistent predictions about future performance.

Correlation

In open systems, however, it is not meaningful to talk about prediction because the introduction of unknown variables and an uncontrollable environment prevents cause and effect from ever truly being certain. Still, a subtle but very important distinction must be made when referencing statistical outcomes from predictive analytics. Specifically, predictive analytics can explore correlation between different items—meaning, how much a piece of data or a series of data is statistically related, or correlated, to another. Using statistical analysis software (IBM's SPSS and SAS are the two most common packages) and in a predictive analytics approach, a dataset will be examined for relationships across items.

How does the process work? A user loads a dataset into statistical analysis software within which an array of different statistical techniques is at his or her disposal. Common analysis techniques include determining correlations, regression to determine impact or weight of the correlations (how different items predict into a specific outcome), analysis of variance or ANOVA (how different items predict into specific group memberships), and structural equation modeling or SEM (how different items relate to each other in a statistical sequence, usually analyzed in a program called AMOS).

When a correlation analysis is run, a software package will objectively review a dataset for when score patterns on a particular item or items tend to correspond, or correlate, with score patterns on another (or others). As an example, say a statistical assessment of a particular province notes that increases in documented adversary numbers is correlated with greater allied intelligence, surveillance, and reconnaissance (ISR) presence but also lower skirmishes. Combined, the analyst might reason the correlations to mean that even in this active war zone—to which enemy fighters may flock—ISR is serving as a sufficient deterrent to the adversary pursuing direct contact with allied forces.

For a correlation to be conducted and have meaningful results, all data must be coded the same way and have similar meaning. There are two main variable types in statistical analysis. The first is called a continuous variable, meaning the data ranges from zero to an infinite number of instances. The second is called a categorical variable, meaning the data is sorted in a range (e.g., 1-10 instances = 1, 10-20 instances = 2, etc.). For instance, while Fahrenheit and Celsius are both continuous variables, comparing their results does not account for differences in measurement criteria. This would render the comparison and results useless. The value of AI/ML/DL in this example is that the computer could automatically clean up the data for statistical

For a correlation to be conducted and have meaningful results, all data must be coded the same way and have similar meaning.

evaluation, or as with the case of NLP, put voice, video, or text data into a form that could be analyzed.

Furthermore, discovering relevant correlations depends on the data that is deemed useful for inputting into the system. Continuing on the example above, suppose sudden changes in temperature (listed by degrees Fahrenheit) also significantly correlate to all of the items above. That is, suppose changes in temperature are statistically and significantly correlated with enemy presence, allied ISR, and skirmishes. One might conclude—and as data from Afghanistan shows—seasonal differences can be a relevant predictor of enemy activity. This finding directly supports a notion of a "fighting season." And, most importantly, this finding advises military strategists on ideal future application of ISR assets and ally presence during specific seasons.

While predictive analytics may produce meaningful outcomes, one should never presume that any correlation—even on an incredibly large mound of big data information—infers causation. Even with the maximum amount of statistical confidence where relationships are found to correlate with each other at an almost 1:1 clip (1.000), in an open social system, one variable never causes another to happen. The variables are simply the things analysts know about, can reasonably measure, and hope contribute to the context in which human decisions are made. The stronger and more persistent the statistical correlations are, the more the analyst can gain confidence in the model.[18] However, as mentioned before, all that statistical correlations accomplish are modeling past behavior and decision structures. New interests, structures, opportunities, and obstacles can arise that alter the decision calculus, making the analyst's model obsolete.

Therefore, big data and predictive analytics should always be deemed a tool and not an answer to strategic questions. Why? Using a classic analogy, if enough spaghetti is thrown against the wall, something is bound to stick. The same can be said for predictive analytics. If one combines enough data, something is bound to correlate with other items. While there might very well be contextual reasons to explain the statistical associations, it is just as likely that the correlations are spurious, or false positives. Undertaking a predictive analysis without sound theories of association and clear research questions in mind is fraught with problems. The spaghetti-on-the-wall version of big data analysis—throwing everything into a dataset to see what fits and where—essentially means expending precious labor hours that SOF do not have in sufficient supply to backwards justify potentially spurious correlations. Starting with the why—an informed theory or hypothesis—and then aggregating appropriate datasets to fit an informed model is generally the better approach.

Without a research question as a roadmap, statistically significant results derived from statistical software create a slippery slope. The danger is that the results might seem objectively important since they are relatively free from analytical bias, and they can be seductive to both planners and leadership looking for a strategic advantage (contextual frame notwithstanding). The results are not useless, per se, but should be further explored via additional data collection (surveys, focus groups, etc.) to determine their utility, if any. If the correlations are determined to have merit, the new models can then marry all of the research questions for re-examination.

The Limits of Predictive Analytics

Predictive analysis is not a magical approach where computers make decisions in place of humans. Rather, "predictive data analytics is the art of building and using models that make predictions based on patterns extracted from historical data."[19] AI, ML, and DL are simply tools that decision makers can use to more confidently base their choices on data. The benefit of ML approaches appears mainly in that these algorithms often see patterns in historical data better and faster than humans.

However, it is important to remember that the outputs are all probabilistic in nature and that choices and assumptions made by programmers may add biases into the models. Therefore, programming teams must be conscious of how their decisions may influence the overall outcome. For example, there is usually more than one model that is consistent with any training data used. The analyst may address this issue by including a bias function that balances the noise added by all the assumptions made by the analyst, guiding the algorithm to choose certain models over others.[20]

The importance of exercising caution with data findings cannot be stressed enough. Recommending any military action from data devoid of a proper framework (e.g., research questions or theory) could literally cost lives, as decisions tied solely to a statistical engine could be made on spurious correlations. In other words, AI/ML/DL

The importance of exercising caution with data findings cannot be stressed enough.

systems cannot—repeat, cannot—replace the need for subject matter experts and seasoned analysts who know their topics. By design, statistical software does not reason consequence or context, and software-led decision-making could lead to a tail-wagging-the-dog scenario. This is why theory and methodology are essential components of predictive analytics. Theory should be the foundation to which research questions are bound and explained. Here, theory refers to social or natural science phenomena previously tested and validated to explain similar problem sets and populations.

Predictive analysis by design does not test theory; rather, its objective is statistical assessment of the variables provided to determine if patterns have a history of repeating or intersecting. As a result, there should always be a theoretical base from which the analyst draws. For example, assume a dataset provides information about a Syrian population vulnerable to radicalization. A predictive analysis of aggregate survey findings (over several months) from these individuals may produce statistically significant results in a respondent's stated perceptions (e.g., on a scale of 1-5, ranging from disagree to agree) of safety, security, access to employment, mobility, and desire to join a terrorist group. A psychological operations professional trained in this doctrine might point to Maslow's hierarchy of needs as a way to make sense of the findings, as these questions all relate to Maslow's theoretical continuum. With a theoretical base to draw from, this also advises the analytical professional on how to conduct future analyses, ones where the theory can be further applied toward a better understanding of similar concepts and populations.

Similarly, methodology should never be taken for granted. A favorite expression of quantitatively leaning social scientists is garbage in, garbage out. This adage highlights that any data that is not properly structured for analytical purposes is apt to produce meaningless, or worse, wrong results. Despite the seemingly obvious nature of this statement, poorly organized and structured data is unfortunately a common occurrence in predictive analytics. Many datasets include questions containing more than one concept (double-barreled) or worded in a leading fashion (toward a desired outcome), items possessing different scale types (e.g., 5 point versus 7 point versus continuous data) and/or an item that does not really encapsulate the problem being addressed, or reverse coding (where a '1' in one question may mean strongly disagree and the exact opposite in another). While statistical analyses of such items can occur—and statistically significant correlations may emerge—these should never be considered predictive or even mathematically sound.

While predictive analysis can be useful in a military context, it is important to understand the strengths and limitations of this tool and when and where it should be applied. It should always be employed to inform versus advise, be grounded in some aspect of theory to enable future use, and be housed within a rigorously designed methodology to ensure compatibility with data analysis software. Finally, one should never infer causation from even the strongest statistically significant correlation.

The best use of predictive analytics is when a diverse, cross-functional team develops a model. Grouping data scientists, operators, planners, social scientists, and statisticians ensures assessments match the operational and strategic intent and that they are used appropriately to inform leadership. Cultural experts can explain findings also derived from statistical outliers. A diverse team also maintains a proper checks-and-balances approach, where one individual does not over- or underestimate findings in light of a bias-driven concept. Lastly, theory, methodology, and most importantly, the purpose of the predictive analysis should be addressed before any data is gathered and ultimately assessed. Working backwards should be done with extreme caution as it risks justifying activities and actions that are more coincidence than circumstance.

Common Artificial Intelligence/Machine Learning Applications

With these basic elements of AI/ML now explained, it is possible to better interpret the common applications SOF are likely to employ.

Social Media Analytics

With information being voluntarily provided by thousands of consumers, focus groups and polls are not needed in the same way as they used to be. Instead, corporations get a sense of users' needs and opinions by using tools such as NLP to identify the meaning and sentiment behind tweets or posts. They may then, for example, analyze the resulting data to gather customer sentiment towards brands and products and tailor their message to attract new audiences. This uses advanced ML algorithms to not only identify hashtags but also understand the context of a user's post. For this specific task, there are key performance indicators that the software will look for and will prioritize posts based on relevance to the client. Relevance is measured by assigning probabilities and weights to the various features. For example, there may be two companies with very similar names. The software will learn to identify cases where a specific company is referenced over another. These indicators (in order of increasing priority) could be the number of views on different platforms or the number of times users interact on a product post. Having access to this information enables companies to tailor their marketing more effectively to each individual.

Biometrics

Digital security based on biometrics like fingerprints, voice, or facial recognition can be improved using AI. Traditional methods have loopholes that can be exploited. For example, a facial recognizer cannot distinguish between a live person and its image. Improvements done by implementing computer vision and advanced facial recognition tools have given us novel software and apps which can recognize faces, expressions, and facial movements. This is done by looking at features such as the distance between the eyes and shape of the nose and lips. By learning such features over time, the software can successfully recognize the person. Voice recognition, however, is different because the features to be looked at are slightly less obvious and unique: voice is characterized by pitch, silence periods, cadence, and tone. Separately, these parameters don't tell us much, but together they can successfully help differentiate between speakers. Most products that use biometrics are designed to recognize users under different circumstances (different hairstyles, applied makeup, changes due to illness, etc.). However, having complete facial or voice recognition software requires large databases which are prone to security breaches. Applying these techniques can eliminate passwords for more secure verification of users. The onset of AI with its ability to learn over time has now made available newer biometric parameters called behavioral biometrics that measure, for example, keystrokes, pattern dynamics, and voice print.

Social Network Analysis

AI is becoming widely used in the commercial world to increase engagement on different social network platforms such as Facebook, LinkedIn, Twitter, and various dating apps. Facebook and Twitter are currently the greatest sources of social media analytics. Analyzing users' internet activity makes it possible to highlight content which is most interesting to them. It also helps write e-mails and messages faster by predicting the remainder of commonly used phrases. Featuring relevant posts requires ML and some data analysis in order to learn the users' responses over a brief time and constantly update it. NLP, as previously discussed, can aid in completing user sentences while typing, and it understands the sentiment behind what has been typed to suggest replies and emojis. Social networks and dating apps can suggest people to be friends, follow, or date based on graphs of mutual connections and interests.

Personnel Management

Recruiting portals and job boards attract employers and employment seekers by trying their best to feature relevant companies to the candidates and best matches to the companies. The success rate of hiring determines the success of that portal. These processes are enhanced by ML, which learns from the company's profile and description about the kind of candidates it is looking for and looks at candidates' resumes to see what experience they bring to the table. When a candidate searches for a keyword, the most relevant jobs show up. When the company receives applications, the portal filters out the best matches. This uses text analysis and data mining to organize the results. Online interviews are now conducted at larger corporations as the first or second round of engagement. Candidates interact with a computer, and their responses are recorded and analyzed for signs of nervousness, stress, authenticity, preparedness, and confidence.

Robotics

Robots have long been fashioned in popular culture as the pinnacle of invention. However, a robot, strictly speaking, is simply a machine that completes a task repeatedly without human intervention and without error. With AI that now learns from its environment, a robot can learn by doing a task in the manner of ML of any kind. This creates possibilities for many new ideas of automation in the industry. Robots can be trained to have conversations with humans and other robots, and they can learn how to walk and perform complicated tasks with programming that helps prioritize tasks when they are faced with decisions.

Behavioral Pattern Analysis

AI is used for things that humans are instinctively good at—things that even science has yet to understand the way they are done by humans—like distinguishing between objects or remembering things. One theory is that in order to achieve human-like intelligence in machines, they also need to be injected with human-like emotions.[21] AI-powered models identify different emotions in speech, facial expressions, and even text, as well as learn how to respond to them. For example, if a robot is to learn how to respond to user emotions on its own, it will learn by observing behavior and deriving patterns from it. Neural networks—computational and mathematical models of the human brain—are used to "teach" a robot, either by explicitly providing rules and supervision or letting it learn from its environment unsupervised. The robot may be taught the various rules of emotions in the form of a graph and how to access the graph. But first the robot needs to deduce what emotion is being displayed. For that, behavioral pattern analysis requires a combination of computer vision, NLP, and speech recognition.

AI-powered models identify different emotions in speech, facial expressions, and even text, as well as learn how to respond to them.

Swarming Technology

Swarm technology is the collective operation of separate, decentralized units, like bees or ants. This technology manifests itself in distributed computing, swarm robotics, and the Internet of Things. Current technologies are primitive and not easily scalable, which means that they cannot be expanded easily and diversified to do different tasks. The flip side to this is that distributed computing can be used to power AI. Companies and research groups are harvesting idle computers all over the world to solve small parts of the bigger, more complicated problems of AI.

Machine Diagnostics

Manufacturing and processing industries all over the world require regular maintenance of their machines and components. AI-based control systems have now begun to predict upcoming failure of machines and alert repair personnel, saving money on human inspectors and avoiding human error. Control systems deployed for this task monitor a machine's usage and predict wear and tear of parts based on their expected life span. Workers are alerted when a critical stage is reached. The AI that does this follows a simple rubric: the machine states are defined and the action of using the machine over a period depreciates the reward for using it. Once the reward becomes less than moving to the next state, the AI moves to the state with higher reward. This state would indicate moderate wear and tear (thus giving warning to the user) and over time, the reward will diminish so much so that the AI will move to the critical stage which will create an alert for immediate requirement of new components. This is known as reinforcement learning and is used to train the model underlying the AI to "think" and work like humans.

Conclusion

Big data is a catchall term used to describe—and unfortunately muddy—a number of precise technical, statistical, and social science concepts. What distinguish big data from very large data files are volume, velocity, variety, value, and veracity. These characteristics require an entirely different handling and analysis process than currently exists in the SOF enterprise. Many people believe that computer-driven automation, AI, ML, NLP, and DL capabilities will provide leaders with precision insight about complex phenomena to improve decision-making. That is, they believe the right data is available, but computers are needed to demystify what really matters so leaders can make the best informed decisions possible. This spaghetti-on-the-wall image of big data analysis is in actuality the opposite of how properly functioning big data analytics operate and is often driven by the myth of predictive analytics.

Although the name *predictive analytics* suggests that AI/ML systems can causally determine relationships between variables and foretell future events, the truth is that they can only correlate relationships and determine a probability of current patterns replicating. These are important distinctions. Moreover, the challenge for SOF occurs in open systems, meaning the number of variables impacting the problems it faces is unknowable. It is impossible, therefore, to accurately model future behavior, though logical inferences can be made based on history. The tools and applications derived from AI/ML techniques can help illuminate aspects of the operating environment that are not obvious, but they require informed design. As a result, big data cannot be an easy button as much as an elegant analytical tool given the right data infrastructure and cross-functional team organization.

Endnotes

1. Annika Richterrich, *The Big Data Agenda: Data Ethics and Critical Data Studies* (London: University of Westminster Press, 2018), 4–6.
2. See Michael Landon-Murray, "Big Data and Intelligence: Applications, Human Capital, and Education," *Journal of Strategic Security* 9, no. 2 (Summer 2016): 92–121.
3. Richterrich, *The Big Data Agenda*, 6.
4. Doug Laney, "3D Data Management: Controlling Data Volume, Variety and Velocity," *META Group File 949* (2001), https://1library.net/document/y4e8w10q-big-data-challenges-big-data-adoption-smes.html.
5. Landon-Murray, "Big Data and Intelligence," 95.
6. *Automic® Workload Automation Benefits of Automating Data Warehouses and Big Data* (San Jose: CA Technologies, 2019), 4, https://docs.broadcom.com/doc/benefits-of-automating-data-warehouses-and-big-data.
7. See Shukla Shubhendu and Jaiswal Vijay, "Applicability of Artificial Intelligence in Different Fields of Life," *International Journal of Scientific Engineering and Research* 1, no. 1 (September 2013).
8. William F. Clocksin, "Artificial Intelligence and the Future," *Philosophical Transactions: Mathematical, Physical and Engineering Sciences* 361, no. 1809 (2003): 1721.
9. Paul Thagard, "Philosophy and Machine Learning," *Canadian Journal of Philosophy* 20, no. 2 (June 1990): 264.
10. Christopher M. Bishop, "Model-Based Machine Learning," *Philosophical Transactions of the Royal Society* 371, no. 1984 (February 2013): 2.
11. Aravind J. Joshi, "Natural Language Processing," *Science* 253, no. 5025 (September 1991): 1242.
12. Fern Halper, *Advanced Analytics: Moving Toward AI, Machine Learning, and Natural Language Processing* (Renton: Transforming Data with Intelligence, 2017), 7.
13. Fernando Maymí and Scott Lathrop, "AI in Cyberspace: Beyond the Hype," *The Cyber Defense Review* 3, no. 3 (Fall 2018): 75.
14. Steven Strogatz, "One Giant Step for a Chess-Playing Machine," *The New York Times*, 26 December 2018, https://www.nytimes.com/2018/12/26/science/chess-artificial-intelligence.html.
15. Victoria Chick and Sheila Dow, "The Meaning of Open Systems," *Journal of Economic Methodology* 12, no. 3 (2005): 367.

16. Chick and Dow, “The Meaning of Open Systems,” 366.

17. Walter Buckley, David Schwandt, and Jeffrey A. Goldstein, “Society as a Complex Adaptive System,” *Emergence: Complexity & Organization* 10, no. 3 (2008): 89–94.

18. Gary King, Robert Keohane, and Sidney Verba, *Designing Social Inquiry: Scientific Inference in Qualitative Research* (Princeton: Princeton University Press, 1994), 6–7.

19. John D. Kelleher, Brian Mac Namee, and Aoife D’Arcy, *Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies* (Cambridge: MIT Press, 2015).

20. Nasser M. Nasrabadi, “Pattern Recognition and Machine Learning,” *Journal of Electronic Imaging* 16, no. 4 (2007): 049901; see also Christian Robert, “Machine Learning: A Probabilistic Perspective,” *Chance* 27, no. 2 (2014): 62-63.

21. See for example Clocksin, “Artificial Intelligence and the Future,” *Philosophical Transactions: Mathematical, Physical and Engineering Sciences* 361, no. 1809 (2003): 1736–1739.

Chapter 2. The Basics of Big Data Modeling

Colonel Mark Zais, PhD; Dr. Karl Aspelund; Mr. Pedro Cesar Lopes Gerum; Ms. Nishka Uberoi

Artificial Intelligence Relies on Models although All Models Are Wrong, but Models Can Be Useful

At their core, artificial intelligence (AI) and machine learning (ML) systems are just mathematical models—really, really complex mathematical models that require computers and code to make the models work. But as mathematical models, they rely on some basic statistical concepts and limitations that are easier to understand. Every model corresponds to a specific problem and accounts for what personnel judge to be the fundamental variables impacting the problem. In many cases, AI/ML projects are comprised of numerous subordinate models that interact with each other. In a general sense, the two main components of such models are objective functions that represent (a) needs, which signify the importance of each aspect of the problem to the client and (b) constraints, representing real-world limitations in terms of knowledge of the problem, access to data, and computational capability. An AI/ML project is not the brainchild of just the data scientist. Rather, most projects come from a client's problems or needs that the data scientist believes can be addressed using cross-functional expertise as interpreted through AI/ML and data analytics techniques.

The objective of chapter 2 is to explain how mathematical models are generated and the problems associated with them as they attempt to represent a portion of reality. The first section reviews some key modeling concepts, namely regression, inductive analysis, and deductive analysis. These elements form the skeleton of every AI/ML project. The second section introduces the issues associated with structured data, unstructured data, and the problem with garbage data. These elements constitute the meat of an AI/ML project, and even the best AI/ML project will fail if the data content it processes is poor. The third section discusses the various types of bias that inevitably become infused within mathematical models. Finally, the fourth section describes the difference between question-driven and dashboard-driven AI/ML systems. Vendors often focus on the elegance of the dashboards they create on top of an AI/ML system, so knowing how they trained their products for demonstration is crucial for making informed AI/ML/natural language processing (NLP)/deep learning (DL) purchasing decisions.

How Artificial Intelligence Models are Designed

AI and ML projects are simplifications of the real world that consider the trade-offs between representing reality accurately and having a model that is trainable and efficient. To be clear, this trade-off means that every mathematical model is a purposefully scoped, incomplete, and partial representation of reality—not reality itself. The seductiveness of AI/ML solutions is that they generate numbers, charts, maps, and other outputs that create the perception of certainty, but in reality, they only present a limited frame of the world generated by the biases of the personnel who created them given the available data. For discrete, replicable tasks operating in fixed time and place environments, AI/ML can be very powerful. Such is the case with the find, fix, finish, exploit, analyze, disseminate process; surveillance functions; mechanical repair and maintenance data; and social media trend analysis, for example. Decision-making with such outputs can most certainly be enhanced. However, mathematical models become less useful in more open, medium- to long-term challenges for which the number of relevant variables is simply unknowable or for which the variables themselves (people) can purposefully change. At the high operational to strategic levels, AI/ML solutions consequently lose their efficacy and should be interpreted cautiously by decision makers.

A common assumption used in many models is that certain features of a problem are linearly dependent. That is, the models assume that there is a causal direction between a dependent variable (the outcome) and the independent variables (or subordinate variables assessed to affect it) that can be known and measured. However, many problems have features that are correlated in a more complex, non-linear manner where instead the variables mutually affect one another in an iterative cycle over time. One example is the geometric relationship between the value of a vehicle and the time the vehicle has been owned. Although clearly correlated, the depreciation of the car's value occurs at a fast rate initially but then slows down after a few years have passed. While it might be possible to correlate the value of a particular make of car over time from aggregated data, it would be impossible to predict on the day it was purchased the value of a specific car in five years because its wear and tear, accident history, and mileage could not be foretold. In the end, a model's assumptions need to be carefully assessed so that the noise produced—the simplification of reality—does not overly affect the quality of the output.

Classification and Regression. Most ML algorithms tackle one of two categories of problems: classification problems and regression problems. The main difference lies in the characteristics of the output sought. Classification models predict outputs with no well-defined order such as identifying the type of tumor a person might have or recognizing which state a vehicle's license plate is from. Similarly, one might consider handwritten digit recognition. The AI/ML output is numbers, but it is not a regression problem. Rather, it is considered a classification problem as the order of the digits is of no importance. The AI/ML project simply tries to connect each image to a specific label, and the number of categories should be very small in relation to the size of the dataset, allowing the computer to confidently pair patterns with each category. Project Maven, for instance, aligns with a classification-oriented AI/ML problem. The system seeks to establish a normal pattern of life in an area so that analysts need not devote their limited time to monitoring, well, nothing. Rather, the AI/ML is supposed to determine what an abnormal incident looks like given the "normal" baseline and then classify it as requiring further analytical attention.

Regression models, on the other hand, are designed to predict continuous outputs such as the price the market is willing to pay for a house or the desired angle of a wheel in a self-driving car. Here, order matters. For example, everyone can agree that a house costing $5 million is more expensive than one that costs $3 million. One number is simply larger. In the military context, a commander might want to know where on the map he or she is most likely to encounter improvised explosive devices (IEDs), hostile populations, or the most concealing terrain. Here, predictive analytics are employed using regression modeling. A typical AI/ML model digests dozens to hundreds of variables to determine which are most correlated with the commander's problem, and an algorithm produces number values, or continuous outputs, that can be compared. When paired with geospatial software, it is possible for heat maps to be created based on the AI/ML model's assessment of the variables used in the model.

In general, regression problems are more efficiently solved when the data is robust and complete, allowing the computer to accurately assess the value of each data point, although the use of sparse or incomplete data can be used in sufficiently large datasets. Unfortunately, good data is not always forthcoming, which means there is a good chance that the models are wrong. Analysts and leaders need to be clear about the strengths and limitations of the models based on their underlying assumptions and data sources.

Unfortunately, good data is not always forthcoming, which means there is a good chance that the models are wrong.

While most problems fall into either the classification or regression category, some problems, such as those that have clearly ordered yet no continuous outputs, are often solved with a mix of the two types of models. For instance, to help a commander determine how to navigate an area, a predictive analytics model might provide the probability of IED strikes along different roads—produced as comparable continuous numbers—and identify a tribe with a history of positive interaction with U.S. forces—produced as a classification output. Both classification and regression models largely benefit from an increase in the size of the dataset used for training, but this depends on time, funding, and data that accurately reflect the environment against which the model is being applied.

There are two main statistical approaches to designing AI/ML models: inductive and deductive analysis. While the average user or leader does not need to be familiar with the algorithms and coding that power the different models, they do need to know the philosophy of how each works to assess the strengths and weaknesses. The next two sections describe the philosophy of science underlying each approach.

Inductive Analysis. Inductive analysis begins with the presumption of ignorance about cause and effect in a system. Some phenomenon is observed—riots, IED attacks, local resistance to extremism—and it is hoped that empirical data can objectively determine the variables contributing to its occurrence. The benefit of inductive analysis is that it attempts to mitigate analytical bias because the data determines relevant correlations, not the analyst. Additionally, analysts and leaders are not required to know anything about the environment beforehand because they are supposed to be agnostic anyway.

To give the AI/ML algorithm a predictive capability, it is first necessary to "train" the data. That is, the model is taught through multiple data points and processing iterations to recognize the variables most highly correlated with the output sought. In an inductive analysis, however, the trick is to bring as many variables into the analysis as possible and let the algorithm determine statistical relevance. When it comes to big data, determining relevant correlations and patterns is impossible for humans to comprehend. The computer, in the inductive step, iteratively finds patterns and matches them to the given examples until it is confident about patterns that most closely match every single data point's output.

ML provides a simple, efficient way for algorithms to learn patterns that might represent correlations instead of coders writing thousands of if-then statements of causation. The algorithms can account for minor patterns that humans often overlook or are not capable of finding. Many variations of common algorithms, such as recurrent or convolutional neural networks, aim to help the computer train faster and more efficiently. As an analogy, the variations represent the different methods a parent might use to teach a child. Inductive analysis of this kind is crucial for the ML algorithms to provide good outputs.

Deductive Analysis. In contrast, deductive analysis begins with the analyst bringing ideas about cause and effect to the table and incorporating the assumptions into the model before statistical analysis is conducted. In this case, the analyst has a hypothesis that certain variables are important and uses the data to test whether the assumption is valid based on statistical testing. Deductive analysis typically derives from deep experience with a problem or from a series of logical inferences. Whereas inductive analysis casts a wide net of variables at the problem, deductive analysis most often utilizes a far narrower range of variables in the model since the analyst already has a sense of causation in mind.

Ideally, AI/ML modeling should incorporate both approaches as each has deficiencies. A completely inductive analysis is likely to result in spurious correlations—statistically relevant results that have no basis in fact. Without some familiarity with the problem, the analyst will have no way to judge whether the algorithm's results make sense. The inductive approach encourages analysts to investigate factors they might ordinarily overlook. On the other hand, a purely deductive approach could result in a model riddled with bias with no way to correct for variables that should be included. An elegant AI/ML approach would leave space for both types of analysis, but rarely is this the case. Dashboard-based AI/ML systems deserve special caution because they tend to be derived from the vendor's deductive analysis and are trained against data that matches the assumptions. While they tend to look impressive in demonstrations, they become prone to missing new, emergent variables over time precisely because they could not have been known to be relevant when the model was designed.

How Machine Learning Enables Artificial Intelligence Models to Adapt

How do AI/ML models adapt to reality? Part of the answer is that not all AI algorithms are designed to adapt, which is why it is essential to distinguish the ML and DL models as subsets of AI. This leads to a confusing, "All machine learning is artificial intelligence, but not all artificial intelligence is machine learning," statement, but this is essential to recognize. The other part of the answer unfortunately depends on whether the problem at hand relates to a closed or open system. For most closed system AI/ML functions, it is possible to record, measure, and code data in a structured way because the system can be controlled and monitored in a fairly precise way. Processing new data and updating the model's algorithm can be written in the code in a relatively streamlined way in such cases. In open systems, this is harder because the data does not come neatly packaged because it cannot be controlled or measured. In other words, the data is unstructured and oftentimes cannot easily be formatted for processing. Since the structured-unstructured aspect of big data is a crucial component, it is worth spending some time discussing this characteristic.

Structured Data. Structured data is data that can be represented by rows and columns and is usually located in relational databases. Spreadsheets often underlie these databases, and examples of data include numbers, dates, and text strings. Some applications, such as those dealing with customer relationship management, may allow for transformations of unstructured data to structured data based on keywords correlated to themes or codes. Structured data tends to be easily searchable by humans using simple search queries or via searchable algorithms. In many cases, basic AI is sufficient for the analytical task at hand because the intention is to identify patterns or trends or to highlight points of interest. The algorithm does not need to adapt, just accurately complete the desired function automatically. In addition, updating structured data tends to be easy; a simple query is usually enough to update one or several data points.

Most early versions of prediction methods relied on structured data because programmers could more intuitively tell the computer how to use each piece of data separately and define their correlation. Some new ML methods still use structured data in their predictive analysis with the advantage of being able to more efficiently find the correlations among several different features—or the different ways patterns interact—that would take humans a considerably longer time to clarify. A major disadvantage of algorithms that use structured data is the need for pre-processing raw data that is usually not clearly structured. Since basic AI models are generally able to tackle problems involving structured data, ML approaches are most effective when dealing with unstructured data.

Unstructured Data. Unstructured data is essentially every other kind of data that can be obtained outside of a spreadsheet format. Some examples include text files, emails, social media content, mobile data such as location or app usage, images, audio files, slideshows, scientific data from atmospheric sensors and space exploration, and traffic data. Although all data has an internal structure of some kind, unstructured data does not have an internal structure clearly defined within the models. Unstructured data tends to be larger and usually takes up more storage space than structured data. Furthermore, updating unstructured data points is usually complicated and often the whole file needs to be replaced. However, unstructured data sources are usually much easier to obtain and need less data processing work unless the intent is to make them structured.

The main purpose of NLP is precisely to efficiently recognize and package unstructured data for further processing. NLP is dedicated to empowering computers to understand spoken commands and the structure of language and accurately interpret the context of conversation whether in audio or text format. The lack of a clearly defined

The main purpose of NLP is precisely to efficiently recognize and package unstructured data for further processing.

internal structure makes it hard for traditional data mining models to define meaningful patterns in such data. Recently, an increasing number of studies have tackled the problem of dealing with unstructured data. ML methods, benefiting from the rapid increase in computational power, have been particularly successful at this. These methods can order unstructured data in an efficient way for computers to understand, even if humans sometimes find it difficult to understand the pseudo structure created by the algorithms.

Garbage In/Garbage Out. Even though AI/ML models have improved markedly in recent years, filtering the good data from the bad is still of major importance for the robustness of any model. AI/ML systems are entirely dependent on the data they are provided, so it is essential that full consideration be given to the collection and processing of data on the front end, not just to the output on the back end. If the intention is for an AI/ML model to find patterns in data to predict future behavior, and if the model is given data that inaccurately represents reality, it will find spurious correlations matching the erroneous information and lead to deeply flawed analysis and decision points despite statistically relevant results. The common expression garbage in, garbage out is thus very true for ML models—the more garbage the model is fed, the more garbage it delivers as output.

Sure, Silicon Valley Is Impressive, but It Cannot Predict the Future

For many leaders in the Special Operations Forces (SOF) enterprise, the allure of big data solutions derives from the experience they have with predictive analytics while shopping online, engaging in social media, or while reading the news. Each of these experiences invariably includes a window exclaiming, "You Might Also Like …" to buy, chat with, or read, respectively—and many times the window is right. These predictive analytics windows are driven by ML systems, and they are impressive. But they have a secret: they work because users voluntarily profile themselves using structured data so that the company can employ inductive analysis on recent purchasing or viewing patterns. That is to say, the ML works because the users are playing right into the hands of the system.

Silicon Valley has become the symbol of any company that utilizes AI/ML techniques and big data analytics to create a seemingly personalized online experience for users. Silicon Valley's most famous companies earn their money not by the product they offer to the user but by selling the marketing, demographic, and preference indicators to which individuals voluntarily give them access in the (mostly unread) user agreements. These companies package the information in structured, geolocated, and time-stamped form and sell them to companies for myriad purposes. Oftentimes, the data serves as key variables in regression models that companies can apply against their product's viewing and sales records to generate a predictive model of what people want to consider based on the statistics generated from the profiles provided by Silicon Valley.

Think about it: what is interesting and valuable to a single twenty-year old with no children would undoubtedly be different from the same person at age thirty-three with two kids, a dog, a mortgage, and aging parents. The combination of search queries, social media chats, and traditional phone or web-based marketing surveys yields truly insightful demographic information about target audience segmentation. When incorporated into an inductive regression analysis with ML algorithms that continuously update the model, the online user experience has the ability to appear persistently relevant and personalized. In truth, the model cannot predict a user's purchasing patterns perfectly, but it can make informed assumptions about the probability the user might be interested in something that others of the same profile found interesting.

The Silicon Valley system in essence creates enough of a closed system that it seems prescient and predictive. While not every user decision can be controlled, aggregate user behavior creates statistically relevant models based on already structured data. Compare this, however, with the environments in which SOF operate. These environments tend to be characterized by weak institutions, poor information, fractured communication systems, language barriers, and cultural differences. In other words, they tend to be extremely complex environments with mostly unstructured data. The Silicon Valley model of predictive analytics breaks down in such circumstances because data does not stream efficiently into the system, and there typically is not the target audience segmentation necessary to make the statistical analysis feasible. This is compounded by the fact that the knowledge management system in the SOF enterprise is built on distributed legacy systems instead of a corporate culture that places data at the heart of the organization as in Silicon Valley. While there are areas and capabilities in the SOF enterprise that currently match the requirements for big data analytics, the system as a whole requires major restructuring before it could even approach the idea of predictive analytics for many of its most pressing challenges.

Bias in Models

As previously mentioned, the issue of bias is intrinsic to the process of developing models since each requires conscious choices to be made in the representation of reality based on limited time, resources, and data. Consequently, algorithmic bias has developed into a significant issue in the evolution of AI/ML modeling. It is a reasonable concern that algorithms might conceal hidden biases that influence consequential decisions a leader might make. Users and stakeholders are often too willing to trust mathematical models because they believe that the models remove human bias. Regardless of how data is controlled and the parameters are set, every algorithm will have some form of bias.

Regardless of how data is controlled and the parameters are set, every algorithm will have some form of bias.

Most algorithms make predictions based on generalized statistics from the most readily available information. When thinking of the term bias, it is natural to think of it as something subjective. In the context of models, analysts learn that bias is not only something to try to avoid but also that it is something that can be accounted for with mathematical modeling and "objective" statistics. Yet model bias is a fairly broad issue that encompasses many sources of bias.

One of the most basic and widely used supervised ML models is the linear regression model previously mentioned. Given a range of independent (or explanatory) variables, the dependent variable (or outcome) is predicted using the model. Linear regression models have an inherent flaw: they assume open systems can be reduced to just the independent variables in the model resulting in bias in the process. In fact, there are other independent variables not represented in the model that impact the outcome and result in deviance from the prediction. These deviations are called error values and are collectively hidden as the standard deviation from the regression model's prediction line.

Very few real-world problems have a simple linear relationship, so the estimate from the equation above reasonably results in some bias due to the rigidity of the linear regression model itself. In general, the more flexible the model becomes, the more capable it is of improving fit and reducing bias. The purpose of ML applications is precisely to adapt and improve the model as more information is ingested.

There is also a natural bias based on the fact that statistics utilizes sample data to make the models. Samples are small portions of an overall population. For example, it would not be prohibitively expensive to model something with a population of just one hundred units, and a sample might not be necessary in this case. However, modeling the attitudes of America's three hundred million person population would be prohibitively expensive, so taking a sample of 1,500 people is much more feasible. In making this time and cost trade-off, the results of the sample are very probably going be wrong to a certain degree, but the error is generally acceptable so long as it is not too large. Developers often estimate the parameters of a model using the method of least squares, also known as ordinary least squares. With parameter estimation, the developer computes a standard error or confidence interval to estimate how well the model represents the overall population. Developers test hypotheses about these parameters by computing test statistics and their associated probabilities (p-values). As previous discussed, it is important to note that bias affects parameter estimates, but it is also important to understand that bias affects standard errors, confidence intervals, test statistics, and p-values. These latter areas are often overlooked.

Assumptions

Bias can be introduced in multiple phases of mathematical modeling. It may come directly in the form of modeling assumptions or indirectly from sources, such as poorly constructed sampling procedures and training datasets. The assumptions that model developers and organizations infuse into the AI/ML systems are the most pernicious because they are so hard to detect and challenge. This is especially true when a vendor utilizes established analytical frameworks, such as DIME or PMESII,[1] because they are culturally comfortable, not because they most accurately model the problem at hand. There are many types of bias that reinforce assumption errors.

Confirmation Bias. Confirmation bias is frequently discussed in the psychology literature and relates to the tendency of people to seek out and interpret evidence in ways that are supportive of existing beliefs, expectations, or a favored hypothesis.[2] It occurs when analysis is conducted to prove a predetermined assumption held by the person performing the analysis. Confirmation bias is an unintentional, or sometimes intentional, desire to prove or substantiate a hypothesis, assumption, or opinion. Cognitive science has proven that there is a natural tendency to search for new information in a way that confirms the current hypothesis and to irrationally avoid information and interpretations that contradict established beliefs.[3]

Sample Selection Bias. Sample selection bias is a type of bias caused by choosing non-random data for statistical analysis. This type of bias occurs when data is selected subjectively or when there is a flaw in the sample selection process, and it is a reason why random data samples are so important. A common assumption in the design of learning algorithms is that the training data consists of samples unrelated to the data being used by the model to make predictions. In fact, the model may get tainted by training against data that is not a random sample representative of the true distribution of the population.

Inductive Bias. In ML, the term inductive bias refers to a set of (explicit or implicit) assumptions made by a learning algorithm in order to perform inductive analysis. Every ML algorithm with an ability to generalize and adapt beyond its training data has some type of inductive bias. Since a model is designed to highlight a portion of reality, it must be pushed in a direction that prioritizes its specific function at the expense of other possible observations. As a result, models can become self-referential and miss emerging trends because they are specifically designed to focus on certain factors over others. Without a bias of this kind in the model, inductive analysis would not be possible, and predictions for new situations could not be made.[4]

Outlier Bias. Outliers are data values or observations that lie at an abnormal distance from other values in a random sample from a population. Sometimes, outliers are classified as extreme outliers if they fall significantly beyond a certain point along a distribution. Outliers are often bad data points, but they can also contain valuable information about the process or system being evaluated. Before considering the possible elimination of these points from the data, it’s important to understand why they appeared and whether it is likely similar values will continue to appear.[5]

Understanding outliers and distributions is critical in analyzing data. Outliers may negatively bias the entire result of an analysis, or the behavior of outliers may be precisely what is being sought. When modelers understand the relationship of outliers to the distribution itself, they can determine the most appropriate way to deal with them. Outlier bias is particularly prevalent in big data because the larger the dataset becomes, the harder it is to identify and remove outliers.

Overfitting and Underfitting. A model that is too simplistic a representation of reality will produce underfitting. As such, the model does not accurately detect signals from the data or learn by adapting the model to fit the new data—it is essentially a one-size-fits-all model. Overfitting is when a model is overcomplicated or too complex. A model that is overfit will memorize noise instead of learning the signal from the data. Overfitting is one of the most common biases, which can come from something as rudimentary as analyzing multiple hypotheses in data. For instance, if a model evaluates numerous hypotheses, each with a probability of being a false positive, the likelihood of a false positive increases significantly.

Confounding Variables. A confounding variable is a variable that is outside the scope of the existing analytical model but influences both the explanatory and dependent variables. In ML, a confounding variable can be described as an important variable that is omitted but should be included in the predictive model. For instance, suppose U.S. Special Operations Command wants to develop a retention model to predict continuation rates (i.e., the estimated length of service) for its Service members but only uses gender, education, and occupation specialty as explanatory variables. It is well established that the branch of Service is an important variable because some branches have statistically different continuation rates than others. For example, the average length of service for members of the Air Force is greater than the average length of service for members of the Marine Corps. Therefore, branch of service is a confounding variable that should be included as an explanatory variable in the model.

It's the Question That Matters, Not the Dashboard

Circling back a moment to the beginning of the chapter, it is important to remember that AI/ML solutions are models, and all models are wrong. More fully stated, every model represents a slice of reality that investigators are trying to analyze. As mentioned earlier, this means that not all variables can be captured, but the model only needs to capture the ones that are meaningful for the problem at hand. Unfortunately, most AI/ML solutions are packaged behind elegant, attractive user interfaces—called dashboards—which make them look impressive. As models, dashboards represent a portion of reality and must therefore be approached with a healthy dose of skepticism. As noted in chapter 1, AI/ML solutions should be driven by a question, problem, or challenge, and multidisciplinary teams should be assembled to tackle them. While dashboards often seem to incorporate different perspectives and data streams, they tend to be difficult to adapt to new and changing circumstances.

What Dashboards Tell You

What a dashboard does for the user is compress disparate information into a structured visual interface. The structured dashboard therefore predetermines what is important for the analyst to focus on, prioritizes certain kinds of analytical tasks over others based on the tools written into the interface, and restricts the information to which the analyst has access based on the model's requirement to structure data in a particular way. This is not to say that dashboard interfaces are not useful in certain cases. Rather, it is merely to note that they are tools with specific purposes, and no single AI/ML-based dashboard system is likely to provide a comprehensive solution for an organization's analytical, forecasting, or intelligence needs.

Dashboards and the Curse of the Black Swan

The danger with AI/ML dashboards is that organizations can become culturally comfortable with them to the point that personnel believe they accurately represent reality rather than just a small slice of it. The allure of charts, graphs, and word maps is that they appear objective on the surface, and they also tend to make for terrific slide presentations. But as intrinsically flawed models, they will miss important aspects of reality because they are developed with bias built into them, and sometimes the obscure error value or outlier initially thought irrelevant to the model becomes the most important data point over time.

The allure of charts, graphs, and word maps is that they appear objective on the surface, and they also tend to make for terrific slide presentations.

This observation was popularized by Nassim Taleb in his book, *The Black Swan*.[6] As a highly trained financial modeler, Taleb came to accept that modeling as a discipline prioritized the average, ordinary, and normal condition and gave little consideration to the low-probability but high-impact outlier. As a result, financial shocks were difficult to discern and forecast because models failed to account for them—doing so would have been irrational since applying the time, resources, and labor against such low-probability events would have been cost prohibitive for nearly all institutions. Nevertheless, Black Swan economic and political crises do strike but often originate in places most modelers would not even think to consider.

As structured interfaces, dashboards are susceptible to Black Swan events. For all their power to synthesize vast amounts of data and produce adaptive analytics, they are nevertheless developed based on a theory of what matters and what does not, and they are limited by the need to structure data for the visual outputs and the cost of data streams. Again, dashboards might have utility for a period of time, but they should always be treated as tools with inherent limitations.

Asking Different Questions Reveals Different Aspects of Data

For leaders in the SOF enterprise, the natural tension with big data will generally come down to the need to treat operations with a nod to military standardization with predictable, efficient, repetitive tasks assisted by technology versus the reality that big data analytics are most powerful when the data is approached from multiple, non-standard perspectives. AI/ML models and dashboards can be helpful, but they are biased representations of reality, not reality itself. In general, however, military culture expects technology to be the solution for telling the Force what it needs to know or care about. Big data analytics start with the exact opposite attitude precisely to avoid modeling bias. That is, the Force needs to engage the data science teams with quandaries or questions, why the questions matter, and how the analysis would likely inform analyses and decisions.

To make the most of a big data capability, SOF leaders will need to carefully consider how to cultivate a culture of asking questions of the data and to create a data architecture open to a range of AI/ML solutions. The SOF enterprise does not have a shortage of data; rather, it has a data storage and access challenge. While operational concerns seem to compel a focus on dashboard interfaces to address the current fight, a true big data culture would prioritize exploring data from different vantage points and ensure that the architecture would enable manipulating the data for future, currently unknowable questions. Whether for predicting breakdown rates for equipment, optimizing personnel dwell time, forecasting expenses, simulating war games, or tracking social media trends, a SOF big data culture should assume that AI/ML solutions require reframing how the data could be utilized to answer new questions or contribute to emerging requirements.

For example, data collected specifically for the purpose of tracking deployments and dwell time could later be accessed for correlating health challenges later in life—same data, two different questions. Similarly, unstructured, text-based civil affairs reports designed to provide village atmospherics for company commanders could later be ingested to correlate the incidents of crop failures on insurgent activity. AI/ML algorithms designed to explore the original use of the data, especially through a dashboard interface, would never be able to reveal the value of the data to other mission sets. This requires SOF to first ask the right questions and then have the data science teams respond by compiling potential data sources through an agile big data architecture.

Summary

All AI/ML solutions are mathematical models and can only represent a small portion of reality. Computational error is a natural and inevitable reality with AI/ML models due to a variety of factors, including hidden assumptions in the model's design, the limitations on available data, insufficient or biased training data, and bias built into a model's design. While a model often appears to objectively reflect reality, choices and trade-offs had to be made behind the scenes in the creation of the model, and those choices ultimately emphasize certain factors while overlooking others. There are rational reasons for doing so, and this is not to assert that AI/ML solutions lack value. It is simply to note that AI/ML models are question and challenge oriented and can be very useful for addressing those specific items but are also constrained by them.

It is also crucial to note that AI/ML models cannot truly predict the future, especially in open systems. Effective inductive modeling often gives the perception of being predictive, but that is only because the data to which they have access is already structured and correlated in a way that allows for near instantaneous ML processing. The SOF enterprise has access to some structured data and makes excellent use of it. However, it has even more unstructured data that requires extensive pre-processing before it can be effective when incorporated in AI/ML solutions. SOF leaders charged with developing a big data capability should consequently recognize the importance of focusing on the architecture and culture that make big data analytics work—a long-term structural endeavor—rather than focus on dashboard interfaces that seem to meet current operational needs. The next chapter discusses the basic elements of big data culture that will enable the SOF enterprise to maximize its data assets, especially once the supporting architecture is in place.

Endnotes

1. The Diplomatic, Information, Military, Economic (DIME) and Political, Military, Economic, Social, Information, Infrastructure (PMESII) analytical frameworks are commonly used intelligence analysis tools.
2. R.S. Nickerson, "Confirmation Bias: A Ubiquitous Phenomenon in Many Guises," *Review of General Psychology* 2, no. 2 (June 1998): 175–220.
3. M. Jeng, "A Selected History of Expectation Bias in Physics," *American Journal of Physics* 74, no. 7 (September 2006): 578–583.
4. E. Hüllermeier and M. Mernberger, "Inductive Bias," in *Encyclopedia of Systems Biology,* W. Dubitzky, O. Wolkenhauer, K.H. Cho, and H. Yokota, eds. (New York: Springer, 2013).
5. Hüllermeier & Mernberger, "Inductive Bias."
6. Nassim N. Taleb, *The Black Swan: The Impact of the Highly Improbable* (New York: Random House Trade Paperbacks, 2010).

Chapter 3. Making Big Data Models Work Right

Dr. Joan Peckham, Lt. Col. Andrew Geyer

Think Team, Not Tool

The foundational unit upon which the Special Operations Forces (SOF) enterprise relies is the team. Although big data is viewed as an enabling technology or a tool, the truth is that it, too, requires a team approach. The good news is that culturally, SOF are good at working in teams. The bad news is that working in teams to write code, adhere to knowledge management requirements, and manipulate databases to feed complex algorithms is generally not part of SOF culture. In reality, most organizations struggle with this very problem, so SOF are not unique in this regard.

Just as the varied SOF mission portfolios require specialized equipment, tactics, techniques, and procedures to meet unique domain requirements, so, too, does big data analytics require specialization of its practitioners. SOF are comfortable with the fact that the complex nature of special operations means that even the most outstanding SOF operators cannot be true experts at everything SOF are required to do. In every type of SOF team across all military branches, team members possess additional specialized training that allows that operator to uniquely contribute to the team. For example, in Army Special Forces teams, there are operators that are individually specialized in weapons, engineering, medicine, communications, or intelligence. While many picture Army Special Forces, Navy SEALs, Marine Raiders, or Air Force Special Tactics personnel when they think of SOF, military professionals know that SOF encompass a much wider set of specialized teams such as psychological operations, civil affairs, special warfare boat teams and special operations aviators to name a few. This variety of SOF teams is then further supported by more traditional military teams that provide logistics, intelligence, maintenance, contracting, etc.[1] Successful senior SOF leaders are able to dynamically combine teams of these teams to fit whatever the mission requires—anywhere, anytime, anyplace.[2] This same philosophy can be applied when creating a data science team to support SOF missions.

Today, problems are emerging that do not yield to traditional functional specialties or disciplinary approaches. Solutions are more easily found when teams of experts with different training and perspectives work together. Outside the military, scholars, practitioners, and teachers have worked to define new disciplines that combine expertise and knowledge from multiple existing disciplines, including experts from the domains in which the problems arise. Data science is a new discipline that emerged in this way. With the tsunami of data that is swamping all sectors, data scientists have responded by forming teams of computer scientists, engineers, ethicists, information technologists, mathematicians, and statisticians to collect, clean, organize, analyze, and present results that support the solution of data-driven problems. To make use of big data, SOF leaders consequently first need to recruit and train SOF-capable data scientists or cultivate data science-capable SOF—or a mix of both. Then, they must decide how to organize SOF data science teams in a manner that makes them most effective at achieving the mission goals.

Solutions are more easily found when teams of experts with different training and perspectives work together.

Fortunately, other sectors have already wrestled with this same problem of workforce development. For instance, new data science programs have recently been developed at educational institutions, and many schools are reaching out to retrain the existing workforce.[3] Industry has made data science training programs and seminars available to their workers. This is similar to the computational thinking movement for which curricula and training materials are now used—from kindergarten through college—to help all students develop computational thinking skills.[4] Students majoring in every discipline now know that they need rudimentary computing skills in order to function and compete in this modern and technology-driven world.[5] The rise of big data has propelled similar developments, but the core of data science is inherently more interdisciplinary than in many other new disciplines. Data science and big data require teams of experts to support functional problem-solving efforts.

Big data and data science, therefore, compel leaders to think more seriously about specialization, domain knowledge, and cross-functional teams. SOF teams require a mixture of skills, capabilities, and perspectives, and question-driven big data analytics similarly require diversity to be effective. While many leaders in the SOF enterprise expect big data to be a time-saving, automated analytics system, the truth is that most artificial intelligence (AI)/machine learning (ML) systems require significant human labor—at least on the front end of development. There are simple, fast AI/ML solutions, but they are likely to be structured dashboards without the flexibility to utilize big data in the way the enterprise will need to keep pace with emerging challenges. Before moving on to the next section describing how to construct a data science team, a few more points should be emphasized about big data analytics in general.

Big Data Is Not a Fire-and-Forget Tool

Statisticians have long informed researchers of ill-designed studies, sloppy data sampling and collection, and poorly configured teams that produce results that threaten lives, ruin economies, and worsen the hardships of vulnerable populations.[6] Given the nature of the SOF mission and influence, great care needs to be taken in the development of big data analytics. This compels consideration about the appropriate approaches for managing diverse perspective teams as well as guiding the problem-solving process. As discussed in chapter 2, bias is inherent to all AI/ML systems, especially ones based on deductive, hypothesis-based analysis. Sometimes analysts become so involved in trying to solve a difficult problem that they forget to consider whether the basic assumptions are correct. A robust problem-solving methodology includes a more complete depiction of the process that includes this first problem-definition step. The process must include early exploration of the problem space, the hypothesis formulation, the communication of the results after problem resolution, and the iterative aspects of the process. Diversity of background, domain knowledge, experience, and perspective help to mitigate issues with bias and highlight incongruities in results.

Stated more explicitly, there are good and bad approaches to big data analytics, and organizations that fail to invest in teams to develop models will experience problems and pain once the limitations of the models emerge. SOF leaders need to become comfortable with the idea that data science techniques cannot simply be developed and allowed to process data in an automated way. Instead, big data analytics requires continuous evaluation, updating, and improved data sources. Expert training and cross-functional teams are needed to explore multiple potential alternatives. Some scientists use an array of analysis tools, each of which has strengths and weaknesses, and then, informed by the existing domain knowledge, converge on an interpretation of the data. That is, if the data is indicating something about a domain that is inconsistent with what the most knowledgeable experts know about the domain, there might be something about the analytical approach that is flawed. Or perhaps the team is on the frontier of a new discovery. But it takes discussion between the analysis experts and the domain experts, and perhaps further analysis, to sort this out.

Once the experiments or questions are designed and the data is collected, organized, and cleaned—that is, structured in a way the AI/ML model can process—data science experts must choose from an array of models, tools, and techniques to analyze the data. Matching the dataset to the statistical, mathematical, or computational analysis technique requires expertise in multiple disciplines. Domain or subject matter experts (SMEs) must explain to the analysis experts—the computational coders—the nature of the data and the questions they are asking. They provide needed links to the knowledge base in the domain of analysis. The analysis expert must explain to the team the nature of their techniques. To construct an AI/ML model, the analysis expert must determine a number of factors: What are the strengths and weaknesses of each technique? What are the assumptions that are needed for a correct and robust result? To which dataset types can the techniques be applied? Where possible, and it is frequently not possible, it is best to have the analytics team at the table when the experiment is designed and before the data is collected.

The problem-solving process should also acknowledge that not every problem can be solved using the gold standard of hypothesis formation before data collection and well-designed comparison groups. Much data today is collected and archived before questions are even asked. Instead of moving from hypothesis to experimental design to solution and replication, data is now coming at such a rapid speed that it might not be feasible or ethical to design comparison groups, nor can the data always be collected again for replication. Quasi-experimental, inductive analytical techniques and ML approaches can now be used on messy datasets after cleaning and organization to get results that are almost as reliable as more traditional techniques. Sometimes, if conditions permit, it is possible to perform early explorations of existing data to support and define a hypothesis, design an experiment, and collect more data using traditional statistical approaches to confirm the trends that have been uncovered.

Powerful If On Target, Costly If Not

The point here is that big data could in theory generate a force-multiplier effect for SOF, but only if the enterprise first creates the culture and conditions for using it correctly. As mentioned in chapter 2, dashboard-oriented and spaghetti-on-the-wall inductive analyses designed to ingest everything and "tell us what we need to know" are the wrong ways to think about big data in many cases. To make big data a powerful tool, the emphasis needs to be placed on creating cross-functional teams of domain experts and analysis experts. Any expectation that big data will save time and effort must first question whether the task is fundamentally about automation or about analyzing a novel circumstance or evolving operating environment. The consequences for decision makers are extraordinary, and those responsible for creating big data structures must take into account the medium- and long-term effects of their approach.

To make big data a powerful tool, the emphasis needs to be placed on creating cross-functional teams of domain experts and analysis experts.

The Cross-Functional Team as a Big Data Necessity

Years ago, many experts and managers in the private sector hoped that the application of data analysis tools to data would yield savings in time and money. There are many companies selling and installing ML dashboards to support these needs. Consumers of these products hope that they can push a button to quickly get results that inform planning, policy, and strategy. The reality today is that these approaches can make an organization more agile, knowledgeable, and competitive but require effort and planning to coordinate interdisciplinary teams and apply and understand the results of data analysis. Human resources and domain experts are as important as software and hardware and are important parts of the ecology of data.

Organizations must train personnel to properly collect and organize the data before applying high-end analysis tools; this effort is significant and often very expensive. Professionals that apply analysis tools to their data must use robust data modeling techniques for organizing and archiving data before accessing and analyzing it. This has always been the case for statistics, but the characteristics of big data compel even more caution.[7] Most data-driven communities have already become good at collecting piles of data. However, few are well organized for collecting, cleaning, archiving, and organizing data in ways that support meaningful analysis. The computing and statistical communities warn of the garbage in, garbage out phenomenon if these steps are skipped. This kind of mistake or oversight can lead to money, lives, and reputations ruined or lost.

Managers in industry and government have long recognized the importance of cross-functional teams. Scientists and engineers might not have created the ability to walk on the moon or sequence the first DNA as early as they did without effective interdisciplinary teamwork. Scholars have characterized the continuum from multidisciplinary through interdisciplinary to transdisciplinary as increasingly engaging teams possessing expertise in multiple diverse disciplines to consider and integrate approaches for solving difficult problems. No-boundary thinking is a next step in which the boundaries among disciplines are erased to bring problem solvers from different perspectives to the table at the very beginning of the process to define the problem and then move through the various stages of solution.[8] The first important step is to be sure that the right questions are asked.

Data Scientists and Analysts

How would a cross-functional team proceed to solve no-boundary problems? The first step is acknowledging that problems do not neatly fall into disciplinary silos. Given that research indicates cross-functional teams of experts usually outperform uni-perspective teams, a difficult problem might be more likely to yield to a cross-functional team.[9] A potential process framework for a no-boundary problem solution might be the following:

1. A vague (or crisp) sense of a problem emerges.
2. Invite a cross-functional team to consider the problem and better define it.
3. Adjust the composition of the team with the diverse expertise needed.
4. Attend to the psychological factors that are unique in teams with diverse perspectives.
    - Consider approaches for flattening the hierarchy to ensure that everyone feels comfortable contributing.

- Develop evidence-based techniques for listening and giving everyone a means to be heard.
- Be respectful and share and define vocabulary across disciplines.
- Acknowledge that teammates will not know everything about one another’s disciplines and that it is their responsibility to provide needed domain information.
- Train the team in empathy, listening, and communication as well as appropriate rhetorical skills.

5. Do not support parallel play unless it is clear and agreed that a distributed approach is best for the problem at hand. If that is the case, there should still be some communication at important milestones during the progress of the project.

6. Develop a unified means for communicating the problem and its solutions among the team and to the stakeholders.

Just as with any other SOF mission, the personnel assigned to the team performing the mission must have the correct set of skills required to achieve the mission objectives. Yael Garten, for instance, argues that most data scientists must effectively specialize into what she calls “decision scientists” or “modeling scientists.”[10] In this framework, decision scientists create analysis for human consumption while modeling scientists create output that serves as input for other machines. For SOF missions, one may be inclined to think that SOF only require decision scientists and not modeling scientists. While what Dr. Garten calls decision scientists would have the skill sets to conduct the analysis needed to create dashboards and other useful tools for commanders from large or incomplete datasets, this is not the only area where SOF data science teams can enhance mission effectiveness. Modeling scientists can and should play a big role in supporting SOF mission sets. Creating big data models that in turn feed other models can be incredibly useful in a number of SOF mission areas. For example, a natural language processing model can aggregate structured data from thousands of unstructured text documents which can then be analyzed by network models to create useful intelligence analysis. Just as SOF require a wide variety of operators, they also require a wide variety of data scientists.

Garten acknowledges that “full stack” data scientists who can do all aspects of data science do exist but assesses that they are rare and exceptional.[11] While Dr. Garten is speaking solely about data science for industry, this assertion is believed to hold true within the military data science community. Most military data scientists tend to specialize in one or more subsets of the total data science skill set. But there are exceptional full stack data scientists in the Department of Defense (DOD) who are true data science generalists. SOF leaders should seek out these data science generalists and recruit them for SOF data science teams. Just as SOF operators must be exceptional at their job, SOF data scientists must also be exceptional at theirs. This is due to the likely size of SOF data science teams as well as the nature of the data that SOF data science teams might encounter. SOF teams are often very small and widely distributed when compared to conventional military units. So, a single data scientist might have to run multiple different types of data science projects at any given time. This scenario is very much like the industry small startup scenario that Dr. Garten describes as ideal for the rare data science generalist.[12]

Since SOF live in the grey zone where data is often unstructured, incomplete, and of varying reliability, a SOF data science team might have to completely change their plan of action depending on the mission situation, changes in available data, and changes in timeline requirements. The SOF data science team leader might have to iteratively apply something like the U.S. Army Rangers' troop leading procedures (TLPs) throughout the project in order to answer the questions he or she is tasked to explore. Eric Colson asserts that this exact type of scenario requires data scientists who are generalists.[13] Colson argues that dividing up labor among many specialized data scientists only serves to slow down the project, make it more expensive, and limit utility of any resulting models or analysis by depriving the data scientists of the larger context of the problem they seek to solve. Data scientists who are able to see the larger picture, he argues, are more effective at obtaining useful results.

Senior leaders in military data science sometimes speak about unicorns—former operators who now possess advanced academic degrees in fields related to data science. They are a rare thing, a single person who is both a data scientist and a SME on SOF operations. Unicorns are even rarer than true data science generalists. While SOF leadership should most certainly make an effort to recruit operators with the requisite background to pursue the education required to become a data science unicorn, it is unlikely there will ever be sufficient funding and manpower to fulfill all SOF data science requirements with unicorns.

Subject Matter Experts

Data scientists cannot by themselves create big data solutions since their expertise lies with statistical analysis, computer coding, and/or data visualization. To make the models in the first place, there must be sensemaking about the issue or challenge, and that can only come from domain or SMEs who study it. Just as SOF operators are most effective as part of an integrated team, data scientists are most effective when working as part of a cross-functional team with relevant SMEs.[14] The complex and delicate nature of SOF missions means that a SOF data science team could require not only regular input from SOF experts but also input from SMEs in areas as varied as international relations, local culture, social science, anthropology, medicine, public health, history, language, physics, or engineering. These SMEs from fields relevant to the project need to be present from the beginning to assist with data collection, data hygiene, and output interpretation. Without SMEs in the loop from the beginning, a data scientist can make a seemingly useful model or conclusion that is, in fact, completely misleading or is not feasibly actionable.[15]

SMEs are typically in short supply in the military just like data scientists. Many leaders' faith in employing big data is due to the perception that the technology will compensate for the lack of domain knowledge—the "gonculator" will give them what they need. While the ideal situation would be for SMEs to be organic to units, which is possible in some cases, as a rule it will be difficult to make happen. Rather than thinking about the process as a matter of possession—the SME is resident in the office—perhaps it is more fruitful to imagine how SMEs can be invited in for a particular purpose, such as issue discovery and framing, and then consulted periodically for analytical purposes. The takeaway here is that big data cannot eliminate the need for SMEs; rather, it might just create more options on when and how they are integrated into the system.

Users

At the end of the day, the data scientists and SMEs are there to serve the needs of the user—the action officer whose job it is to translate the analysis into effect. The user is not expected to have domain knowledge or data science skills, but he or she knows why the information is being requested in the first place. The user's perspective is ultimately the one that must be accommodated, but to do so correctly, the data scientists and SMEs must work with the user to correctly frame the questions and determine what the available data can actually address. Moreover, the data scientists must generate analytical and visual products that enable the user to communicate vast amounts of information intuitively.[16] Without the user's perspective, the data scientists and SMEs are likely to create elegant but ultimately misaligned results. Users with general awareness of the principles underlying big data are all that is required. Armed with the concepts in chapters 1 and 2, a user could very easily interact with the rest of the cross-functional team.

Without the user's perspective, the data scientists and SMEs are likely to create elegant but ultimately misaligned results.

Which Comes First: The Subject Matter Expert or the Capability?

In the U.S. military, the natural impulse is to buy the technology and hire the staff to make it work. Sustainable big data solutions, however, require a different mindset. With technology evolving so rapidly, expenditures on hardware and software could become potential liabilities if the personnel cannot use them due to a lack of skills. The allure of dashboards is precisely because they appear to have done all the hard data science and SME work up front. But to generate an appropriate capability for SOF, the enterprise needs adaptable, not static, systems. Lessons from industry suggest that investing first in the personnel and human resources aspect of big data will yield greater dividends than purchasing hardware and software.

Considerations for Developing a Capability

It should be evident by now that making good decisions with big data is not about purchasing a tool and pushing a button. A ML, statistical, or mathematical technique is a tool and not an oracle. Experts with diverse training are needed to know when to use a tool, how to use a tool, and which tools to use for a particular problem and dataset. There are many important steps that proceed and follow the use of such tools, and different types of expertise are needed at each step. As with all things SOF, people are more important than hardware, so investing in the human resources of data science concepts and big data cross-functional team behavior constitute baseline undertakings for an enterprise solution. At a minimum, human resources approaches to developing a big data capability should:

1. Invest in rudimentary training in the data science core—math, science, computing, analytics, and ethics.
2. Create a process for cross-functional team development that:
   a. Picks the right tools for the problem and dataset. This means understanding an array of tools and having access to experts in multiple domains who are knowledgeable about these tools.
   b. Recognizes when teams need deep expertise in a domain that they do not possess.
   c. Recognizes when the proper analysis tool for the problem does not exist and knows whom to call to tweak existing tools or develop new ones. There is a difference between learning how to use existing tools and knowing how and when to build new ones.
   d. Considers information technology tools to support communication among teammates.
3. Include multiple perspectives at the problem formulation stage. Deep integration of design is important.[17]

To further ensure the effectiveness and resourcing of SOF big data teams, they should likely be embedded in operational and strategic headquarters with sufficient infrastructure to support high-performance computing. Team leaders need to have access to senior leaders to ensure the team is properly resourced and working tasks that matter to the overall SOF objectives for the headquarters. The personnel on the team should be uniformed or DOD civilian workers, as near-peer adversaries have shown a tendency to infiltrate or exploit technology companies for their own objectives.[18]

Big Data Pitfalls: Expectation Management in the Transition to a Big Data Capability

A properly constituted big data cross-functional team would minimize the risk of falling into one or more of these three pitfalls of data science: data hubris, hidden bias, and failing to capture uncertainty. Data hubris is assuming that traditional data collection and analysis can be avoided when using a massive dataset that can be fed into a data science modeling tool. One example of this is Google Flu Trends (GFT). The GFT prediction tool was designed to predict doctor visits for influenza-like illness using 50 million search terms to fit 1,152 data points. The GFT failed as a stand-alone tool but became more useful when combined with more traditional science-based techniques.[19] In this case, data hubris was a problem because the Google data scientists had extensive experience working exclusively with large and complete datasets. Data scientists who have worked exclusively in internet-based companies may be prone to this type of data hubris because they are used to having all of the data they need collected from every possible customer. The dataset used for the GFT turned out to be big but incomplete. As previously discussed, SOF live entirely in a world where the data may be incomplete, unavailable, or unreliable. For data scientists used to working on large, complete datasets, data hubris may be a problem. Conversely, data hubris may also be a problem for personnel with minimal training in data science who are acting as a data scientist. Data science tools are user friendly and widely accessible. It is all too easy to use these tools without possessing the ability to discern if the resulting output has merit.[20] There is nothing more dangerous than a half-trained soldier with a powerful weapon.

The second major pitfall that a team might face is hidden bias. Hidden bias can enter in three possible stages. The first stage is when the team decides what question(s) they really want to answer. How the problem is framed can "bake in" bias. Karen Hao uses the example of a bank trying to find the most creditworthy clients morphing into a model that targets customers for predatory, subprime loans due to the mathematical model being formulated solely to maximize profit.[21] The bank intended to ask, Who are our best customers? but ended up accidentally asking, Who are the best targets for predatory loans? The second stage where bias can be introduced is when the data is collected. A noted example is from facial recognition datasets. Too many companies working in this area have relied too heavily on images of people with lighter skin tone, thereby biasing their facial recognition results so that people with darker skin are not recognized accurately.[22] The last stage where bias may be introduced is when the data is scrubbed for analysis. If the data science team fails to include the appropriate SMEs and statistical techniques, it is possible that the wrong set of variables will be selected for inclusion in the resulting data science model, thereby creating a model with poor predictive or analytical utility.[23]

The third major pitfall for a SOF data science team to avoid is the failure to capture uncertainty in their results. Traditional statistical regression models provide rigorous methodologies for quantifying the impacts that input variable values have on the resulting output variable values.[24] When the problem calls for more complex techniques such as deep learning, the contributions of the input variables to the output variable values are much harder to discern. This is a scenario where experienced data scientists are needed to work alongside relevant SMEs to ensure that the uncertainty in any conclusions is properly quantified or, at the very least, commented upon.[25]

Leading the Big Data Team

Most SOF ground operators use some variant of the eight-step TLPs found in the Army's *Ranger Handbook* when conducting tactical mission planning. While TLP is designed for infantry squad or platoon tactics, it can be applied to any small team, or team of small teams, tasked with achieving an objective. The utility of TLP and SOF operators' familiarity with it make TLP a convenient methodology to use for applying data science techniques to big data for SOF mission planning and execution. The rest of this section will walk through the eight steps of TLP and explain how they can be applied to data science and big data.[26]

Step 1. Receive the Mission

This is the step in TLP when the unit leader receives an operations order (OPORD) or fragmentary order that specifies the mission. The leader then conducts a hasty analysis centered on preparation and planning. The leader of a data science team conducts the same steps when presented with a new problem to tackle. The data scientist's hasty analysis involves first understanding exactly what question(s) is being addressed. Is the analysis required descriptive, diagnostic, or predictive? Descriptive analysis extracts insight about what has happened in the past. Diagnostic analysis looks to determine why some known event has happened or is happening. Predictive analysis uses data to gain actionable insights about what is most likely to happen in the future.[27] Each type of analysis requires a different approach using similar skills, but understanding which type of analysis is required up front is critical to mission success. There is no benefit to answering the wrong question.

Step 2. Issue a Warning Order

In this step, the leader issues preliminary instructions in a warning order (WARNORD) so that the team can begin preparations. In an infantry squad or platoon, this is accomplished using the five-paragraph OPORD format. Paragraph one provides information to subordinates so that they can understand the current situation. A data science team leader does the same when first relaying the description of the analysis required to the rest of the data science team. The second WARNORD paragraph explains the unit's mission. This is where the data science team leader coordinates with a diverse team the metrics that will be used to assess the team's effectiveness at addressing the analysis problem. The third paragraph discusses mission execution to include the desired end state. In data science terms, this is where the team leader explains why the analysis is needed and what minimum values of the performance metrics are required for the resulting analytic model to be declared useful or appropriate for the task. Paragraph four lays out logistical requirements. This is the same for a data science project. The team leader needs to describe what resources in computing, data storage, and personnel are available to address the task(s) at hand. Finally, the fifth paragraph in an infantry WARNORD states where command and control elements are located during the planned operation. For a data science team, this is where the team leader spells out the team organization, who is the point of contact for specific parts of the project, and any external resources that may be available for assistance in the event that the team meets a stumbling block.

Step 3. Make a Tentative Plan

This is the step where the infantry leader makes his own assessment of the situation and begins making his plan. A data science team lead does the same by scoping the project. Like his infantry counterpart, the data science team leader first examines mission, intent, and concept for the question(s) he or she is tasked to answer. Next, the team leader examines unit tasks. Just as it is for the infantry, the data science team has to examine both specified and implied tasks. For example, Connaboy et al. had the sole specified task to create a model that predicts lower-extremity injury in U.S. Special Forces.[28] That led to a number of implied tasks, some of which are listed in table 1.

Table 1. Example Specified and Implied Tasks. Source: *Medicine & Science in Sports & Exercise*

Specified Tasks	Implied Tasks
Create model that predicts lower-extremity injury in U.S. special forces	Determine set of measurable variables related to lower-extremity injury
	Determine method for quantifying lower-extremity injury rates
	Obtain appropriate volunteer subjects for injury study

The next part of the planning process involves specifying unit constraints. This is where an infantry leader spells out the prohibitive and proscriptive constraints on the team. A data science team leader does the same. Some examples of these constraints for data science teams are listed in table 2.

Table 2. Unit Constraints for Data Science Teams. Source: Author, *Forbes*, *Wake Forest Law Review*

Prohibitive Constraints	Proscriptive Constraints
Laws and ethics governing data collection	Document all assumptions made in analysis
Access to relevant data	Review published literature for current techniques
Computer hardware limitations	Practice good data hygiene

The data science team leader then identifies the mission essential tasks based on the constraints and the specified and implied tasks. These tasks are specific to each data science problem the team faces. For the data science team leader, this TLP step culminates in restating the question(s) under investigation in terms of mathematical models with quantifiable input and output variables. If applicable, this is where the data science team also discusses limitations of the techniques or mathematical models they intend to use including model assumptions and uncertainty in the model results.

Steps 4–6. Initiate Movement, Conduct Reconnaissance, and Complete the Plan

In these steps, infantry teams make physical movements to scout the situation and adjust the tentative plan as needed. These steps can occur simultaneously with the previous three steps of the TLP. The leader may rely on other team members to provide information required to adjust the plan to the current situation. Data scientists and engineers recognize this as a parallelized planning process. Like an infantry squad or platoon, a data science team also conducts reconnaissance and adjusts the plan as the situation evolves or new perspectives are added to the team. However, the data science team's reconnaissance is of the computing resources, mathematical techniques, and data sources available for the project versus the terrain and personnel in the objective battlespace. Most data science team leaders delegate tasks whenever possible to members of the data science team with the applicable skill or subject matter expertise just as a SOF team leader delegates appropriate mission tasks to the various specialized operators on the team.

Step 7. Issue the Operations Order

In an infantry squad or platoon, this is where the leader ensures every member of the team understands the plan and his or her role within it. This may involve quizzing each team member until every member knows their part by heart. While it would be unusual for a data science team leader to quiz individual team members on their role in a project, data science team leaders do enact a version of this step of the TLP. In the data science version of this step, the team usually gathers around a large screen or dry-erase board to walk through the plan for addressing the data science problem and everyone's roles in the project. This usually also entails determining timelines, potential bottlenecks, and constraints on the project. Some data scientists and engineers refer to this step as "whiteboarding" because of the common use of dry-erase marker boards. It is very much the data science equivalent of the Army's rock drills prior to an operation.

Step 8. Supervise and Refine

This final step is where an infantry squad or platoon leader conducts rehearsals and inspects the team's equipment to ensure they are able to successfully conduct the planned operation. While inspections and rehearsals are not necessary for tackling a data science problem, data scrubbing and screening experiments are almost always required, especially when dealing with big data. Data scrubbing is the act of practicing good data hygiene. It is the data scientist's version of an equipment and personnel inspection. Good data hygiene requires the steps in table 3.

Table 3. Data Hygiene Steps. Source: *Forbes*

1. Audit the data for accuracy and completeness
2. Standardize data formats across all sources
3. Determine guidelines for which data to include and exclude
4. Automate rules for removing future "bad" data
5. Ensure data remains up-to-date
6. Eliminate silos within the organization (cross-feed information)

Steps three and four of the data hygiene process are usually conducted via one or more screening experiments. Screening experiments can be thought of as analogous to the infantry platoon rehearsal prior to an operation. Screening experiments are smaller data science projects that are deliberately designed with the objective of determining which variables in the dataset should be included in the larger data science project.[29] While these screening experiments are conducted to narrow down the data to just those variables that matter for the data science project at hand, they also allow the data science team leader to assess the strengths and weaknesses of the data science team members. This kind of preliminary work can sometimes serve as a way to sharpen team skills before beginning the larger project depending on how large the big data may be. In addition to screening experiments, steps three and four may also require consulting data SMEs in the field(s) from which the data originates. These SMEs can assist with determining which input variables are more likely to provide insight into the output variable(s). SMEs can also provide expertise regarding the validity of any data science models created. SMEs in combination with data scientists are also most effective at ensuring the data science project avoids "baking in" any biases into the resulting model that maybe make the data science model look valid when, in fact, it is missing critical data required to make accurate predictions, inferences, or decisions.

SMEs are also very useful at assisting data scientists with steps five and six of the data hygiene process. SMEs provide insight into what the data means and how the variables in the data are expected to relate with each other. They know relevant data sources, where current research is headed in their respective fields, and how often data should be refreshed. While data scientists provide expert insight on how to formulate variables from raw data in a manner that makes them interpretable by the applicable mathematical or statistical models, SMEs play a critical role in helping the data scientists determine which data at which resolution to include in data science projects relative in their field.

Conclusion

Contrary to popular conception, big data is a personnel-intensive analytical process, not a labor-saving efficiency tool. The real benefit of big data is the ability to make evidence-based decisions. While it takes more work, the benefit is not speed or lower cost, but better evidence—if it is done well. The purpose of Part I of this monograph has been to provide the basic concepts on how to conduct proper big data analytics.

When approached properly, big data analytics require data scientists, domain SMEs, and users to work together in collaborative teams to deal with often novel datasets to answer novel questions. Leaders who assume big data can save labor effort are likely imagining automated functions or dashboard interfaces that are relatively static. When thinking about big data as an enterprise solution, however, it is crucial to recognize that the cross-functional team lies at the heart of the undertaking and getting the human resources component right from the beginning is the first critical step in making big data capability the force-multiplier it could possibly be.

Endnotes

1. Joint Chiefs of Staff, *Special Operations*, 3-05 (Washington, D.C.: Joint Chiefs of Staff, 2014), https://fas.org/irp/doddir/dod/jp3-05.pdf.
2. S. A. McChrystal, T. Collins, D. Silverman, and C. Fussell, *Team of Teams: New Rules of Engagement for a Complex World* (New York: Penguin, 2015).
3. *Data Science for Undergraduates: Opportunities and Options* (Washington, D.C.: The National Academies Press, 2018), http://nap.edu/25104.
4. Jeanette Wing, "Computational Thinking," *CACM* 49, no. 3 (March 2006): 33–35.
5. Shriram Krishnamurthi and Kathi Fisler, "Data-Centricity: A Challenge and Opportunity for Computing Education," *Communications of the ACM* 63, no. 8 (August 2020): 24–26, https://cacm.acm.org/magazines/2020/8/246361-data-centricity/fulltext.
6. Cathy O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (New York: Crown Publishing Group, 2016).

7. O’Neil, *Weapons of Math Destruction*.
8. Xiuzhen Huang et al., “No-Boundary Thinking in Bioinformatics Research,” *BioData Mining* 6, no. 1 (November 2013): 19.
9. Scott E. Page and Lu Hong, “Groups of Diverse Problem Solvers Can Outperform Groups of High-Ability Problem Solvers,” *Proceedings of the National Academy of Sciences* 101, no. 46 (2004): 16385–16389.
10. Yael Garten, “The Kinds of Data Scientist,” *Harvard Business Review*, 6 November 2018, https://hbr.org/2018/11/the-kinds-of-data-scientist.
11. Garten, “The Kinds of Data Scientist.”
12. Garten, “The Kinds of Data Scientist.”
13. Eric Colson, “Why Data Science Teams Need Generalists, Not Specialists,” *Harvard Business Review*, 8 March 2019, https://hbr.org/2019/03/why-data-science-teams-need-generalists-not-specialists.
14. Mario Konschake, “Embedding Data Science in Cross-Functional Teams,” *Medium*, 21 December 2018, https://medium.com/@Infinite_Monkey/embedding-data-science-in-cross-functional-teams-7bfce9283ad2.
15. M. Loukides, “The Unreasonable Necessity of Subject Experts,” *O’Reilly*, 20 March 2012, http://radar.oreilly.com/2012/03/subject-matter-experts-data-stories-analysis.html.
16. Mark Ballora, “Opening Your Ears to Data,” TEDxPSU, 5 December 2011, YouTube, video, 13:45, https://www.youtube.com/watch?v=aQJfQXGbWQ4.
17. Notes on effective interdisciplinary teams are taken from a talk by Caroline Guttschalk Druschke to a graduate class at the University of Rhode Island. Professor Druschke is currently an assistant professor of English, focused on rhetorical theory and freshwater science and management at the University of Wisconsin–Madison.
18. Z. Budryk, “Top U.S. General to Meet With Google on China Security Worries,” *The Hill*, 21 March 2019, https://thehill.com/policy/defense/435160-top-us-general-to-meet-with-google-on-china-security-worries.
19. David Lazer, Ryan Kennedy, Gary King, and Alessandro Vespignani, “The Parable of Google Flu: Traps in Big Data Analysis,” *Science* 343, no. 6176 (2014): 1203–1205, https://science.sciencemag.org/content/343/6176/1203.full.
20. Maurice Ewing, “Is Data Science Too Easy?” *Forbes*, 6 February 2017, https://www.forbes.com/sites/quora/2017/02/06/is-data-science-too-easy/#5347ddef292a.
21. Karen Hao, “This Is How AI Bias Really Happens—And Why It’s So Hard to Fix,” *MIT Technology Review*, 4 February 2019, https://www.technologyreview.com/s/612876/this-is-how-ai-bias-really-happensand-why-its-so-hard-to-fix/.
22. Stephanie Ruhle, “About Face: One Woman’s Quest to Make AI Less Biased,” *NBC-News*, 1 February 2019, https://www.nbcnews.com/better/video/about-face-one-woman-s-quest-to-make-ai-less-biased-1436205635783.
23. Hao, “This Is How AI Bias Really Happens.”

24. Max D. Morris, *Design of Experiments: An Introduction Based On Linear Models* (Boca Raton: CRC Press, 2011).

25. Maria Temming, "Why a Data Scientist Warns Against Always Trusting AI's Scientific Discoveries," *Science*, 20 February 2019, https://www.sciencenews.org/article/data-scientist-warns-against-trusting-ai-scientific-discoveries.

26. Headquarters, Department of the Army, *Ranger Handbook* (Washington, D.C.: Headquarters, Department of the Army, 2017), https://armypubs.army.mil/epubs/DR_pubs/DR_a/pdf/web/ARN3039_TC%203-21x76%20FINAL%20WEB.pdf.

27. Anjali Uj, "Differentiating Descriptive, Diagnostic and Predictive Analytics," *AnalyticsInsight*, 3 June 2018, https://www.analyticsinsight.net/differentiating-descriptive-diagnostic-and-predictive-analytics/.

28. Chris Connaboy et al., "Employing Machine Learning to Predict Lower Extremity Injury in U.S. Special Forces," *Medicine & Science in Sports & Exercise* 15, no. 5 (May 2019): 1–31, https://journals.lww.com/acsm-msse/Fulltext/2019/05000/Using_Machine_Learning_to_Predict_Lower_Extremity.28.aspx.

29. Morris, *Design of Experiments*.

Part II: Management Issues with a Big Data Capability

Chapter 4. The Management Culture for Big Data

Mr. Gaurav Tanwar

As the previous chapters indicate, a big data management paradigm must be instituted if the force is to evolve with and capitalize on disruptive technological developments and make the most of its personnel, not just adapt to the times. This calls for something beyond strategy; it requires a cultural update because "culture contains strategy."[1] (This phrase is sometimes colloquially altered to "culture eats strategy for breakfast.") The management culture for big data requires moving beyond thinking of technology as an add-on enabler; rather, it contains a set of values, attitudes, and behaviors that often conflict with military culture. This chapter provides an overview of the management culture traits that best align with big data use and contrasts it with military culture and challenges. It begins by discussing the values and attitudes necessary for a productive big data culture, highlights some challenges military leaders are likely to face, and concludes with some models that have been tried in the government for leading big data culture change.

Big Data Management Values and Attitudes

Of the many values exemplified by the force, the three values most applicable to the formation of proper big data culture are agility,[2] adaptability,[3] and technical humility. The first and foundational big data culture value is agility, which is distinguished from mere speed. Anyone can move fast, but speed paired with control creates agility that serves a purpose. This mentality is exemplified by agile methodology, the current industry standard in big data project management, which emphasizes rapid testing to find what works, minimizes paperwork, and fails fast but early.[4] While parts of this methodology are in natural tension with government requirements, such as minimizing paperwork because the government must document its activities, there are ways it can be incorporated into a military big data culture.

The second value, adaptability (often considered flexibility), is defined in this context as comfort in working with the unfamiliar, finding asymmetric payoffs, and exploiting opportunities for rapid assimilation.[5] Russian actions are a prime example of this value. Those in the counter-disinformation/misinformation space have seen Russian trolls change tactics and approaches, adapting when they find themselves flagged or countered by tech platforms on which they operate, or boosting a message if the target audience is found to be receptive. This ability to pivot and press an advantage is extremely valuable, showcasing adaptive fluidity in the face of obstacles. This approach has yielded results for relatively little cost from Ukraine to the United States.[6]

The third value, technical humility, is a caution reminding managers that technical solutions should not displace human decision makers nor is faster always better.[7] Human judgement should be informed by technical tools and amplified by the systems at their command. To forget this precept is to abdicate responsibility to algorithms, an unacceptable reality when it comes to the particular responsibilities the interagency bears. Managers must remember that big data is based on models, and while models have utility, they do not perfectly reflect reality, and overreliance on them leads to a blinding of human understanding. A conceptual model's primary objective is to convey the fundamental principles and basic functionality of the system it represents in an easily understood manner. Models can thus support frameworks in which the user can think and make decisions by refuting misconceptions, better highlighting constrained actions, and more precisely targeting objectives.[8] Moreover, managers must remember that context matters in the application of big data; successful application in one instance does not necessitate successful or appropriate application in another instance.[9] [10] It is very easy to simply follow a technical system's advice thinking that it knows better, mistake a model's output for truth, or apply a great new method to every problem regardless of the context surrounding that problem. Ingraining the above values and cautions can help organizations grow and maximize big data culture, navigating the bright ideas and new fads that inevitably appear. In developing an enterprise capability, however, managers should also be mindful of the attitudes most suited for a big data architecture for the long haul. These attitudes are patience, willingness to fail properly, awareness of the layered and distributed Special Operations Forces (SOF) user base, and sensitivity to the governance practices in big data.

Patience

One of the largest hurdles in developing an enterprise-oriented big data capability is a common problem in many large organizations: teams are overburdened with responsibilities which leads to a lack of patience. As an example, SOF are not positioned to capture the gains from big data advances because they have been fighting for over 20 years in two extended combat campaigns and engaged in multiple other operational theaters of various kinds worldwide. According to a review by the U.S. Special Operations Command (USSOCOM) commander, this has affected the force's preference for combat-focused capabilities over other investment opportunities.[11] The growth of USSOCOM in the past two decades has largely been in response to all the things it has been forced to take on instead of deliberate growth to achieve targeted objectives.

Whether it is the need to meet operational requirements, fiscal years, or quarterly reports, patience remains in short supply. After all, patience in a world of rapid change, competitive pressure, and an unrelenting pace is a difficult thing. It is understood that big data culture cannot be quickly implemented in an existing large organization and requires careful consideration. However, this lies in natural tension with the pace at which most organizations find themselves operating. Unfortunately, "win now" culture leads to dashboard-type solutions, not enterprise architecture development. There is an old maxim referring to the iron triangle of Good/Fast/Cheap.[12] In short, pick two of the three, if lucky. In this instance, what is being created has to be good, and it is not going to be cheap, so fast is off the table unless managers settle for dashboards and all the limits inherent to them.

> *After all, patience in a world of rapid change, competitive pressure, and an unrelenting pace is a difficult thing.*

To take one example from the past decade, there has been a continued urgency related to open-source intelligence (OSINT) collection and processing, but there is yet to be a true enterprise solution. Part of the issue is certainly attributable to the various authorities and titles inherent to organizations conducting this mission. However, part of this is due to the fact that organizations want to get something now with the idea that anything they can do to meet the mission's needs will be better than doing nothing. Once the tool is obtained, there is then little pressure to create an enterprise solution as other priorities take over and people make do with what they have. In the publicly available information OSINT world, that could be everything from not pricing in maintenance and upgrades to the lack of hiring the right personnel to failing to account for the steps necessary to harmonize the type of data being collected with the product secured. While this is a simplified account of a common Gordian knot, the observable end result is a range of single-purpose dashboards across the military and interagency, often without the ability to speak to each other, but no real enterprise solution.

The aim here has to go beyond building the tools and instead to the cultural foundation, which takes time to instill. The manager's thinking must be shaped by the considerations that create time constraints, key events that should shape the timeline: How long will it take to execute correctly? How long will it take to inculcate the right skills into the force? Will that go beyond my tenure here? How vulnerable does this shift make the force? And for how long? To whom? Once objectives are benchmarked against external and internal factors, the framework is set for patience.

Failing Properly

The second hurdle is the acceptance of failing properly, a concept derived from the hard sciences wherein even negative results or failed experiments are supposed to be published,[13] so that others in the field may learn from the experiment's failure. Leaders who wish to inculcate the required big data culture must become comfortable with the promotion of failing properly. This goes beyond accepting failure. It is about shepherding unorthodox thinkers who take risks up the ranks,[14] encouraging fast failure in conjunction with the continuation of successful programs.[15] Every organization will have to, at some point, take large risks, and if they have not held on to the personnel who have taken large risks before, the organization will not have the diversity of expertise needed to see them through the riskiest of times. They will have people who fit within the existing organizational culture, not those who are best suited to adapting to and addressing the problem at hand.[16] This is an uncomfortable position for many institutions, as success in all things is now an expected cultural norm. But if institutions are unwilling to judiciously fail, they are not only unwilling to reach as ambitiously as needed, but they will be unprepared for when they will be forced to act creatively in the face of unorthodox challenges. It is not enough to say fail fast; there must also be commendation and promotion of those who fail properly.[17]

Failing properly in the big data paradigm does not have to be a mysterious process. Existing practice with this process already exists within the hard sciences. It means documenting failure in detail, coalescing lessons learned, and sharing the accumulated knowledge widely in excruciating detail for review by peers. This is done so that others may learn to not pursue those same avenues, shed light on what happened, and then do better. This way is no different than how the scientific method is employed.[18] This acknowledgement of failure is not easy, it is not comfortable, and it takes a great deal of moral courage, but it is absolutely necessary if organizations, large and small, are to move forward with a big data capability. Leaders must not only acknowledge this but put themselves on display. In the same way that others follow when leaders are the first through the breach, the first to the fight, leaders must be willing to say "follow me" when acknowledging failures. A properly documented failure is as important as, if not more than, an improperly documented success.

A Distributed, Layered User Base

Many organizations are stuck in the processing-exploitation balance in which they take in a tremendous amount from the oceans of data being produced every day,[19] yet they are only able to exploit a fraction of it. The challenge of full exploitation and analysis has yet to be completely overcome partly due to the sheer scale of data creation and the fact that leaders have a plethora of questions that need answers, resulting in a prioritization issue. Rather, with big data, the objective ought to be how to enable organizational culture to pivot when and as needed to answer new and novel questions.[20] Enabling this pivot requires a distributed, layered user base with access to the appropriate data at the appropriate level of classification. Together, this provides the foundation that a proper big data culture can use to answer the questions sought in the related but multifaceted pursuit(s) of their mission. This does not answer the question of what knowledge the organization seeks to generate; rather, it answers the more fundamental question of how the organization creates a shareable, universally exploitable, enterprise-wide knowledge base.

Effective big data requires continuity of data entry, not just the old garbage in-garbage out conversation, but also refreshing the data so as to not let it rot due to the velocity aspect of the five V's.[21] Achieving enterprise-wide standards of entry despite the multiple headquarter (HQ), component, and unit cultures of SOF would result in the fluid creation of new knowledge for the force to answer existing questions or for querying in the face of future unpredictable, emergent challenges. This is where artificial intelligence (AI) ultimately enters into the organization's observe–orient–decide–act loop.[22] The combination of data volume, velocity, variety, veracity, and value[23] form a clean river of data from which the organization can feed machine learning models, which can then grow algorithms quickly and efficiently. Without the proper big data culture, potential will be squandered. Without an enterprise approach, reach will be stymied. Without the best possible data stream, results risk being tainted, outdated, biased, and/or without value.[24] Utilizing this data requires full organizational buy-in to create a unified system, training pathway, and quality control process despite the distributed and layered SOF user base.[25]

Governance

Even with the first three attitudes in place, there are some specific hurdles to overcome related to issues of governance and classification. It is impossible to discuss one of these issues without addressing the other as they are so interwoven as to be tangled into each other. Separating them is like untangling barbed wire: difficult, full of sharp edges, and hard to know where to start. The first facet is governance, or who gets access to what. Maintaining control over data access and results is not just a good idea, it is codified in law, policy, and authorities.[26] For example, the health research field bears particular examination due to the legal parallels. For health research, it is imperative to have clear, accurate data upon which to model outcomes, yet for very good reasons, that data must be protected to preserve the privacy of individuals despite resulting analytical faults.[27] Striking a proper balance in a world where organizations can mine tons of personal data is very difficult, especially when dealing with small populations (subsets with certain traits, diseases, or other issues). The health research field works on this issue by creating abstractions, having those who work with the data sign waivers, and by having people voluntarily share information.[28] In contrast, Silicon Valley does not take these steps, which allows it to make enormous gains, exceptional products, and tailor make solutions but often also violate privacy.[29]

The second facet, classification according to security clearance levels, clearly does feed into the governance realm. It not only determines who can and cannot see certain types of data and analytical products but also requires an important analysis independent from the larger permissions discussion. Further, classification issues constantly crop up in conversations pertaining to managing access and entry—who has access to what data for modeling purposes? As a rule, it is generally easier to move things up the classification chain rather than down.[30] As such, it is possible to use the same sets of data as things go up the classification chain, though additional data garnered from classified sources and finished results may be difficult to work down. While there is work in this area,[31] no solution has yet been found to overcome the limitation of seamless access across systems, though that may change. Overcoming this hurdle, then, is a matter of creating the correct data-flow structure.

This messy combination of obstacles related to governance forms a single, overarching hurdle to the management of a distributed, layered, enterprise architecture, which directly affects how data is processed in a standardized format for use by the force and eventually AI algorithms. The way to overcome these hurdles, then, is to structure the data in a way to obviate the entire issue. As always, form follows function. Some practitioners would fall back on the data lake vs. data warehouse dichotomy. In the field of big data, there is a rigorous debate as to how big data should be organized, with the data lake and data warehouse options being the most often cited paradigms.[32] In a data lake, all of the unstructured data is kept together, there is ample access by those who need it, and there is no loss of data no matter how many elements draw from the lake. In the data warehouse model, the data is structured in a way that is easily accessible and can be delivered to those who need it.[33] While each of these paradigms is useful within specific organizations, because of the hurdles of governance and classification, there are problems with each of these paradigms. The data lake cannot overcome the issues of the unique governance constraints imposed by special access classification issues while the data warehouse comes already structured, making it difficult to combine at a higher level with differently structured data or with information gathered through separate authorities. Like many other central repositories of information, both systems also present a tempting cyber-attack, sabotage, or espionage target.[34]

Current Deficiencies in Military Communication and Technology Culture

The previously mentioned steps have all been about building the required cultural foundation and technical pillars to support an organization's leading-edge elements. All of those steps, from cultivating the right big data culture to creating the correct enterprise solution, takes a tremendous amount of time, effort, and energy. As no organization has all three values simply laying around, there comes a perceived tension in allocating very precious resources away from current responsibilities. As mentioned above, the current operations orientation creates the perception of little room to focus on a futures orientation. So how do managers create a proper balance as they develop an enterprise solution?

Current Operations versus Futures Orientation

Part of the answer is realizing the false dichotomy of having to trade off between current operations and future operations. This is not the case. Like the other systems which support joint all-domain command and control operations, the proper execution strategy is to work the big data usage into current operations, including iterative deployment coupled with rigorous, end-to-end testing and evaluation cycles. Implementing practices today and persistently employing them day by day on the big data ready applications reinforces the culture the organization must cultivate in its personnel. With experience and diffusion, the culture will begin to transform with the correct values and attitudes as a foundation. That is how individuals and organizations build the habits which see them through anticipated hard times.[35]

Part of the answer is realizing the false dichotomy of having to trade off between current operations and future operations.

The other part of the answer is realizing the biggest tradeoff comes in the habits, attitudes, and actions of senior leadership.[36] This has been discussed in oblique terms, from the need to grow the right culture to leaders demonstrating their own failures to support for unorthodox thinkers. However, there is a finer point to be made that a leader's team responds to his actions not words.[37] Sometimes that is difficult for senior leaders to remember or understand—that they are acting in ways that do not jive with their words. That dissonance is keenly felt by their teams, and those teams act accordingly.

As leaders seek to build the right big data management culture, they at times forget to check their own role in the process.[38] The onus upon senior leaders is greater because of their position and because they are trying to drive a change. Dissonance is detected by teams far, far faster than senior leaders realize, even if they do not realize there may have been any.[39] Examples of this dissonance range from claiming the aim is to institute a big data institutional culture but simply overruling data-driven results based on experience; not incorporating the output alongside (or even above) other analysis; not giving it the same time, attention, resources, or status as others; or not providing a pathway for those in the field to succeed and grow. By contrast, when senior leaders act consistently by asking questions, caring about the methodology, and requesting more nuanced or detailed information, teams respond by working enterprise tools into current operations to find those answers, leveraging their culture to find innovative ways forward. Over time, this everyday response propels the team forward to the desired end state of "future operations."[40]

Legacy Platforms and J-Code Structures

Structurally, there are also significant hurdles. How can a cross-functional team/big data culture evolve in the context of the current SOF Joint Staff culture while simultaneously breaking free from the legacy platforms/technologies that place data as an enabling function and not the center of operations? Big data requires cross-functional teams but the J-code creates silos of excellence. The mechanics of a fast break from the J-code may be under SOF control, but it would certainly be met with skepticism or resistance as it goes against the organizational culture of the force. It also would be difficult to have people conceptualize the new paradigm. As of now, anyone with experience in the military knows what each J-code does and does not do. If the military were to evolve beyond the J-code, what would a cross-functional team look like? It is obvious to say that form would follow function, but how would that actually play out?

To take an information operations team as an example, it could iteratively evolve to maximize output speed, accuracy, and insight. The team's structure could look something like the following: a leader to authorize actions, an intel specialist to keep the team informed of fresh intel, a message shaper/public affairs/influence operations specialist to hit just the right note, a cyber delivery specialist to maximize delivery, and a legal/policy point of contact to clear the actions in accordance with existing legal or policy structures. Thus, this cell could be authorized to make decisions and push out messaging with automated tools and an underlying data structure from which to gather information, allowing it to respond quickly and accurately, breaking down stovepipes, and streamlining the bureaucratic decision-making processes.

Models for Adapting the Military for a Big Data Capability

There is still a question as to which management paradigm is best suited for inculcating a big data culture. Below are three management paradigms for senior leadership consideration in the move to a big data cultural paradigm. In the first two, it is a choice between mitigating the burdens or emphasizing the strengths from above, while the third explores a radical departure from the norm. While each of the models is dependent on which of the previously identified constraints can or cannot be changed/accommodated, as form follows function, they each have their strengths and drawbacks, support victory in their own ways, have red flags to avoid, and require multiple commanders to cleave to the same vision.

Cheerleader

The first paradigm, called the cheerleader, is all about support. In this paradigm, there is no institutional authority mandating a singular path. Instead, it relies upon access and placement to motivate followers voluntarily to move toward a common objective; in other words, it is a social movement. Before applying this to a particular example, it is important to examine the pros and cons. The benefits of this paradigm rest in its lack of direct responsibility. Without direct responsibility, the organization is able to focus on clearing the path for other organizations to press forward by removing obstacles, providing advocacy, and especially resolving friction. As multiple organizations press forward in the same general direction, there is bound to be friction between and among them or novel circumstances with which each organization must deal within the framework of the overall direction. A strength of the cheerleader paradigm, then, is the neutrality with which managers can provide clear-eyed advice on contentious issues to fractious institutions since they are invested in the process as a whole to get to an overarching objective, not a particular project or single pathway. This ability, to remain engaged and encouraging yet neutral, may be the greatest strength of this paradigm.

As with all things, there are drawbacks to pursuing this management culture. First and foremost, without money, expertise, or other resources directly applied to the issue, there is little direct control over the project. That same neutrality from above also removes the organization from the direct levers of control, leading it beholden to the agendas of others along with their timelines. This is often a frustrating experience, requiring a high level of trust, commitment, and understanding of the other organizations and their processes.

First and foremost, without money, expertise, or other resources directly applied to the issue, there is little direct control over the project.

Empowered Authority

In the next paradigm, empowered authority, there is a high level of institutional authority, but there is also a large reliance on subordinates to execute the mission. The pros and cons are almost the exact opposite of the cheerleader model. While this paradigm takes far more resources and is another responsibility on top of the existing list, it also allows for control over the product and the ability to lead the agenda. However, it tends toward a standardized, fairly rigid implementation scheme, which might be problematic for SOF with their numerous component specialties and Service support programs. Here the tension between standardization of data for efficient big data exploitation and component resourcing comes into clear view.

One thing to keep in mind is that supporting this paradigm would require a massive change in hiring practices, recruiting, and retention of the force. It would not be enough to onboard people quickly; it would require them to be constantly trained and protected from being poached by other departments and by private companies. This does create friction with the current up-or-out system from the Department of Defense (DOD) as the paradigm encourages risk taking with a high chance for failure, builds expertise, and aims to increase retention. Having someone stay for ten years at a certain level or function is critically important in this paradigm because the expertise and institutional knowledge is not easily replaced or grown. One way to meet this objective is to increase use of the chief warrant officer system, which allows focused expertise in particular technical areas without the same needs demanded by the officer career pathway.

One way to overcome the issues inherent to either paradigm is to implement the two management paradigms in conjunction at the levels they would make most sense. In the SOF context, this would mean a split between USSOCOM HQ and the components, but this HQ/component split would be of use in any large organization. Keeping to the SOF context as an example, USSOCOM HQ takes up the cheerleader paradigm, and the components take the empowered authority management paradigm. This would allow for USSOCOM HQ to remove the obstacles for the components while allowing for decision-making authority to be pushed out to the edge. There would be drawbacks, such as the inevitable divergence from the desired standardization in approach as the components roll out the implementation. However, if there has been investment in an effective big data culture and a true enterprise system, then this divergence in approach could be minimized while improving the overall ability of the enterprise to ingest and digest its available data. These two paradigms used in conjunction largely fall under existing operational capabilities, authorities, and titles, making them generally complementary to the current system. However, there is another option not to be underestimated for its effectiveness but also not to be pursued lightly given its destructive potential.

Radical Change

The last paradigm is the radical change option, at times dubbed the nuclear option. It is a decision to completely break with previous technology platforms despite the significant pain and disruption that would occur in order to make a revolutionary leap in capability. One example is Apple moving from its traditional Mac computer operating system (OS) in favor of the iOS system used on its iPads. In doing so, Apple has wagered that the future of its customer base will be working on tablets or similar smart devices. It has chosen to eschew updating its laptops in favor of its more mobile products—in contrast to Windows—attempting to bypass the pain and investment needed to update the Mac OS system.[41] If successful, this one change would save tremendous resources while placing it ahead of its competitors in a valuable market space. If unsuccessful, it will paint Apple into a corner, ceding significant ground to Windows.

The radical change option is best employed when there is a need to change the game, forcing the hand of all other competitors. Some paradigmatic examples in military history are the introduction of the firearm,[42] the carrier task force,[43] and atomic/nuclear weapons.[44] Each case represents a complete break in the existing paradigm, leading to a revolution in warfare others could not effectively counter. When thinking about the radical option, there are other important questions: When is it appropriate? Why is the organization undertaking this course of action? What objective is hoped to be achieved? How will this unfold? Who is going to lead execution? Is there the will to stay the course no matter what?

Radical change is appealing because it fits within overall American and DOD cultural paradigms, making it a tempting option: there is immediate action, it is dramatic, a fresh start is appealing to many leaders, it limits any bureaucratic stakeholders' ability to slow momentum, and it is demonstrable in briefs. However, it also threatens to destroy the organization's understanding of valuable lessons learned and resets the clock on any valuable work which was underway. Make no mistake, resorting to this option is a gamble that carries huge risk, making it imperative to think through the stakes.

Conclusion

Managing big data requires a supporting culture, not just technical expertise capable of using the technology. Big data culture in many ways runs counter to standard military culture, and the mission- and action-oriented culture of SOF could lead them to prioritize the current mission over a true enterprise solution. By adopting the values and attitudes underpinning big data, the SOF enterprise can transform the culture, but it will not likely happen for the current fight, and managers should make this an expectation as they consider technology and hiring policies. Senior leaders and managers will soon need to choose the type of management paradigm to be pursued but getting to the point of making that choice effective takes a tremendous amount of time, effort, and resources. Inculcating the foundational culture, building the enterprise tools, and maintaining cognizance of hurdles to be overcome together ensure that once a paradigm is chosen, its full potential can be realized.

Endnotes

1. Edgar H. Schein, *Organizational Culture and Leadership* (San Francisco: Jossey-Bass Publishers, 1985), 33.
2. Michael Schrage, "How the Big Data Explosion Has Changed Decision Making," *Harvard Business Review*, 25 August 2020, https://hbr.org/2016/08/

how-the-big-data-explosion-has-changed-decision-making?referral=03759&cm_vc=rr_item_page.bottom.

3. Matthew Corritore, Amir Goldberg, and Sameer B. Srivastava, "The New Analytics of Culture," *Harvard Business Review*, January-February 2020, https://hbr.org/2020/01/the-new-analytics-of-culture.

4. S. Augustine, B. Payne, F. Sencindiver, and S. Woodcock, "Agile Project Management: Steering from the Edges," *Communications of the ACM* 48, no. 12 (December 2005): https://dl.acm.org/doi/pdf/10.1145/1101779.1101781.

5. William R. Burns, *Adaptability: Preparing For and Coping With Change in a World of Uncertainty* (Alexandria: Institute for Defense Analyses, 2013), https://www.ida.org/-/media/feature/publications/a/ad/adaptability-preparing-for-and-coping--with-change-in-a-world-of-uncertainty/p-5069.ashx; Brian Dickerson, *Adaptability—A New Principle of War* (Carlisle Barracks: U.S. Army War College, 2003), 23–25, https://apps.dtic.mil/dtic/tr/fulltext/u2/a415124.pdf.

6. *Freedom on the Net 2013: A Global Assessment of Internet and Digital Media*, eds. S. Kelly et al. (Washington, D.C.: Freedom House, 2013), https://freedomhouse.org/sites/default/files/resources/FOTN%202013%20Summary%20of%20Findings.pdf; Olga Khazan, "Russia's Online-Comment Propaganda Army," *The Atlantic*, 9 October 2013, https://www.theatlantic.com/international/archive/2013/10/russias-online-comment-propaganda-army/280432/; Olga Bugorkova, "Inside the Kremlin's 'Troll Army,'" *BBC News*, 19 March 2015, https://www.bbc.com/news/world-europe-31962644.

7. Lee Rainie and Janna Anderson, "Theme 3: Humanity and Human Judgment Are Lost When Data and Predictive Modeling Become Paramount," *Pew Research Center*, 8 February 2017, https://www.pewresearch.org/internet/2017/02/08/theme-3-humanity-and-human-judgment-are-lost-when-data-and-predictive-modeling-become-paramount/.

8. C.H. Kung and A. Solvberg, "Activity Modeling and Behavior Modeling," in T. Ollie, H. Sol, and A. Verrjin-Stuart, *Proceedings of the IFIP WG 8.1 Working Conference on Comparative Review of Information Systems Design Methodologies: Improving the Practice* (New York: Association for Computing Machinery, 1986), 145–71.

9. Pedro Manrique et al., *Context Matters: Improving the Uses of Big Data for Forecasting Civil Unrest: Emerging Phenomena and Big Data* (Seattle: 2013 IEEE International Conference on Intelligence and Security Informatics, 2013), 169–172, https://ieeexplore.ieee.org/document/6578812/authors#authors.

10. Alissa Lorentz, "With Big Data, Context Is a Big Issue," *WIRED*, accessed 20 October 2021, https://www.wired.com/insights/2013/04/with-big-data-context-is-a-big-issue/.

11. Jeff Schogol, "Special Operations Command Review Finds Deployment and Leadership Issues but No 'Systemic Ethics Problem,'" *Task & Purpose*, 28 January 2020, https://taskandpurpose.com/news/USSOCOM-no-ethics-problem.

12. M.J. Wolf and F.S. Grodzinsky, "Good/Fast/Cheap: Contexts, Relationships and Professional Responsibility During Software Development," *Proceedings of the 2006 ACM Symposium on Applied Computing* (Dijon: Association for Computing Machinery, 2006), 261–266, https://dl.acm.org/doi/abs/10.1145/1141277.1141339.

13. Eileen Parkes, "Scientific Progress Is Built On Failure," *Nature*, 10 January 2019, https://www.nature.com/articles/d41586-019-00107-y.

14. Jessica Lahey, "How to Teach a Non-Conforming Child," *The Atlantic*, 12 February 2016, https://www.theatlantic.com/education/archive/2016/02/educating-an-original-thinker/462468/.

15. Brian Klapper, "Free Yourself from Conventional Thinking," *Harvard Business Review*, 6 May 2013, https://hbr.org/2013/05/free-yourself-from-conventiona; G. Plimmer, R. Norman, and D. Gill, "Skills and People Capability in the Future State: Needs, Barriers and Opportunities," in *Future State: Directions for Public Management in New Zealand*, eds. B. Ryan and D. Gill (Wellington: Victoria University Press, 2011), 281–305.

16. Corritore, Goldberg, and Srivastava, "The New Analytics of Culture."

17. Jim Moffat, "Helping People Learn by Letting Them Fail Is Essential," *Forbes*, 18 November 2015, https://www.forbes.com/sites/forbesleadershipforum/2015/11/18/helping-people-learn-by-letting-them-fail-is-essential/#59ca88acc439.

18. Parkes, "Scientific Progress Is Built On Failure."

19. Jeff Desjardins, "How Much Data Is Generated Each Day?" *Visual Capitalist*, 19 April 2019, https://www.visualcapitalist.com/how-much-data-is-generated-each-day/; Branka Vuleta, "How Much Data Is Created Every Day? 27 Powerful Stats," *SeedScientific*, 24 March 2020, https://seedscientific.com/how-much-data-is-created-every-day/.

20. Dirk J. Roux, Kevin Murray, and Ernita van Wyk, "Principles Enabling Learning Environments for Good Ecosystem Governance," in *Governance as a Trialogue: Government-Society-Science in Transition*, ed. A.R. Turton et al. (Berlin: Springer, 2007), 253–280.

21. "Big Data in Action: Definition, Value, Benefits, Context," *i-scoop*, accessed 20 October 2021, https://www.i-scoop.eu/big-data-action-value-context/; Anil Jain, "The 5 V's of Big Data," *IBM Watson Health Perspectives*, 17 September 2016, https://www.ibm.com/blogs/watson-health/the-5-vs-of-big-data/.

22. Observe, Orient, Decide, Act was developed by U.S. Air Force Colonel John Boyd. John R. Boyd, *A Discourse on Winning and Losing*, ed. Grant T. Hammond (Maxwell AFRB: Air University Press, 2018), 383–385, https://www.coljohnboyd.com/static/documents/2018-03__Boyd_John_R__edited_Hammond_Grant_T__A_Discourse_on_Winning_and_Losing.pdf.

23. Jain, "The 5 V's of Big Data."

24. The JAIC, "Why Data Governance is Critical to a Successful JCF," *AI in Defense*, 6 May 2020, https://www.ai.mil/blog_05_06_20-why_data_governance_is_critical_to_a_successful_jcf.html.

25. Danah Boyd and Kate Crawford, "Critical Questions for Big Data," *Information, Communication & Society* 15, no. 5 (May 2012): 662–679, DOI: 10.1080/1369118X.2012.678878.

26. Stephen P. Mulligan, Chris D. Linebaugh, and William C. Freeman, *Data Protection Law: An Overview*, CRS Report No. RL45631 (Washington, D.C.: Congressional Research Service, 2019), https://fas.org/sgp/crs/misc/R45631.pdf.

27. Mulligan, Linebaugh and Freeman, *Data Protection Law*; Jordan Harrod, "Health Data Privacy: Updating HIPAA to Match Today's Technology Challenges," *Science in the News*, 16 May 2019, http://sitn.hms.harvard.edu/flash/2019/health-data-privacy/.

28. Vincent C. Hu, Tim Grance, David F. Ferraiolo, and D. Rick Kuhn, "An Access Control Scheme for Big Data Processing," *10th IEEE International Conference on Collaborative Computing: Networking, Applications and Worksharing* (Miami: IEEE, 2014), 1-7, http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7014544&isnumber=7011734.

29. Jamie Condliffe, "The Week in Tech: Huge Fines Can't Hide America's Lack of a Data Privacy Law," *The New York Times*, 26 July 2019, https://www.nytimes.com/2019/07/26/technology/facebook-data-privacy.html.

30. Ross D. Arnold, *Strategies for Transporting Data between Classified and Unclassified Networks* (Fort Belvoir: Defense Technical Information Center, 2016), https://apps.dtic.mil/docs/citations/AD1005160.

31. Arnold, "Strategies for Transporting."

32. "Data Warehouse vs. Data Lake: Which Is Right for Your Enterprise App Development Effort?" *Medium*, 19 February 2019, https://medium.com/@distillerytech/data-warehouse-vs-data-lake-which-is-right-for-your-enterprise-app-development-effort-fa9f046d47ca.

33. Chris Campbell, "Top Five Differences between Data Lakes and Data Warehouses," *Blue Granite*, 26 January 2015, https://www.blue-granite.com/blog/bid/402596/top-five-differences-between-data-lakes-and-data-warehouses.

34. Kalyan Yedakula, "Identifying Security Challenges against Data Lakes: Cyware Hacker News," *Cyware*, 25 October 2019, https://cyware.com/news/identifying-security-challenges-against-data-lakes-c4e19226.

35. Brent Dykes, "Creating a Data-Driven Culture: Why Leading By Example Is Essential," *Forbes*, 26 October 2017, https://www.forbes.com/sites/brentdykes/2017/10/26/creating-a-data-driven-culture-why-leading-by-example-is-essential/.

36. Dykes, "Creating a Data-Driven Culture."

37. Ben Martz, Frank C. Braun, and Michel Avital, "Action-Centered Team Leadership Influences More Than Performance," *Team Performance Management* 18, no. 3/4 (June 2012): 176–195, https://doi.org/10.1108/13527591211241015.

38. Jelena Lukić, "Leadership Challenges for the Big Data Era," in *Challenges to Promoting Entrepreneurship, Leadership and Competitiveness*, eds. M. Radović

Marković and S. Ilieva (Belgrade: Visoka škola za poslovnu ekonomiju i preduzetništvo, 2015), 293-309.

39. Lukić, "Leadership Challenges for the Big Data Era."
40. Martz, Braun and Avital, "Action-Centered Team Leadership Influences More Than Performance."
41. Jason Perlow, "Forget MacOS: The IPad Is Now Apple's Mobile Computing Future," *ZDNet*, 6 June 2019, https://www.zdnet.com/article/forget-macos-the-ipad-is-now-apples-mobile-computing-future/.
42. Naval Doctrine Command, *Revolutions in Military Affairs, Paradigm Shifts, and Doctrine* (Norfolk: Naval Doctrine Command, 1995), https://apps.dtic.mil/dtic/tr/fulltext/u2/a292013.pdf.
43. Naval Doctrine Command, *Revolutions.*
44. Naval Doctrine Command, *Revolutions.*

Chapter 5. Ideas for Cultivating Big Data Personnel

Dr. Justin Brunelle

MITRE: APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED. CASE NUMBER 18-2725-8.

This technical data was produced for the U. S. Government under Contract No. FA8702-19-C-0001 and is subject to the Rights in Technical Data-Noncommercial Items Clause DFARS 252.227-7013 (FEB 2014).

Creating, acquiring, maintaining, and retaining a staff with expertise in big data is an ongoing challenge within both commercial industry[1] as well as the government.[2] Doing so in the military and intelligence domains is especially challenging given the potential roadblocks of acquiring appropriate security clearances, the (in)ability to offer adequate monetary compensation, and the opacity of the work to the external world.

As a result, there is a greater challenge within the Department of Defense (DOD) and intelligence community (IC) than in industry to maintain a properly skilled and cost-effective data workforce. Government organizations have turned to creative models of attracting top talent and often rely heavily on external organizations (e.g., contractors) to provide the requisite talent to handle, manage, and solve big data challenges. Particularly with the maturation of government agencies' understanding of big data challenges, there is an increasing focus on big data as a catalyst for mission success and an increased emphasis on maintaining an appropriate big data staff.

This chapter discusses several aspects of the challenge and potential solutions with establishing or maintaining a big data workforce within the military. It also describes opportunities and challenges with recruiting experienced employees from non-governmental organizations (e.g., private industry) and the aspects of a reliance on contractor support to accomplish government missions.

Options within the Military

While both industry and the government face challenges with hiring data practitioners, the DOD and IC face unique challenges in recruiting and maintaining a big data workforce that is not prevalent in private industry. For example, security clearances, inability to offer competitive salaries, inability to share work outside of closed environments, and location-based work all present barriers to hiring data experts. However, the DOD has several benefits to draw upon to enhance its ability to recruit and reload a skilled data workforce.

Skills and Compensation

Big data specialists were once extremely rare, but with exploding salaries[3] and increasing academic investments in big data programs,[4] data scientists and data analysts are increasingly entering the workforce. Despite increased supply, the demand remains high for data scientists within industry and government.[5] As a result, even private industry is focused on training the existing workforce to fill the gap along with hiring skilled graduates from training and traditional degree programs.[6]

Big data practitioners are often a rare blend of skillset, mindset, capability, and experience in a variety of domains (e.g., programming, statistics, and social sciences). Highly sought-after practical skills in a data practitioner include the following:

- Scientific, scripting programming languages such as Python and R
- Experience with cloud and other high-performance computing (e.g., Hadoop)
- Strong statistics and mathematics background
- Big data theory, such as data management, data quality, and wrangling
- Knowledge of machine learning algorithms
- Data visualization experience
- Technical and non-technical communication

Data teams often seek social scientists, such as psychologists or human factors engineers, to ensure that data is used effectively and does not introduce bias into decision-making.

Given the specific skillset, along with soft skills such as intellectual curiosity and ability to navigate communication between technical and non-technical communities, data scientists are often able to command salaries in private industry well above typical government salaries. Entry level salaries may exceed general schedule (GS)-13 or GS-14 salary ranges. This creates a recruiting challenge within DOD and IC stakeholders due to the potentially large discrepancy in monetary compensation for staff members. As a result, the government is facing a hiring challenge and must focus on the non-monetary benefits and incentives to attract government employees. Government organizations are turning to broader organizational training (i.e., at all levels from senior leadership down to practitioners) on big data practices and skills to augment existing staff with requisite knowledge and work to fill the skills gap in the government workforce.[7]

Along with experts skilled with managing and manipulating data should come a variety of skillsets to round out a high-quality data team. These team members enable and ensure the appropriate insights or decisions are derived from the appropriate data. Additionally, data experts and their team members must work to ensure the appropriate caveats are associated with data-driven decisions to ensure that the information in data is used in appropriate ways. A multidisciplinary team should consist of data management, data processing, data architecture, statistics, software, and domain experts. A single team member may serve in multiple roles, and a challenge space may require varying numbers or varying skill and experience levels of these experts. Depending on the domain, other experts may be required, such as social scientists, legal experts, or psychologists.

The goal of a multi-disciplinary data team is to equip the organizational team with the relevant domain knowledge to understand how to use the data appropriately, determine which assumptions or conditions for use may be appropriate, and determine how (i.e., under what conditions) to apply the information derived from the data. These aspects of data analysis are as—if not more—important than the ability to apply algorithmic analysis of datasets.

Recruit Motivations

Government is not unique in its challenge with recruiting data experts. Commercial industry has challenges as well. A 2018 study by the National Association of Colleges and Employers shows that there are a variety of challenges facing organizations hiring data experts[8] (e.g., lack of advancement opportunities and lack of suitable tools available in-house to enable job function). The study cites that organizations are pursuing internal training and creative compensation packages (among other incentives) to attract the appropriately skilled employees.

Despite the monetary compensation, DOD and IC careers have unique benefits that can attract talent to government careers. Primarily, big data practitioners within the government are often drawn by the mission—the opportunities to affect the success of the nation, their fellow citizens, and improve the lives of people potentially across the world. The activities being performed by individuals in the government are contributions to efforts that very literally influence the entire nation and world. The career opportunities at U.S. Digital Service and Defense Digital Service (DDS)—and the Silicon Valley tycoons[9] taking advantage of them[10] for less monetary reward than in private industry[11]—are prime examples of top technical experts joining government service to contribute to an important mission.[12]

Data practitioners specifically in the DOD and IC are afforded the responsibility and opportunity to work with datasets and on challenges that very few individuals experience. As such, the opportunity to hone skills, understand challenges, and operate on live and important datasets is not available in all career paths. This is an additional lure to government work.

The development opportunities available to staff entering the government workforce are unique to government service. There are opportunities to be immersed in highly diverse and topically broad teams both from the perspective of technical topics as well as domain experts. This exposes team members to a variety of topics and applications of their domain expertise. These opportunities are enhancements to an employee's resume; even if the data practitioner is not destined for a full career in government service, the time spent in government positions will provide additional opportunities for success in future positions.

Retention and Attrition

The National Geospatial Agency (NGA) Silicon Valley Outpost provides various government service positions to assist the NGA in recruiting and potentially maintaining data and other in-demand technical expertise.[13] The effort is specifically focused on early career contributors but has opportunities for other members as well. Some of the careers are intentionally term limited (i.e., one- to three-year terms), allowing employees to work remotely or offsite and return to academia or private industry at the conclusion of their term. This incentivizes applications in a similar manner to a post-doctoral position for PhD researchers—contribution to the employer and domain during the tenure as well as mentoring peers and gaining technical expertise. These positions create learning and culture exchange opportunities for both the employee and employer.

Efforts such as the NGA outpost are being undertaken to improve the recruiting efforts to fill the government big data skills gap. However, with salaries outside of government being typically larger than is feasible within the General Schedule (GS) scales, retaining top talent is an outstanding challenge within the government. Adding to this challenge is the often-prohibitive bureaucracy within government projects. Employees may become frustrated with the inability to move at the speed or autonomy of industry and opt to leave for a faster-paced environment. Again, the NGA outpost and other government efforts (e.g., Kessel Run[14]) are challenging those norms and challenging government projects to move faster.

However, with salaries outside of government being typically larger than is feasible within the General Schedule (GS) scales, retaining top talent is an outstanding challenge within the government.

As previously discussed, data experts are most effective in multi-disciplinary teams. Ideally, a data team gains experience and knowledge of a domain; the data experts become closer to domain experts, as well. Data experts must maintain expertise in various approaches and current state-of-the-art approaches. As such, a team must balance the need to be deeply immersed in a domain with revisiting training on new approaches. Opportunities to increase skillsets, maintain training, and increase exposure to domain challenges are valuable for data experts. Opportunities to increase individual skills and practical applications are appealing to data experts in both private industry as well as government.

To increase retention, creative incentives could be explored. For example, technical career paths could be created for individuals that want to remain in government service but receive promotion motivation for technical work as opposed to a traditional leadership position. In this case, promotion and other incentives would be based on technical contributions. For example, providing a clearer path to reach a science technical service position would help demonstrate a commitment to retaining excellent technical contributors in government service. Similarly, the USG and military are experimenting with "reskilling" programs to enable existing personnel to learn data science skills through certificate programs rather than requiring undergraduate and graduate degrees in the field.[15]

Another opportunity to increase retention is to offer post-service benefits for civil servants. For example, government organizations may explore a post-service organization or professional networking group (e.g., Armed Forces Communications and Electronics Association[16]) that is limited to those that served a minimum amount of time in government positions. This would encourage engagement within the community, demonstrate the value of government service, and enhance the non-monetary benefits of government employees. Similarly, continuous training through these organizations could be offered as a way to retain skill sets and knowledge exchanges similar to the models used by volunteer firefighters (i.e., professional training opportunities) or the Illuminate Thinkshop[17] (i.e., peer training exchanges).

Recruiting from the Civilian Sector

The demand for data scientists is not unique to the government. Private sector companies are striving to acquire highly sought-after data talent. Smaller companies are unable to compete with the salaries of the larger, perennial powers in the tech landscape. As a result, they use techniques such as incentive packages, training programs, and exchange programs (e.g., with government) to enhance the appeal and improve recruiting.

Skills and Compensation

Entering government service from the private sector is appealing when considering the opportunities and mission. However, government can improve its ability to recruit using similarly creative incentive packages. For example, DDS uses a “straight to GS-15” program which places top technical talent at the top of the GS pay and responsibility scale. This can be an opportunity for individuals to gain valuable experience and interact at high levels of government. The program has been successful at attracting otherwise unobtainable talent from Silicon Valley startups.

Training incentives within the government can help attract talent. For example, education programs or partnerships with organizations in academia or industry can help enhance the skill sets of potential candidates. Training for software engineers within the Air Force’s iLab effort in the 548th Intelligence, Surveillance, and Reconnaissance Group[18] and Kessel Run[19] are examples of government partnerships with industry for training.

Recruit Motivations

Civilian data scientists in general are puzzle solvers and innovators, and they like to have a high degree of input in any discovery process. As with most creative minds, they like flexible work schedules and the ability to surge when inspiration hits. As long as the problem or challenge they are working on is meaningful, they will retain interest in the work. That said, it is common for civilian data scientists to stay with a job for only three years or less as their interests adapt along with changing personal and technological circumstances.[20]

Retention and Attrition

Motivating experienced civilian personnel to stay will likely be a function of multiple factors. First, without leadership who understand and accommodate the data scientist culture, civilians will quickly become frustrated and leave for more interesting work. Second, without the right tools, civilian data scientists will feel underutilized and unproductive. Here again, government purchasing protocols and security requirements could become a source of frustration unless expectations are clearly managed on the front end of the hiring process. Third, data scientists are highly sensitive to the rapid pace of technological change. A position that prevents them from maintaining currency or working with the latest technology could cause serious harm to their future careers. As a result, they will need the ability to stay current either through their jobs or dedicated time for continuous learning.[21]

Motivating experienced civilian personnel to stay will likely be a function of multiple factors.

Contractor Support

Because of the highly competitive hiring and recruiting environment around big data practitioners, the government often relies on contractor support to supply big data expertise. In many ways, this is beneficial to all parties: the government is able to immediately acquire the skillsets required without investing in training and education that takes time to achieve a return on investment, and the employees may be potentially paid at a higher rate than possible within government. However, this may lead to a gap in institutional knowledge and an inability to operate in all capacities. For example, contractors may not be suitable for highly sensitive tasks or may not be able to access data of particular sensitivities, leaving the responsibility for processing these edge cases with the government employees.

Contractor support and employee tenure are not guaranteed, especially during times of fiscal constraint. Therefore, there is a risk of losing institutional knowledge or domain expertise if a structured hand off is not planned for when the contract is let. Despite this, contractor support is essential to augmenting the government data workforce until billets become available for military or civilian positions.

Conclusion and Recommendations

As the government moves toward adopting big data and continuing to leverage it as a catalyst for mission success, it is essential for the government to identify a reliable path for staffing big data efforts. Organizations are beginning to explore creative incentives and recruiting mechanisms. Despite the benefits of working in the government data domain (e.g., access to data, training, and experience), the government remains commonly challenged by bureaucracy and less lucrative salaries when compared to industry.

The government should focus on immersing their data experts in multidisciplinary teams and providing training options. Working in the government should be a career boost and create open opportunities for candidates. Creative incentives can help enhance the appeal of working with government. For example, creating explicit career paths for technical experts, tenure-based incentives for civil servants, training and mentoring programs, and post-service benefits should all be explored for feasibility in attracting top talent. Properly incentivized, trained, and equipped (including with appropriate team members), a data expert can help make the best use of the data within an organization. Government hiring organizations should work to leverage their strengths to attract the top talent available. Despite the challenges, government missions—the ability to affect the globe and improve the lives of fellow citizens—remains a prime driver for technical leaders to join government service.

Endnotes

1. "Challenges of Recruiting for Data Analytics, Cybersecurity," *NACE*, 5 March 2018, https://www.naceweb.org/talent-acquisition/trends-and-predictions/challenges-of-recruiting-for-data-analytics-cybersecurity/.
2. Justin F. Brunelle et al., *October 2018 Federal Data and Analytics Summit Summary*, 18-2725-6 (Washington, D.C.: The MITRE Corporation, 2019), https://www.mitre.org/sites/default/files/publications/17-3231-6-january-2018-federal-ciso-summit-report.pdf.
3. Bill Snyder, "How to Make $120,000: Get a Job in Big Data," *Infoworld*, 4 May 2015, https://www.infoworld.com/article/2922065/how-to-make-120k-get-a-job-in-big-data.html.
4. Wesley P. Hester, "UVA Plans New School of Data Science; $120 Million Gift is Largest in University History," *UVAToday*, 18 January 2018, https://news.virginia.edu/content/uva-plans-new-school-data-science-120-million-gift-largest-university-history.
5. Michael J. Garbade, "LinkedIn Workforce Report: Data Science Skills Are in High Demand Across Industries," *TowardsDataScience*, 17 October 2018, https://towardsdatascience.com/linkedin-workforce-report-data-science-skills-are-in-high-demand-across-industries-1510b06382a6.
6. Adi Gaskell, "Organizations Striving to Close the Data Science Skills Gap," *Forbes*, 18 July 2018, https://www.forbes.com/sites/adigaskell/2018/06/18/organizations-striving-to-close-the-data-science-skills-gap/#47f281c31d50.
7. Christine Harvey et al., *December 2017 Federal Big Data Summit Report* (Washington, D.C.: The Mitre Corporation, 2017), https://atarc.org/wp-content/uploads/2019/01/2017-12-05-ATARC-Federal-Big-Data-Summit-White-Paper.pdf.
8. "Challenges of Recruiting for Data Analytics."
9. "DOD Expands Tech Talent Initiative to Develop Critical Cyber Capabilities," *Defense.gov*, 25 October 2018, https://dod.defense.gov/News/News-Releases/News-Release-View/Article/1672698/dod-expands-tech-talent-initiative-to-develop-critical-cyber-capabilities/.
10. Jim Garamone, "Defense Digital Service Emphasizes Results for Service Members," *Defense.gov*, 26 June 2018, https://dod.defense.gov/News/Article/Article/1560057/.
11. Lauren C. Williams, "Defense Digital Service Looks to Retool Tech Recruitment," *FCW*, 8 February 2018, https://fcw.com/articles/2019/02/08/dds-recruitment-rfi.aspx.
12. "Our Mission," *U.S. Digital Service*, accessed 27 July 2019, https://www.usds.gov/mission.
13. Frank R. Konkel, "NGA Launches Bold Recruitment Plan to Hire Silicon Valley's Best," *Nextgov*, 8 December 2017, https://www.nextgov.com/analytics-data/2017/12/nga-launches-bold-recruitment-plan-hire-silicon-valleys-best/144410/.

14. “Home,” Kessel Run, United States Air Force, accessed 20 October 2021, https://kesselrun.af.mil/.

15. Jory Heckman, “Census Bureau to Launch Data Science Reskilling Pilot as Template for Other Agencies,” *FederalNewsNetwork*, 3 December 2019, https://federalnewsnetwork.com/hiring-retention/2019/12/census-bureau-to-launch-data-science-reskilling-pilot-as-template-for-other-agencies/; Andrew Eversden, “The Army Kicks Off a New Reskilling Program for Civilian Employees,” *C4ISRNET*, 14 July 2020, https://www.c4isrnet.com/battlefield-tech/it-networks/2020/07/14/the-army-kicks-off-a-new-reskilling-program-for-civilian-employees/.

16. “Home,” *AFCEA.org*, AFCEA International, accessed 20 October 2021, https://www.afcea.org/site/.

17. Benjamin Gallo, “Learning and Doing: Changing the Status Quo,” Department of the Navy, 3 March 2017, https://www.secnav.navy.mil/innovation/Pages/2017/03/LearningandDoing.aspx.

18. “548th Intelligence, Surveillance and Reconnaissance Group,” Beale Air Force Base, 21 May 2012, https://www.beale.af.mil/Library/Fact-Sheets/Display/Article/279967/548th-intelligence-surveillance-and-reconnaissance-group/; Benjamin Bugenig, “Beale AFB ISR Group Connects Airmen to Commercial Innovators at Collider Demo,” United States Air Force, 18 April 2019, https://www.af.mil/News/Article-Display/Article/1817666/beale-afb-isr-group-connects-airmen-to-commercial-innovators-at-collider-demo/.

19. “Home,” Kessel Run.

20. Eva Murray, “How to Attract and Retain the Important but Elusive Data Scientist,” *Dataconomy*, 13 June 2019, https://dataconomy.com/2019/06/how-to-attract-and-retain-the-important-but-elusive-data-scientist/.

21. Murray, “How to Attract.”

Chapter 6. Why Silicon Valley Is a Poorly Suited Model for Special Operations Forces

Dr. Mark Grzegorzewski

> The reason that "guru" is such a popular word is because "charlatan" is so hard to spell. – William J. Bernstein

Organizational Change

Given its proficiency with big data, could the organizational culture of Special Operations Forces (SOF) adopt the so-called Silicon Valley model? As the above quote indicates, many companies bring in outsiders (gurus) to fix what is wrong in their organization.[1] Their fixes may appear to work initially, but in reality what the organization is experiencing is a small change to the existing system rather than a complete organizational overhaul. Over time, the change that is introduced becomes less effective as the system reverts to the mean and the guru's one-size-fits-all change program does not take.[2] Why is this the case? What is it about organizational culture that makes it resistant to change? For that matter, what is organizational culture, and how does it impact innovation? In the following sections, this chapter will tackle these questions and explain four barriers to SOF adopting the Silicon Valley model. These barriers include the entrenched functional organizational culture of SOF, their inability to capture the *zeitgeist* of Silicon Valley, the fact that they draw upon closed-system concepts while operating in a dynamic environment, and that they lack the freedom to innovate to the degree that Silicon Valley enjoys.

Organizational culture is defined in a number of ways, but there are some common elements to culture in general:

- Some combination of artifacts (also called practices, expressive symbols, or forms), values and beliefs, and underlying assumptions that organizational members share about appropriate behavior[3]

- Patterns of meaning that weave human experience together into a coherent whole[4]
- A body of solutions to problems which have worked consistently and are therefore taught to new members as the correct way to perceive, think about, and feel in relation to those problems[5]

Some definitions for organizational culture that try to capture these ideas include:

- Holistic, historically determined, and socially constructed beliefs and behavior existing at a variety of levels and manifesting in a wide range of features of organizational life[6]
- A complex set of values, beliefs, assumptions, and symbols that define the way in which a firm conducts its business[7]
- Shared common values and beliefs that guide organizational members' actions by providing a perception of goal congruence and by helping employees to determine what is in the best interest of the collective[8]

Having a common organizational culture, however defined, is important for several reasons, such as the need to convey a sense of identity for organizational members, facilitating the generation of commitment to something bigger than oneself, enhancing social system stability, and serving as a sense-making device to guide and shape behavior.[9] How the operational culture is conceptualized will determine the rules followed either consciously or subconsciously. For instance, if the organizational culture is viewed only in terms of ideational and symbolic aspects, it will be difficult to discover the deeper meanings within the unconscious mind. These surface level analyses will not lead employers to guide their decisions when hiring the right people, strategizing, or changing the organization.[10]

The Silicon Valley Model

Despite exhortations from senior leaders to adapt to the Silicon Valley model, there is no clear conception within the literature of what this model actually entails. Therefore, this model may in fact be more of a frame of mind than an explicit representation. In interviews with cyberspace, artificial intelligence, and emerging technology companies based in Austin, Texas, several different commonalities about the concept of the Silicon Valley model were revealed. Homa Bahrami and Stuart Evans (2011) describe Silicon Valley as an "ecosystem that can be best described as Khunian inversion—long periods of frenzied change punctuated briefly by stable interludes. Technologies, products, markets, and competitors are in a state of flux."[11] The authors also describe the need of Silicon Valley modeled companies to be "super flexible," meaning the "dialectical capacity of withstanding while transforming."[12] To achieve this super flexibility, the authors argue that organizations must achieve five interlocking principles:

- Strategizing by maneuvering, which entails developing a variable portfolio of initiatives—encompassing pre-emptive, protective, opportunistic, and corrective maneuvers—and changing gear between them on a real-time basis
- Executing by experimenting, prototyping, iterating, and recalibrating assumptions, initiatives, and actions as new realities unfold
- Organizing by setting up a distributed, multipolar nodal architecture and clarifying federal rules of engagement
- Leading by aligning and realigning knowledge workers around dynamic realities and by redeploying peer-to-peer practices
- Innovating by recycling know-how, talent, and failures in a multipolar ecosystem[13]

The Silicon Valley model also encompasses a fluid labor market in which workers are given freedom from the employer. Conversely, there is not an expectation that the employer is responsible for the development of the employee. The employee is expected to take the initiative and constantly improve his or her marketability.[14] Further, the employee is given freedom from the employer on how to realize a project. It only matters that the project is completed on time.[15] This freedom to explore different ways in which to achieve the objective in turn leads to innovation within the organization. However, this same freedom can also lead to burnout amongst workers as they have no defined time when the workday ends, and therefore their job becomes their life.[16]

Another aspect that defines the Silicon Valley model is its insistence on radical innovation rather than incremental innovation.[17] However, this insistence on radical innovation is more easily said than done. In order for an organization to be as innovatively successful as Silicon Valley, it must be supported by the same ecosystem in which Silicon Valley thrives. To be clear, the Silicon Valley model is only successful due to the copious amounts of government and corporate money spent on innovation plus the location of Silicon Valley right next to Stanford University.[18] Henry Etkowitz describes this as the "triple helix format," where each helix is only as successful as the other helixes.[19]

Amongst interviewees for this project, there was a common theme that a Silicon Valley-based system is flexible and willing to bring in the right talent to make the change that the organization needs. By bringing in digital natives, companies bring in a fresh perspective to look at old problems. These employees will fail, perhaps multiple times, but it is important for employers to not get upset at an employee but rather to acknowledge that a new path forward was attempted, it failed, and that the organization learned from it. Respondents also added comments like, "failure is 99 percent of the process," and "adopt failure as an expectation."

In the test-and-learn culture that makes up part of the Silicon Valley model, interviewees held some version of the belief that their organizations have adopted failure as part of the business model. That is to say, if an organization is not failing, they are not being innovative and taking on risk to improve the product. One respondent did give a more nuanced view on failure by dividing his organization into the software and hardware divisions in which the former does not get instant feedback since it engineers a product which takes many months to complete and test. The latter, however, gets instantaneous feedback due to the nature of the work. Despite these differences, this individual did think that both divisions need to have people who can overcome setbacks.

In order to set loose independent teams that make agile Silicon Valley organizations hum, respondents were largely in agreement that as long as managers set clear guidance for their employees, they should have faith that the objective will be achieved. That is not to say that managers should only check in periodically; rather, as one interviewee remarked, they should "trust but verify." Moreover, managers should give periodic guidance to facilitate the actions of their employees. By not micromanaging their employees, managers allow them to innovate and get to the objective in previously unforeseen ways.

In agile organizations, which are typically associated with the Silicon Valley model, there is not a great need for managers so long as the values of the organization are clearly articulated and accepted by the employees. Moreover, because these organizations only hire people that are a fit for the organization, managers do not have a trust deficit that their employees are accomplishing the mission. Therefore, the organization becomes a mix of flatness and hierarchy wherein there are managers at the top to articulate the company's vision, but peer-to-peer management self-polices implementation in case something goes wrong. Within the peer-to-peer management level, it is important to have someone who can push information up the management chain in the event that challenges cannot be solved at the worker level.

According to respondents, leaders and managers in agile organizations should focus on the overall long-term strategy of the company and not the day-to-day minutiae. When they are compelled to involve themselves in day-to-day work, it is important that they focus on finding solutions to problems and not expending the majority of their energy on placing blame. In fact, it is recommended that managers shield their employees from the blame and focus on removing any barriers to their employees' success. By protecting them, the employee in an agile organization is allowed to focus on innovating solutions while avoiding what would otherwise become high stress and mentally and physically draining undertakings.

Interview participants were split as to whether it is better to focus on a data-driven culture or a culture built around a data model. Those who saw the problem in terms of data noted that there are tons of data available to sift through. For the group that saw a model as being more important than the abundance of data available, it was noted that it does not matter how much data there is unless there is also a concept about how the data is going to be used. Individuals went on to state that having too much data can actually constrain an organization if it does not have the ability to store it. Rather, they argue having a model up front is important for determining the data needed, and, as a result, avoids trying to collect everything, which is a strain on resources. Participants noted that, in both data-driven and model-driven cultures, it is important to articulate to the employees how the parts of the system fit together, and that every employee should know his or her individual role within the system and how he or she contributes to the overall success of the organization.

Silicon Valley companies build their cultures around transparency and the proper alignment of resources. In such an organization, everyone is responsible for calling out small problems before they turn into large messes. At this level, each employee has a different perspective within the company, and when amalgamated, these perceptions can give leadership a holistic view of the company. By taking in the feedback from multiple vantage points, the business becomes a learning organization. Therefore, according to respondents, it is important to emphasize the importance of a feedback loop within the Silicon Valley model.

Finally, employee freedom was an important topic to the interviewees. Each emphasized the importance of freedom being baked into the organizational culture. They also emphasized that there are always multiple ways to achieve an objective, so management should not force a one-size-fits-all strategy. If the right people for an organization are found, defined as a cultural fit, constant supervision is not necessary. In a somewhat radical remark from an interviewee, it was stated that he hired a top-notch individual to work for his company knowing full well that he or she would complete a project ahead of schedule. The manager also knew that the individual would spend the rest of his time working on a side project. This did not bother the manager since the job was completed and since the highly skilled worker would add value to the company in other ways outside of the project. The moral here is that employers should only focus on whether the job gets done on time, not how it was accomplished.

If the right people for an organization are found, defined as a cultural fit, constant supervision is not necessary.

While the literature is sparse on what comprises the Silicon Valley model, the interviews conducted provided some more insights into how people within the tech industry view the model that governs their work activity. They focused on the test-and-learn culture, independent teams and agile organizations, management structures and transparency, and the debate between data-driven and model-driven cultures. Surely there are other aspects to the Silicon Valley model, but these insights provide a sufficient baseline against which to compare the organizational culture of SOF and whether they can adapt to becoming more like Silicon Valley.

Barriers to Special Operations Forces Adopting the Silicon Valley Model

The Entrenched Functional Organizational Culture of Special Operations Forces

Every organizational culture has values that guide it. As discussed by Yoash Wiener, these values generally take two forms: functional and elitist.[20] These

Four Barriers to SOF Adopting the Silicon Valley Model:

1. The entrenched functional organizational culture of SOF
2. Inability to capture the *zeitgeist* of Silicon Valley
3. A closed system operating in a dynamic environment
4. Lacking freedom to innovate

are also identified as the negative (functional) and positive (elitist) forces of organizational culture.[21] Functional values address the modes of conduct of organizational members whereas elitist values speak to the status, superiority, and importance of the organization.[22] Put another way, functionalist values guide behaviors through explicit rules while elitist values channel the pride within the organization to achieve an end state. These values are not mutually exclusive, and both can likely be found in organizations. These values often start with an original charismatic leader within the organization or the founder.

For example, within the SOF community, the original leader that shaped the organization as we know it today could be considered General William Donovan or even the service members involved in Operation Eagle Claw. These leaders instilled in the organization a sense that every mission is "no fail" (elitist) and that SOF gets its mission completed by doing whatever is necessary (functional). These same value characteristics can be found in Silicon Valley organizations. For example, Facebook's founding member, Mark Zuckerberg, has instilled in his organization five core values that guide their work (functional).[23] In addition, Mark Zuckerberg also has the charismatic value of "move fast and break things" (elitist).[24] Functional values are longer lasting and harder to change than elitist values since the former begets an imbedded behavior.[25]

These functionalist values can encourage or discourage dissent within an organization, thereby guiding employee behavior.[26] Scholars have noted the importance of having divergent points of view within an organization and the ability for information to flow up to senior leaders.[27] When organizational silence is part of the culture, lower-level employees fear telling management about problems. This climate of silence works to the detriment of organizations.[28] Management may perpetuate this unhealthy climate for several reasons. One such reason is that management views it as an attack on their power and credibility.[29] In fact, the more homogenous senior leadership is, the more likely leaders are to remain cohesive against the perceived threat of dissent.[30] They also may dismiss information from below since they believe employees are self-interested and untrustworthy.[31] A final reason is that organizations may view dissent as unhealthy and therefore encourage a culture of uniformity and consensus.[32] When compared against the characteristics of the Silicon Valley model, these stifling functionalist values would never be found in a successful technology company. However, some of the characteristics are found in the military, and, indeed, are part of military culture.

In addition, the flatness of an organization, which is determined by both its functional and elitist values, impacts the information flow within an organization. In hierarchical organizations, creativity is stifled in that information flows mainly from the top down. In these tiered organizations, innovation is not seen as an end in itself but rather a means to an end.[33] In the Silicon Valley model, innovation is an end itself and organizations are generally flat. In fact, one organization, Zappos, eliminated all managers and hierarchy.[34] To be clear, Zappos over the past few years has backed away from this "holacracy" and started to return to slightly stronger hierarchy, but it still demonstrates a potential, if not radical, conception around which an organization can be organized. Surprising no one, the military could never implement a flat organizational system. The military's relatively rigid information and decision-making flows run counter to the Silicon Valley model in which information flows in all directions. In fact, extensive research has shown that when information flows in multiple ways, it improves the quality of the organization's decision-making.[35] Therefore, the first barrier to SOF adapting the Silicon Valley model is that for over 30 years, the SOF community—even longer in the context of larger military culture—has done things a certain way. In order to change its functionalist way of doing things, SOF requires a leader to adopt wholesale the mindset of Silicon Valley leaders.

Inability to Capture the *Zeitgeist* of Silicon Valley

Even if a charismatic leader does take leadership over the SOF community, it does not guarantee that the shift to the Silicon Valley model can be accomplished. Replicating successful organizational cultures is difficult. This explains why not all organizations have shifted to the Silicon Valley model. Certainly, other firms have wanted and tried to replicate the model. However, their failures demonstrate that certain organizational cultures are not perfectly imitable.[36] A firm that has "a valuable, rare, and imperfectly imitable culture enjoys a sustained competitive advantage that reflects that culture."[37] As such, just as SOF would like to emulate the Silicon Valley model, it will be unable to perfectly emulate a culture that has a sustained competitive advantage over its rivals. This is due to Silicon Valley having that "triple helix" quality that cannot be perfectly described and therefore cannot be perfectly imitated.[38] This inability to capture the *zeitgeist* (spirit or mood) of Silicon Valley is perhaps the biggest barrier to emulation for SOF. Instead, SOF should focus on emulating a successful culture within its industry rather than try to imitate a foreign organizational culture. Research demonstrates that firms that adopt organizational changes from within their industry are more successful than those industries that adopt changes from outside of their industry. This is mainly due to the adopting organization's ability to identity the correct cultural characteristics that they want to emulate.[39] In the case of SOF, it may better for U.S. Special Operations Command (USSOCOM) to adopt the Joint Artificial Intelligence Center's organizational culture as opposed to the nebulous Silicon Valley culture, though problems with scalability would have to be addressed.

Lacking the Freedom to Innovate

What motivates human beings within organizations to act? This philosophical question is tied to the conception of freedom.[40] Do organizations need paternalistic office structures to manage people at all times? Undoubtedly, workers do not prefer to work in places where they are constantly supervised and given only one way in which to achieve a task. This view is validated in the literature in that workers are more productive when they are part of an environment in which they do not have to be coerced to do their job; equally, they are likely to show less motivation when they have no control over an organization's processes and systems.[41] This view is amplified with the finding that individuals are more than willing to give up their freedom to an organization if they perceive that their values align with the organization's, thereby leading individuals and organizations to superior outcomes.[42]

The other side of organizational innovation is determining how much risk an organization is willing to take when giving employees the freedom to innovate.[43] Too much uncertainty within an organization, revealed as too much input from employees, can overload the organization and paralyze its decision-making process.[44] However, if the amount of input into the organization is precisely calibrated and appropriate risk assumed, employees will be free to deviate from the established cultural norms and innovate.[45] If risk taking is not baked into the organizational culture, workers will feel that any innovation that is not successful will be punished. Such an organizational culture is certainly not conducive to innovation and will leave the organizational system static as the external environment changes around it.[46]

Research on innovation has also taken issue with organizations that call for no failure. The critique of no-fail cultures is that such organizations do not fail because they do not take real risks. Annika Steiber and Sverker Alainge argue that organizations should take on riskier missions and fail. As such, the true mark of innovation is failing early on and learning from that failure.[47] Tom Peters recommends that management "better be trying stuff at an insanely rapid pace. You want to be screwing around with nearly everything."[48] Harry Boer and Frank Gertsen take the argument a step further and argue that organizations should strive for continuous innovation, which consists of "continuous improvement, learning and innovation, and implying an effective ongoing interaction between incremental improvement and learning and more radical innovation and change."[49]

A Closed System in a Dynamic Environment

The environment around an organization is constantly changing, meaning the organizational culture may be outpaced by the environment.[50] Only by continuously sensing the environment and enacting continuous learning do cultures such as those in Silicon Valley adapt their own organizational culture to the surrounding dynamic environment.[51] Rather than waiting until it is too late and the organization becomes obsolete due to the changing environment, it is recommended that leaders should implement the most effective cultural change in a semi-incremental approach in which they keep their organization stable and then implement "revolutionary periods of change."[52] Moreover, change leaders need to take into consideration the ways in which their employees might perceive the change, make sure the employees are ready for the change, and have the employees be an active part of that change.[53] If not enacted carefully, organizational change can result in a loss of external legitimacy for the organization as well as a questioning of the organization's internal identity.[54] That said, successful organization change cannot be implemented haphazardly. Rather, there must a detailed plan on how the change will be implemented and an expectation that the change may not happen quickly.[55]

Part of the reason Google, a typical Silicon Valley organization in an open system, is so successful is that it chooses the right people to succeed within their organizational culture. For example, Google's founders, Sergey Brin and Larry Page, cultivated a culture around three values: "do no evil, have a large impact, and change the world."[56] Moreover, Brin and Page wanted to create a culture in which employees would feel that they are working for one of the best companies in the world. By centering their workforce on these values, it allows Google to constantly change with the environment around it.

Further, to match employees that fit the Google culture, it starts with human resources. To test whether individuals are a match, interviewees are asked questions on four pre-defined areas: "cognitive ability, role-related knowledge, leadership, and 'Googliness.'"[57] The last area is a test between the individual and Google to see whether the individual is a cultural fit.

When a person is determined to be a good fit for Google and comes on board with the company, leaders within Google empower the individual to make change within the organization. Leaders provide this freedom to innovate by placing wide parameters on how a project can be completed and removing bureaucratic obstacles to change.[58] That is to say, Google's leaders lead by getting out of the way and trusting that their employees do not need constant supervision. Once inside the organization, Google employees form connections with others and share as much as possible, even across departments, with the aim that sharing meets their individual goals.[59] This process of sharing at the individual level helps the overall organization in that it can achieve innovation and efficiencies through the new synergies that occur. These dynamic and unpredictable synergies are best characterized by Steve Jobs when he said, "… creativity comes from spontaneous meetings, from random discussions. You run into someone, you will ask what they're doing, you say 'Wow,' and soon you're cooking up all sorts of ideas."[60]

How Special Operations Forces Can Learn From but Not Become Silicon Valley

The barriers outlined in this chapter (the entrenched functional organizational culture of SOF, inability to capture the *zeitgeist* of Silicon Valley, lacking freedom to innovate, and a closed system operating in a dynamic environment) will not likely be overcome by SOF. The gulf between the Silicon Valley model and SOF is simply too large. The values within the special operations community are too entrenched. However, that does not mean the SOF community cannot learn from the Silicon Valley model. Below are three takeaways flowing from the discussion above that the SOF community can apply today to move them closer to the Silicon Valley culture and processes:

The values within the special operations community are too entrenched.

1. Do not try to be Silicon Valley. SOF is SOF, and that is okay. SOF will never be the most efficient organization but it can remain the most effective. That does not mean that it could not stand to change some of the ways in which it manages data and technology. For one, SOF leaders could delegate more freedom to employees, allowing them different ways to achieve an objective. As long as it is not immoral or illegal, SOF leaders could incentivize information sharing across directorates and encourage problem solving at lower levels. As part of that, SOF leaders and managers would need room to fail on administrative and analytical activities. Agile organizations stay current by aggregating small innovations that add up and make the organization more efficient over time.

2. Break down silos between J-codes and components and inform people how they fit into the larger whole. Within USSOCOM, between J-codes, and across components, individuals have little understanding of what others do on a daily basis. As noted in chapter three, big data in particular requires cross-functional teams that transcend the J-code structure. Give the participants a problem to solve and wide latitude to arrive at a solution. Have managers provide top cover to these small teams as they innovate. In addition to coming up with a solution (or failing, which is okay), these individuals will learn about one another's departments or Components. In turn, they will see the organization from a new perspective and more clearly understand their role within it.

3. Figure out the purpose of the data before starting collection on everything. A model-driven approach to big data will guide what data is

collected and save community resources. Once the model is ready and the data is obtained, adjust the model appropriately. No model should remain static in a dynamic world. This process will take longer than anticipated and more energy will be expended than anticipated. Try out the process before implementing it large scale. Start small, perhaps on a single project. Find out the setbacks at the micro level to anticipate them at a larger level. These setbacks can include finding out that the right people are not currently on the team. It can also mean that changes to project leadership might be in order. Make the corrections and then scale up. As Steiber and Alainge argue, organizations only escape failure because they are trying to change.

Endnotes

1. A. Sorge and A. Van Witteloostuijn, "The (Non) Sense of Organizational Change Continued: A Rejoinder to Armbrüster and Glückler," *Organization Studies* 28, no. 12 (December 2007): 1887–1892, https://doi.org/10.1177/0170840607084968.
2. A.M. Pettigrew, R.W. Woodman, and K.S. Cameron, "Studying Organizational Change and Development: Challenges for Future Research," *Academy of Management Journal* 44, no. 4 (November 2001): 697–713, https://doi.org/10.5465/3069411.
3. E.H. Schein, *Organizational Culture and Leadership* (San Francisco: Jossey-Bass, 2004).
4. Clifford Geertz, *The Interpretation of Cultures* (New York: Basic Books, 1973).
5. E.H. Schein, "Defining Organizational Culture," *Classics of Organization Theory* 3, no. 1 (1985): 490–502, http://commerce.du.ac.in/web/uploads/e%20-%20resources%202020%201st/MBA%20HRD/Ms.Swati_Organizational%20Culture%20and%20Leadership%20by%20Edgar%20H%20Schein.pdf.
6. A. Pettigrew, "Conclusion: Organizational Climate and Culture," in *Organizational Climate and Culture*, ed. Benjamin Schneider (San Francisco: Jossey-Bass, Inc., 1990), 413-434.
7. Jay B. Barney, "Organizational Culture: Can It Be a Source of Sustained Competitive Advantage?" *Academy of Management Review* 11, no. 3 (July 1986): 656–665, https://doi.org/10.2307/258317.
8. W.G. Ouchi and A.L. Wilkins, "Organizational Culture," *Annual Review of Sociology* 11, no. 1 (1985): 457–483.
9. L. Smircich, "Concepts of Culture and Organizational Analysis," *Administrative Science Quarterly* 28, no. 3 (September 1983): 339–358, https://doi.org/10.2307/2392246.

10. M.S. Schall, "An Exploration into a Successful Corporation's Saga-Vision and Its Rhetorical Community," in *ICAl SCA Conference on Interpretive Approaches to Organizational Communication* (Alta: ICAl SCA 1981).
11. H. Bahrami and S. Evans, "Super-Flexibility for Real-Time Adaptation: Perspectives from Silicon Valley," *California Management Review* 53, no. 3 (2001): 21–39.
12. Bahrami and Evans, "Super-Flexibility for Real-Time Adaptation," 21–39.
13. Bahrami and Evans, "Super-Flexibility for Real-Time Adaptation," 21–39.
14. J. Shih, "Project Time in Silicon Valley," *Qualitative Sociology* 27, no. 2 (June 2004): 223–245, https://doi.org/10.1023/B:QUAS.0000020694.53225.23.
15. R.M. Kanter, *When Giants Learn to Dance* (New York: Simon and Schuster, 1990).
16. Shih, "Project Time in Silicon Valley."
17. Howard E. Aldrich and Martin Ruef, "Unicorns, Gazelles, and Other Distractions on the Way to Understanding Real Entrepreneurship in America," *Academy of Management Perspectives* 32, no. 4 (2017): 458–472.
18. Gordon Moore and Kevin Davis, "Learning the Silicon Valley Way," in *Building High-Tech Clusters: Silicon Valley and Beyond*, eds. Timothy Bresnahan and Alfonso Gambardella (New York: Cambridge University Press, 2004), 7–39.
19. Henry Etzkowitz, "Silicon Valley at Risk? Sustainability of a Global Innovation Icon: An Introduction to the Special Issue," *Social Science Information* 52, no. 4 (December 2013): 515–538, https://doi.org/10.1177/0539018413501946.
20. Yoash Wiener, "Forms of Value Systems: Focus on Organizational Effectiveness and Cultural Change and Maintenance," *Academy of Management Review* 13, no. 4 (October 1998): 534–545, https://doi.org/10.5465/amr.1988.4307410.
21. Joseph W. Whorton and John A. Worthley, "A Perspective on the Challenge of Public Management: Environmental Paradox and Organizational Culture," *Academy of Management Review* 6, no. 3 (July 1981): 357–361, https://doi.org/10.2307/257371.
22. Wiener, "Forms of Value Systems."
23. "Facebook's Five Core Values," *Facebook*, accessed 15 June 2020, https://www.facebook.com/media/set/?set=a.1655178611435493.1073741828.1633466236940064.
24. Nick Statt, "Zuckerberg: 'Move Fast and Break Things' Isn't How Facebook Operates Anymore," *CNET*, 30 April 2014, https://www.cnet.com/news/zuckerberg-move-fast-and-break-things-isnt-how-we-operate-anymore/.
25. Wiener, "Forms of Value Systems."
26. Charlan Jeanne Nemeth, "Managing Innovation: When Less Is More," *California Management Review* 40, no. 1 (October 1997): 59–74, https://doi.org/10.2307/41165922.
27. Nemeth, "Managing Innovation."

28. Elizabeth Wolfe Morrison and Frances J. Milliken, "Organizational Silence: A Barrier to Change and Development in a Pluralistic World," *Academy of Management Review* 25, no. 4 (October 2000): 706–725, https://doi.org/10.2307/259200.
29. M. Audrey Korsgaard, Loriann Roberson, and R. Douglas Rymph, "What Motivates Fairness? The Role of Subordinate Assertive Behavior on Manager's Interactional Fairness," *Journal of Applied Psychology* 83, no. 5 (October 1998): 731.
30. Irving L. Janis, *Groupthink: Psychological Studies of Policy Decisions and Fiascoes* (Boston: Wadsworth Cengage Learning, 1982).
31. Jeffrey Pfeffer, *New Directions for Organization Theory: Problems and Prospects* (New York: Oxford University Press, 1997).
32. Gibson Burrell and Gareth Morgan, *Sociological Paradigms and Organizational Analysis* (New York: Routledge, 2016).
33. Thorsten Büschgens, Andreas Bausch and David B. Balkin, "Organizational Culture and Innovation: A Meta-Analytic Review," *Journal of Product Innovation Management* 30, no. 4 (April 2013): 763–781, https://doi.org/10.1111/jpim.12021.
34. Aimee Groth, "Zappos Has Quietly Backed Away From Holacracy," Quartz at Work, 29 January 2020, https://qz.com/work/1776841/zappos-has-quietly-backed-away-from-holacracy/.
35. Marvin E. Shaw, *Group Dynamics: The Psychology of Small Group Behavior* (New York: McGraw-Hill College, 1981).
36. Barney, "Organizational Culture."
37. Barney, "Organizational Culture."
38. S.A. Lippman and R.P. Rumelt, "Uncertain Imitability: An Analysis of Interfirm Differences in Efficiency under Competition," *The Bell Journal of Economics* 13, no. 2 (1982): 418–438; Barney, "Organizational Culture."
39. Lynne G. Zucker, "The Role of Institutionalization in Cultural Persistence," *American Sociological Review* 42, no. 5 (1977): 726–743.
40. Geert Hofstede, *Organizations and Cultures: Software of the Mind* (New York: McGraw Hill, 1991).
41. W. Edward Deming, *Out of the Crisis* (Cambridge: MIT Press, 2018).
42. J. C. Anderson, M. Rungtusanatham, R.G. Schroeder and S. Devaraj, "A Path Analytic Model of a Theory of Quality Management Underlying the Deming Management Method: Preliminary Empirical Findings," *Decision Sciences* 26, no. 5 (September 1995): 637–658, https://doi.org/10.1111/j.1540-5915.1995.tb01444.x.
43. Bahrami and Evans, "Super-Flexibility for Real-Time Adaptation."
44. Michael J. Glauser, "Upward Information Flow in Organizations: Review and Conceptual Analysis," *Human Relations* 37, no. 8 (August 1984): 613–643, https://doi.org/10.1177/001872678403700804.
45. Nemeth, "Managing Innovation."
46. Paul R. Lawrence and Jay W. Lorsch, *Organization and Environment* (Boston: Harvard Business Review Press, 1967).

47. Annika Steiber and Sverker Alänge, "A Corporate System for Continuous Innovation: The Case of Google Inc.," *European Journal of Innovation Management* 16, no. 2 (2013).

48. Tom Peters, "Tom Peters on Leading the 21st Century Organization," *McKinsey Quarterly* 3 (September 2014): 91.

49. Harry Boer and Frank Gertsen, "From Continuous Improvement to Continuous Innovation: A (Retro) Perspective," *International Journal of Technology Management* 26, no. 8 (January 2003): 805–827.

50. M.T. Hannan, "Rethinking Age Dependence in Organizational Mortality: Logical Formalizations," *American Journal of Sociology* 104, no. 1 (July 1998): 126–164, https://doi.org/10.1086/210004.

51. Carole Lalonde, "Managing Crises Through Organisational Development: A Conceptual Framework," *Disasters* 35, no. 2 (November 2010): 443–464, https://doi.org/10.1111/j.1467-7717.2010.01223.x.

52. Danny Miller, "Evolution and Revolution: A Quantum View of Structural Change in Organizations," *Journal of Management Studies* 19, no. 2 (April 1982): 131–151, https://doi.org/10.1111/j.1467-6486.1982.tb00064.x.

53. Dennis A. Gioia, Shubha Patvardhan, Aimee L. Hamilton and Kevin G. Corley, "Organizational Identity Formation and Change," *The Academy of Management Annals* 7, no. 1 (February 2013): 123–193, https://doi.org/10.1080/19416520.2013.762225.

54. Gabriele Jacobs, Arjen Van Witteloostuijn and Jochen Christe-Zeyse, "A Theoretical Framework of Organizational Change," *Journal of Organizational Change Management* 26, no. 5 (August 2013).

55. Rosabeth Moss Kanter, *Challenge of Organizational Change: How Companies Experience It and Leaders Guide It* (New York: Simon and Schuster, 2003).

56. Steiber and Alänge, "A Corporate System for Continuous Innovation."

57. Steiber and Alänge, "A Corporate System for Continuous Innovation."

58. Steiber and Alänge, "A Corporate System for Continuous Innovation."

59. Steiber and Alänge, "A Corporate System for Continuous Innovation."

60. Walter Isaacson, *Steve Jobs: The Exclusive Biography* (New York: Simon & Schuster, 2011), 340-352.

Part III: Advanced Concepts in Big Data

Chapter 7. Ethics and Big Data

Dr. Bohyun Kim

Ethical Questions Raised by Military Robots

Artificial intelligence (AI) is a research area where its military application is being actively pursued. On 12 February 2019, the U.S. Department of Defense released a summary and supplementary fact sheet of its AI strategy.[1] AI is used to generate intelligence from analyzing data in satellite imagery, terrain information, and data from multiple sensors by applying deep learning, statistical analysis, and probabilistic algorithms to such data. AI can also be used for designing targeted missiles, sophisticated weapons, and technology-intensive fighter planes.[2]

Many predict that developments in AI will lead to highly autonomous weapons. This trend is already seen in unmanned aerial vehicles (UAVs) and other military robots. UAVs such as the MQ-1 Predator, MQ-9 Reaper, and RQ-4 Global Hawk can track and/or attack targets from the air with laser-guided bombs. These UAVs are embodied and situated in the world, take sensory input from the environment, process it, and take action based upon the input and the rules in their programs (strike functions retain a human in the loop). For these reasons, they are often referred to as a "robot."[3] They are also cited as examples of what AI researchers call "an intelligent agent."

There are many different types of military robots. Some identify and dispose explosive devices. Others perform scouting tasks. Some military sentry robots are capable of automatic targeting and shooting. Automated defense systems, such as the Goalkeeper close-in weapon system and Aegis, protect military ships by automatically surveilling, detecting, and destroying incoming threats. Many military robots are remotely controlled by human operators. But some of them are capable in theory (if not practice) of engaging in military actions without the human operator being involved in the process.

Military robots have human benefits. They can significantly reduce the risk to soldiers in a military operation and even save their lives. Powered by sophisticated AI technology, capable military robots are likely to keep military personnel away from the battlefield and even replace them in the near future. Military drones, UAVs, and other robots functioning independently without a human operator having to determine and authorize each and every decision and action would bring vast savings in efficiency and cost. This is one of the reasons why military robots will become more autonomous and independent. The U.S. military considers robots to be a great asset. The Pentagon's spending on UAVs increased from approximately $300 million in the 1990s to $2 billion in 2005 and over $6 billion by 2011.[4]

However, autonomous military robots raise new and unsettling ethical questions. For example, a sentry robot cannot tell whether those crossing the border are unarmed civilians or enemy soldiers. Should the robot fire at those people? What happens if it turned out that those killed were unarmed civilians? Whether a military robot is to be given the authority to make its own decision or not is a controversial topic. If a military robot makes a mistake, who should be responsible for the mistake: a robot manufacturer that built the robot, engineers and computer programmers who designed and programmed the robot, the commander who deployed the robot, the soldier who was supposed to monitor the robot, or the robot itself?

One may believe that these are all merely theoretical questions. But with today's rapidly advancing AI technology and strong interests in its military use, these questions may become real issues that require a solution much sooner. That the U.S. is already planning to incorporate AI into its cyber defense systems is one indication of this.[5] Ethics is a discipline from which many seek insight in these matters. It is important to take a look at what moral philosophy has to offer regarding the morality of AI agents such as military robots.

The Trolley Problem

The trolley problem is a philosophical puzzle first introduced by Philippa Foot in 1967.[6] The recent development in a self-driving car brought new spotlight to this old philosophical problem. A runaway trolley barrels down a track where five unsuspecting people are standing. A bystander happens to be standing next to the lever that switches the trolley onto a different track when pulled. The other track, however, is not clear either. There is one person on it. Those who are on the either track will be killed if the trolley heads down that way. Should the bystander pull the lever to change the track for the trolley so that it would kill one person rather than five people?

Needless to say, in real life there will be little time for any deliberation. If one does not freeze and acts at all, it is likely to be more from a reflex than a conscious decision. With the autonomous car, however, the trolley problem becomes a matter of programming—a required decision-making process in advance. A machine can act much more quickly than a person. It will not panic or hesitate. It will simply follow and execute the given instruction. The engineers of an autonomous trolley now have an opportunity to program a moral action if an unfortunate case such as the trolley problem materializes. The question remains, however, should one act to make the autonomous trolley swerve or leave it to stay on course?

Moral philosophers have been discussing the trolley problem for a long time, and different moral theories take different positions. Utilitarianism, for example, argues that the utility of an action is what makes an action moral. That is, what generates the greatest amount of good is the most moral thing to do. Since five human lives are a greater good than one, one acts morally by pulling the lever and diverting the trolley to the other track. By contrast, deontology claims that what determines whether an action is morally right or wrong is not its utility but moral rules. If an action is in accordance with the rules of morality, the action is morally right. Otherwise, it is morally wrong. That one is not to kill another human being is one of those rules. Therefore, killing someone violates the rule and is morally wrong under all circumstances, including one in which it may result in saving more lives.

It appears that the utilitarian and the deontological position both appeal to and go against U.S. moral intuition in different aspects. On one hand, if a trolley cannot be stopped, is it not clearly better to choose saving five lives rather than only one? On the other hand, if killing people is wrong in the first place, how can one justify sacrificing someone's life even if it is to save five people? Isn't killing morally wrong no matter what?

At first glance, utilitarianism looks promising. But consider the case in which the bystander in the trolley case freezes and does nothing, thereby killing five people. Has the bystander committed a moral wrong? Originally, the question was whether one ought to switch the track of the runaway trolley or not. Now, a different but no less challenging question arises. Is a moral failure equivalent to a moral wrong? Suppose that one ought to act to maximize the greatest good. But how far should one go for that goal? For example, if one can lift up and throw

Isn't killing morally wrong no matter what?

a really large person onto the track to stop the trolley from running over the five men standing on the track, is this solution as morally permissible as pulling the lever since the result is the same as the loss of one human life?[7] Utilitarianism would count the outcome to be the same. But nearly no one—including those who argue that one should pull the lever to divert the trolley—would say that throwing a person onto the track to stop the trolley is a morally permissible act.

The problem with utilitarianism is that it treats the good as something inherently quantifiable, comparable, calculable, and additive. But not all considerations that people must factor into moral decision-making share those traits. What if the five people on the track are helpless babies or murderers who just escaped from prison? Would or should that affect our decision? For this reason, the utilitarian position is not necessarily the most persuasive view.

Deontology does not fare too well, either. Deontology emphasizes one's duty to observe moral rules. But what if those moral rules conflict with one another? Between not killing a person and saving lives, which one should trump the other? Conflict of values is common in life. Given this, deontology will have just as hard a time as utilitarianism in deciding what an intelligent agent is to do in a tricky situation such as the trolley problem.

Artificial Intelligence in a War and Its Moral Risk

Philosophical discussion around the trolley problem is likely to disappoint those looking for practical guidelines and solutions. Suppose that some engineers program an autonomous unmanned vehicle in a battlefield to always choose to do whatever maximizes the chances of victory, interpreting that as the utility in utilitarianism. That may include decisions such as sacrificing a great number of civilians when it can be avoided, which many would consider morally wrong. Now imagine that other engineers program the vehicle to act to minimize the number of casualties at all costs. This will not be always strategically beneficial to win a war.

If the engineers ask military commanders what to do, what should they choose? While these are certainly unrealistic simplifications, it is clear that moral philosophy does not provide an easy answer. However, this should not lead anyone to dismiss ethical issues related to AI-powered military robots. AI-powered military robots are not just another kind of weapon. When widely deployed, they can change the nature of war. Below are some of the prominent ethical issues that autonomous military robots present.

a. There are already many remote-operated weapons, but AI-powered military robots go one step further. They can identify a target and initiate an attack on their own. Due to their autonomy, military robots can significantly increase the distance between one who decides and acts to kill and the other who gets killed.[8] This increase, however, may lead one to surrender one's own moral responsibility to a machine resulting in the loss of humanity in a war, and this is a serious moral risk.[9] The more autonomous military robots become, the less responsibility for taking life-and-death decisions comes to rest with humans.

b. Furthermore, as military robots make killing easier, small conflicts may more quickly escalate to war. The side that deploys AI-powered military robots is likely to suffer many fewer casualties for itself while inflicting many more casualties on the enemy side. With this, the military may become more inclined to start a war. Ironically, when everyone thinks and acts this way, the number of wars and the amount of violence and destruction in the world will only increase overall.[10]

c. Another issue with the military robot is that it may fail to distinguish between combatants and innocents. In that case, the moral problem this creates becomes two fold. First, is it justifiable to let robots take the lives of other human beings? Second, can we risk robots killing innocents? Some would argue that only people should decide to kill other people, not machines.[11]

d. Lastly and as noted in chapter two, many modern AI systems use machine learning (ML) techniques that generate algorithms on their own from a large amount of data without being given a set of preprogrammed rules. This contrasts with a more traditional, symbolic AI system in which a set of explicit rules are programmed into a machine. Unlike those explicit rules in a symbolic AI system, the process that a ML system goes through to reach a conclusion from data is opaque to humans, including to the designers of the system itself.[12] For example, a ML system may generate an algorithm that successfully recognizes a cat in a photo after going through a large number of photos containing cats from a variety of angles and in many different postures. But the

> resulting algorithm, a complex mathematical formula which identifies a cat in a photo, is not something that humans can easily decipher. The algorithm is trusted based upon the accuracy of its predictions. But why and how it exactly works remains a mystery, and even the designers of such an AI system are unable to supply an explanation.[13] This means that reasons behind some of the decisions made by military robots and AI systems may not be easily explained and that the military decision-making involving those military robots and AI systems may lack sufficient justification as a result. Considering the gravity of these decisions, the lack of explainability and justification is a serious issue.

Given these issues, some may argue that in all actions by AI-powered robots, humans must be kept in the loop in order to prevent unexpected behavior and to ensure that all life-or-death decisions still rest with humans. A "human-in-the-loop" AI system is autonomous up to the point of selecting a target and even action to take. But a human-in-the-loop system will execute the action only when a human confirms it. By contrast, a "human-on-the-loop" system acts on its own unless it is overridden by a human.[14] A "human-off -the-loop" is most autonomous. It does not require any confirmation to engage, and its action cannot be aborted once activated.

The idea of making all AI systems human-in-the-loop or human-on-the-loop, however, does not directly address or resolve the moral issues outlined above. This is because moral decisions are something that humans struggle with without clear answers. Machines can behave as ethically as humans manage to do so. In this sense, the moral challenge raised by military robots and other ML-driven AI systems is more to do with how humans should act rather than machines.

Military Decision-Making with Artificial Intelligence

Ethicists pursue generalizable abstract principles. For this reason, they are interested in borderline cases that reveal subtle differences in varying moral theories. Their goal is to define what is moral and investigate how moral reasoning works. By contrast, engineers desire practical solutions to real-life problems and look for guidelines that will help with implementing those solutions. Their focus is on creating a set of constraints and if-then statements, which will allow a machine to identify and process morally relevant considerations so that it can determine and execute an action.

On the other hand, the ultimate goal for commanders and soldiers is to end a conflict, bring peace, and facilitate restoring or establishing universally recognized human values such as freedom, equality, justice, and self-determination. In order to achieve this goal, they are tasked with making the best strategic decisions and taking the most appropriate actions in a battlefield. In deciding on those actions, they are also responsible for abiding by moral codes and not abdicating their moral responsibility, protecting civilians, and minimizing harm, violence, and destruction as much as possible.

This chapter has discussed some of the ethical issues related to the use of military robots and other autonomous AI systems. Unlike ethicists and engineers, commanders and soldiers often face complicated situations where they must decide and act quickly with potentially life-or-death consequences. The complications that they face will only increase as more military robots and other AI systems are adopted.

Will military robots and other AI systems be able to help untangle some of those complications, or will they only add more difficulty to military decisions? Before selecting and adopting military robots or other AI systems, all military decision makers will need to consider if those robots and other AI systems follow the values that the military abides by in doing their job and further help commanders and soldiers to do the same. There are already several ethical guidelines for AI development and use.[15] But guidelines are only recommendations for voluntary adoption at varying levels, if adopted at all. To ensure the safe and ethical use of AI, the military will also need to start developing an appropriate and enforceable regulatory framework for the military application of AI.[16]

Endnotes

1. Terri Cronk, "DOD Unveils Its Artificial Intelligence Strategy," U.S. Department of Defense, 12 February 2019, https://dod.defense.gov/News/Article/Article/1755942/dod-unveils-its-artificial-intelligence-strategy/.
2. For more examples, see Paul Scharre, *Army of None: Autonomous Weapons and the Future of War* (New York: W. W. Norton & Company, 2018); Naveen Joshi, "4 Ways Global Defense Forces Use AI," *Forbes*, 26 August 2018, https://www.forbes.com/sites/cognitiveworld/2018/08/26/4-ways-the-global-defense-forces-are-using-ai/.
3. Interestingly, there is not one widely agreed-upon definition of a robot even among roboticists. One working definition of a robot by George Bekey is that a robot is a machine situated in the world that senses, thinks, and acts. See George

Bekey, "Current Trends in Robotics: Technology and Ethics," in *Robot Ethics: The Ethical and Social Implications of Robotics*, eds. Patrick Lin, Keith Abney, and George Bekey (Cambridge: MIT Press, 2012), 18.

4. Scharre, *Army of None*, 14.

5. U.S. Defense Science Board, *Summer Study on Autonomy* (U.S. Department of Defense, 2016) quoted in Mariarosaria Taddeo and Luciano Floridi, "Regulate Artificial Intelligence to Avert Cyber Arms Race," *Nature* 556, no. 7701 (April 2018): 296, https://doi.org/10.1038/d41586-018-04602-6.

6. Philippa Foot, "The Problem of Abortion and the Doctrine of Double Effect," *Oxford Review*, no. 5 (1967): 5–15.

7. Judith Jarvis Thomson, "Killing, Letting Die, and the Trolley Problem," *The Monist* 59, no. 2 (April 1976): 204–217.

8. Noel Sharkey, "Killing Made Easy: From Joysticks to Politics," in *Robot Ethics: The Ethical and Social Implications of Robotics*, eds. Patrick Lin, Keith Abney, and George A. Bekey (Cambridge: MIT Press, 2012), 111–128.

9. Daniel L. Davis, "Who Decides: Man or Machine?" *Armed Forces Journal*, 1 November 2007, http://armedforcesjournal.com/who-decides-man-or-machine/.

10. Kahn also argues that the resulting wars increased in number by the use of military robots will be morally bad. See Leonard Kahn, "Military Robots and the Likelihood of Armed Combat," in *Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence*, eds. Patrick Lin, Keith Abney, and Ryan Jenkins (New York: Oxford University Press, 2017), 274–292.

11. Davis, "Who Decides."

12. Will Knight, "The Dark Secret at the Heart of AI," *MIT Technology Review*, 11 April 2017, https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/.

13. This black-box nature of AI systems powered by machine learning raised great concern among many AI researchers in recent years. This is problematic in all areas where these AI systems are used for decision-making, not just in military operations. The gravity of decisions made in a military operation makes this problem even more troublesome. Fortunately, some AI researchers, including those in the U.S. Department of Defense, are actively working to make AI systems explainable. But until such research bears fruit and AI systems become fully explainable, their military use means accepting many unknown variables and unforeseeable consequences. See David Gunning, "Explainable Artificial Intelligence," *DARPA*, accessed 28 May 2019, https://www.darpa.mil/program/explainable-artificial-intelligence.

14. Regarding how the "human-in-the-loop" and the "human-on-the-loop" system work differently in AI weapons, see Sharkey, *Killing Made Easy*, 110–111.

15. See "Ethics Guidelines for Trustworthy AI," European Commission, 8 April 2019, https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai; "The Toronto Declaration: Protecting the Rights to Equality and

Non-Discrimination in Machine Learning Systems," *Access Now*, 16 May 2018, https://www.accessnow.org/the-toronto-declaration-protecting-the-rights-to-equality-and-non-discrimination-in-machine-learning-systems/.

16. For related discussions, see Mariarosaria Taddeo and Luciano Floridi, "Regulate Artificial Intelligence to Avert Cyber Arms Race," *Nature* 556, no. 7701 (April 2018): 296–298, https://doi.org/10.1038/d41586-018-04602-6; Alison Berthet, "Why Do Emerging AI Guidelines Emphasize 'Ethics' over Human Rights?" OpenGlobalRights, 10 July 2019, https://www.openglobalrights.org/why-do-emerging-ai-guidelines-emphasize-ethics-over-human-rights/?lang=English.

Chapter 8. The Future of Systems Architecture

Dr. Justin F. Brunelle, Mr. David Bryson, Dr. D.J. Shyy, Dr. Yaakov Weinstein

MITRE: APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED. CASE NUMBER 20-2282.

This technical data was produced for the U. S. Government under Contract No. FA8702-19-C-0001 and is subject to the Rights in Technical Data-Noncommercial Items Clause DFARS 252.227-7013 (FEB 2014).

U.S. Special Operations Command has pushed out on developing the hyper-enabled operator (HEO) concept, which seeks to link the war fighter to the internet of things (IoT) to achieve cognitive overmatch "at the edge."[1] The underlying assumption is that the war fighter will be able to make more informed decisions with the ability to locally assess or engage reach-back support to understand the operating environment through man-portable computing and communications technology.[2] While technically feasible, there are many challenges associated with underlying assumptions. First, for the system to work effectively, information must be easily queried, but current collection, storage, and dissemination practices are not uniform. Overcoming this basic infrastructure and architecture issue is a prerequisite for the HEO to work. Second, the bandwidth necessary to support the HEO is also problematic. Generating secure communications depends on a number of factors, some of which might be cost prohibitive. Clearly understanding the factors surrounding localized bandwidth is a necessity in determining when and how the HEO should be employed. Alternatively, it is important to note that a HEO could also be a hyper-surveilled operator due to the signatures generated from the capabilities. And third, the HEO concept presumes security through encrypted communications.

The sections in this chapter discuss each of the three challenges. The first section discusses the current state of the art in cloud computing, the issues surrounding access permissions, and the basics of blockchain technology as a mechanism of encrypted data transfer. The second section reviews the considerations associated with communication "at the edge." Mobile and wireless networks have long existed for citizen and government use but are more nuanced when deployed for tactical (i.e., outside the continental U.S. [OCONUS]) applications. There are significant cost, logistics, and risk calculations that go along with enabling the IoT down range. The third and final section reviews the basics of quantum computing—a disruptive technology that could render contemporary encryption extremely vulnerable and make the HEO concept untenable.

The Current State of the Art

While the government frequently lags behind industry in adopting emerging technologies, a variety of technologies that are well-established in commercial industry are being adopted with increasing success by practitioners in the government domain. This section investigates cloud computing as a technology for enabling high-volume storage and computational processing, as well as enhancing the agility of government information technology (IT) services. Blockchain is famous due to its cryptocurrency applications, and it is also being considered for a variety of government applications.

The Cloud—Myths and Realities

Cloud computing traditionally has a large number of myths and realities regarding the technologies associated with its use, particularly in the government. This section provides a broad introduction to the topic of cloud computing and then addresses a few common myths about it.

Realities. In non-technical terms, *cloud* computing is a term that refers to a method of delivering technical services from a pool of resources that are aggregated, scaled, and utilized to match a specified need. A *cloud* is a series of connected computers utilizing software that enables services to be executed more efficiently and reliably than using a single computer; using a cloud is similar to using someone else's bigger computer to accomplish a task. For example, Gmail[3] and Google Docs[4] are examples of programs—hosted on the Google computers—that store data in a cloud environment and only deliver data to the users' computers upon request.

More technically, cloud computing refers to the infrastructure for elastic, on-demand provisioning of storage, computation, and software service delivery from a shared set of resources. When recalling the non-technical description, it may be similar to using someone else's bigger computer, but in reality, the size and location of the resources being used are hidden from the user. More formally, the National Institute of Standards and Technology (NIST) definition of cloud computing is "a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction."[5]

The 2011 NIST report also establishes five essential characteristics of cloud computing:

> **On-demand self-service.** Cloud users can request computing resources from the larger pool of cloud resources without human intervention.
>
> **Broad network access.** Cloud resources are accessible through reliable networks, both from thick clients (e.g., servers) and thin clients (e.g., mobile web browsers).
>
> **Resource pooling.** A cloud's resources are pooled for use by multiple cloud users (or customers) and are dynamically allocated based on requests and workloads; resources that are no longer required are returned to the resource pool.
>
> **Rapid elasticity.** Resources and services can be scaled up when required and scaled down as workloads reduce.
>
> **Measured service.** Cloud users are charged by the cloud provider in accordance with the level of usage of the resources; users that use less will be charged less and users that use more will be charged more.[6]

In short, cloud architectures should provide the ability to rapidly provision, access, expand, and release computation resources. The NIST cloud service models considered in this work include infrastructure as a service (IaaS), software as a service (SaaS), and platform as a service (PaaS). The following are descriptions of each service model:

> **Infrastructure as a service.** The cloud service provides users the option to establish a computing infrastructure that includes the operating system, storage, processing features, etc., with which they can instantiate their services. This is closest to operating a data center but one in which the cloud provider owns the hardware.

Platform as a service. The cloud user provisions a computing infrastructure (e.g., operating system, storage, and computation) from the cloud provider and deploys user-owned applications onto the provisioned platform.

Software as a service. The cloud user provisions applications running on a cloud platform with the application being owned by the cloud provider and accessed by the cloud user(s) through a web browser or other thin client device.[7]

NIST also lists four deployment models:

Private cloud. A cloud infrastructure provisioned, managed, and owned within an organization.

Community cloud. A cloud infrastructure provisioned and owned by related organizations with common use cases and requirements (e.g., common security needs or common access protocols).

Public cloud. A cloud owned by an organization (often a commercial company) that provisions services to the general public.

Hybrid cloud. The cloud infrastructure comprised of more than one of the above cloud types that can be bound together by policy and technology to enable their operation.[8]

Common services offered by cloud models include elastic storage (storage dynamically provisioned based on user need), high performance or parallel computing (leveraging multiple computing resources to increase processing efficiency), and robust software services (software provided to an expanding and contracting user base). Services such as these allow elastic information storage, processing, and distribution. Cloud technologies allow consumers of cloud services to request resources as needed from a larger pool of resources and release those resources after task completion. These interactions are fulfilled automatically by the cloud and (in commercial cloud environments) users are often charged in accordance with their consumption.

To users, cloud services appear to be centrally located geographically and architecturally. In reality, cloud computing allows the services being consumed by the user to be geographically, physically, and architecturally distributed while allowing a single point of user interaction. For example, when accessing elastic storage services in a cloud environment, the data may be distributed over multiple data stores in multiple countries or even continents while maintaining the appearance of being consolidated to a single data store at a single location. The distributed nature of the cloud nodes allows rapid, seamless failure recovery, as well. In short, clouds allow resources to be physically distributed while being logically consolidated.

With cloud computing in the U.S. Government (USG) comes several perennial challenges:[9]

Security. Third party control of government data in non-government systems is a primary cloud computing challenge mitigated by the advances of the Federal Risk and Authorization Management Program,[10] securing cloud access points, and utilizing government-tailored cloud services. Outstanding challenges exist for insider threats, encryption, vendor lock-in, and mitigating attack surfaces.

Acquisition. Government acquisition processes are well suited for enterprise-wide, physical systems but not well-suited for metered utilities and services such as a cloud computing service. Government practices are shifting to adopt more agile practices for acquiring metered utilities and services.

Governance. Identifying the roles, responsibilities, and program management of incorporating cloud services conflicts with traditional government processes. Identifying the systems to migrate to a cloud (independent of migration complexity), the organizations or other groups that own and pay for migration and maintenance, and the process of managing a potentially multi-year development effort pose challenges to typical government cloud programs.

Cultural resistance. Issues of owner and operation responsibilities shift and cloud computing disrupts traditional government cultures of data and process ownership. Adopting proper development operations practices, utilizing brokers, and educating leadership are mitigations for the cultural challenges of adopting cloud computing.

Cost management. Measuring, monitoring, and restricting growth of cloud operations costs is a concern that can be mitigated with cloud monitoring tools, cost estimation techniques, service-level agreements (SLAs) with cloud providers, or reliance on a cloud broker.

Migration complexity.[11] Not all applications are cloud-ready and must be modified for migration purposes. This incurs cost, time, and expertise that must be invested to perform an effective cloud migration.

Cloud services are routinely used to process large datasets due to the ability to scale up. Technologies, such as Hadoop,[12] enable suitable datasets to be split between computational nodes (e.g., machines) and processed in parallel, making the data processing that would otherwise take hours or days possible in minutes. Because of cloud services' ability to scale, operate in parallel, and provide increased computational resources, they are extremely useful in big data processing.

Myths

From this broad introduction to the cloud computing domain, we can dispel a few of the common myths regarding cloud computing. These myths are not unique to the USG but are applicable and represent some of the common misconceptions held by some government consumers.

Myth 1: Projects in need of information technology must adopt the mindset of cloud first. Successful cloud practitioners within the government recommend that prospective cloud adopters consider the business and technical needs and value of cloud services rather than strict adherence to "cloud-first"[13] mandates. There are some applications and missions that are not well suited to leveraging or being hosted in a cloud. While it is necessary to consider cloud services as part of the data and hosting service options, it is important for cloud adopters and IT managers to consider the requirements of their effort and the associated suitability of cloud services for the mission.[14]

Myth 2: Cloud is cheaper than stand-alone data centers. Cloud is not always cheaper than a private or stand-alone data center. There are a variety of factors that influence the cost of operating in a cloud environment such as the utilization, rate of data moving into and out of the environment, and provisioning of the applications. It is sometimes cheaper to operate in a data center (e.g., when an application has a very steady and well-known operating load without a need to scale up for surge handling) as opposed to a cloud; a cloud would be better suited for an application with widely varying utilizations and loads to take advantage of scaling within the cloud. There are a variety of cost estimation tools that can help navigate the factors and compare the cost of operating in a cloud versus a data center both in industry and government.[15] Regardless of cost considerations, cloud computing offers other benefits that may outweigh cost impacts. For example, programs may become more flexible in their operation, more efficient in their computation, or more accessible across an enterprise. Cloud adopters are often encouraged to consider aspects of cloud benefits beyond cost when considering a migration or other cloud adoption.

Myth 3: Cloud migration is simple; "lift-and-shift" is suitable for legacy-to-cloud migrations. "Lift-and-shift" refers to the act of taking virtualized applications and moving them from a legacy data center to a cloud environment without further modification. Cloud practitioners have cited that this practice is useful for cloud adopters early in their cloud lifecycle,[16] but that these applications (sometimes referred to as forklifted) do not make use of the benefits of operating in a cloud environment, such as the ability to seamlessly scale or make use of distributed storage. Often, applications must be refactored to be optimized for cloud. In summary, lift-and-shift is beneficial for cloud novice organizations to refine their cloud policies and practices, but applications often must be refactored to make use of cloud features.

Cloud Is Less Secure Than Private Data Centers

Early in the U.S. Federal Government's venture into cloud adoption, security was a major concern. While these concerns still exist,[17] many organizations have cited improved security after moving to a cloud from a data center. In many cases, the security risks in a cloud are the same as those faced by traditional IT solutions. However, government organizations and the commercial cloud providers they use are more frequently establishing agreements to share the responsibility for safeguarding applications and monitoring data and application security. Often, the security of the cloud solution is not less than that of a traditional data center but rather an issue with the ability of cloud adopters to maintain awareness of security procedures and risk mitigations.[18] While cloud solutions may be less secure in some cases, it is a fallacy to say—ubiquitously—that clouds are less secure than data centers.

With an understanding of distributed and large-scale computing established in this section, a discussion can begin about technology in which government is interested based on its promise of decentralized authority—blockchain.

Blockchain

Blockchain technology has been successfully implemented outside of government, most famously as the technical underpinning of popular cryptocurrencies such as Bitcoin.[19] Blockchain is the foundational technology enabling cryptocurrencies used by the public.[20] A blockchain allows a decentralized network of nodes (i.e., connected computers) to agree upon a record of transactions. As a result, a blockchain can provide a method of decentralized trust in a peer-to-peer network. The nodes within the network work to synchronize the state and record of the blockchain transactions. Further, the nodes each keep a replicated copy of the blockchain state, adding to resilience in the event of node failure or attack. This allows applications—such as those in cryptocurrencies—to succeed in a reliable fashion despite lacking centralized control.

Given the high-profile successes of blockchain-based cryptocurrencies, USG organizations—along with private industry and open-source communities—are investigating the applicability of blockchain technologies for government use cases (e.g., implementing decentralized trust and authority). Blockchain adoption within the government is just now emerging and less mature than cloud technology due in part to the relative newness of the technology. This section describes the foundational concepts behind blockchain technologies and discusses the details, opportunities, and potential challenges of government adoption of blockchain.

A blockchain is frequently compared to a ledger of transactions similar to those used at traditional banks. For example, each transaction entry in a cryptocurrency blockchain provides information regarding the exchange of a digital asset. For a supply chain blockchain, each transaction entry may have information about tracked objects such as the location and custody of shipping containers. These blockchain transactions can be considered the fundamental unit of work in a blockchain. The transactions are permanently recorded, synchronized, verified by every node in the peer-to-peer network and cryptographically bound into tamper-resistant blocks replicated to each node. This distributed synchronization among nodes, redundancy, replication, and cryptographic bound blocks provides significant protection against tampering.

Blockchain technology relies on two cryptographic primitives to secure the contents of the ledger—cryptographic hash functions (e.g., SHA-256) and digital signatures. Cryptographic hashes give nodes in the peer-to-peer network the ability to efficiently detect changes to the ledger. Digital signatures are used to verify the cryptographic authenticity of transactions submitted to the network and, in some cases (depending on the type of blockchain), to verify agreement on a block of transactions. The combination of the two, along with the consensus algorithm, form the basis for providing the tamper-resistance and security properties of blockchain and its ability to operate across a trustless network.

This distributed synchronization among nodes, redundancy, replication, and cryptographic bound blocks provides significant protection against tampering.

Public and Permissioned Blockchains

There are two kinds of blockchains depending on the needs of the blockchain stakeholder community—public or permissioned. Public blockchains such as Bitcoin and Ethereum[21] operate in a completely trustless stakeholder environment in which anyone in the world can participate in the blockchain. To operate in this type of environment, public blockchains use a cryptocurrency to enhance security through a combination of game theory[22] and economic incentives. On the other hand, permissioned blockchains are more restrictive and designed to operate across a pre-defined group in which the group decides who can participate in the network. Within a permissioned blockchain there is a stronger sense of identity and control and therefore no need to use a cryptocurrency as way of providing economic incentives to help secure the network. Interest in blockchain technology across the USG tends toward permissioned blockchain due to the level of control over membership achievable by the participating organizations. However, permissioned blockchain implementations require increased planning and governance to establish the network. Failure to properly implement these aspects could lead to a less secure solution.

Benefits, Tradeoffs, and Considerations

Adopting blockchain for the government has a variety of benefits and tradeoffs. The primary interest from government adopters is the ability to establish trust in a decentralized and trustless environment. Similarly, the ability to share information between organizational boundaries has benefits, particularly when no existing mechanisms to establish trust between organizations has been established. Even when mechanisms to establish trust between organizations exist, the mechanisms usually take the form of expensive federation and/or centralized capabilities. Despite the benefits, common barriers to blockchain adoption within the government exist, such as the need for privacy and confidentiality on the blockchain, transaction scalability, and blockchain-to-blockchain connectivity. However, the high level of interest in the technology is driving research to address the gaps.

Government adopters have several aspects to consider prior to implementing blockchain to solve a challenge.[23]

Transaction throughput. The transaction throughput of current blockchain implementations vary widely. For example, the Ethereum public blockchain averages approximately 12-15 transactions per second.[24] A permissioned blockchain can potentially achieve thousands of transactions per second[25] depending on the application and platform choice.

Information privacy. Information stored on a blockchain is not private. This is by design to provide auditability. Government adopters must consider the privacy and security impacts of adopting blockchain.

Existing architectures. Many current methodologies and processes are designed around a centralized architecture making them incompatible with a decentralized blockchain. Adapting existing business processes to blockchain technology will—in most cases—require redesign of the process to take advantage of the technology.

Algorithms and security. There are a variety of consensus algorithms used by permissioned blockchain implementations, and not all may provide the same level of security. Adopters should consult blockchain experts prior to selecting an algorithm or blockchain platform.

Adoption preparation. Governing a permissioned blockchain requires planning (e.g., selecting validator/consensus nodes in the network). Failure of blockchain adopters to carefully plan and prepare for a permissioned blockchain could negate some of the security properties of the blockchain.

Blockchain type selection: When selecting the type of blockchain (i.e., public versus permissioned), adopters should examine several aspects of their challenge domain:

- What level of trust exists between participants?
- Is there a need for a cryptocurrency?
- Is there a need to control who can participate in the network?
- Are there data privacy concerns?
- Will control of smart contracts be limited to a subset of the participants?

The Potential Impact of Blockchain for Government

While blockchain technologies have been implemented in the private sector for trusted transactions, there is high interest but few successful implementations of blockchain for data management. A successful implementation for blockchain-enabled data management is likely coming in the near future and may potentially involve storing data on blockchains or using blockchain for linking/referencing data external to blockchain itself. In the future, blockchain technologies may be able to help assert trust for datasets that are maintained in a distributed environment, but care should be taken to select the appropriate implementation; this is performed through blockchain's ability to provide an audit trail of transaction activity on data.

With increased emphasis on organizations' agility and interoperability, the ability to establish trust in a decentralized manner for the exchange of data is useful. Blockchain technologies can help establish—in a decentralized fashion—mutual trust of data without surrendering data (and associated control) to a centralized party. For scenarios in which all participating parties are fully trusted, blockchain technologies are not needed; in scenarios in which parties are not inherently trusted (either from a technical or procedural standpoint), blockchain can help establish the trust.

With the increased interest in blockchain technologies from the Federal Government, it is critically important to understand the fundamentals, usage, and concerns with adopting blockchain. Selecting the appropriate implementation, model, and algorithm to be used in the blockchain system is essential to government adoption. Selecting the appropriate blockchain features can provide the ability to operate trusted transactions in a trustless environment. Based on the current state of the art, permissioned blockchain solutions are most often best suited to government use cases. The government should use care to provide proper planning and preparations when establishing a permissioned blockchain system. Additionally, with the increased research, development, and emphasis on blockchain throughout the industry, government adopters should use care to monitor the state of the art and evolutions with blockchain applications.

Bandwidth and Mobile Network Access—CONUS and OCONUS

Data exchange is an aspect of big data that is taken for granted in traditional cloud computing environments. Often, network connectivity, bandwidth, and network availability are challenged within environments where data collection is taking place. That data must be exchanged across a network to enable decision-making and integration of the collected data with other datasets. Information derived from collected and fused data must be delivered to operators and decision makers in potentially disconnected environments. Because of these constraints and challenges (coupled with the rate at which the size of data is increasing), bandwidth in various environments greatly impacts how data can be leveraged by those that require information.

It is increasingly common for data to be collected and exchanged via mobile devices and mobile networks. As a major contributing aspect of mobile data transmission, this section provides an overview of long-term evolution (LTE)[26] deployment strategy for CONUS and OCONUS as well as the LTE spectrum strategy that should be considered.

There are four use cases the government deploys or makes use of: mobile networks enterprise, tactical, special mission, and limited access.

1. The enterprise use case involves acquiring cellular wireless services from a LTE provider for government personnel and paying monthly subscription fees for a data plan. There is also an upfront cost of purchasing the mobile devices.

2. The tactical use case is to build out infrastructure (including both radio access network (RAN) and core network) for cellular mobile networks, typically for OCONUS deployments. The RAN consists of base stations (eNBs in LTE terminology) as well as their supporting

infrastructure (e.g., cell towers). Since the government does not own spectrum overseas, agreements with host nations must be in place first before the network can be built.

3. The special mission use case is for the government to conduct a special mission utilizing mobile networks within CONUS in which the government owns the network and spectrum. One of the typical use cases is for government organizations to build their own infrastructure for mobile networks with their own spectrum for video surveillance purposes. Another example is the 700 MHz public safety networks also known as First Responder Network.[27]

4. The limited access use case can be considered as a subset of the tactical use case. Limited access deployments would be in remote locations without access to typical LTE infrastructure such as cabling and fiber. In this case, the government can use deployable cellular system in a box (which contain both base station and the core) or use host nation cellular infrastructure.

For tactical, special mission, and limited access use cases, the government owns and builds the cellular infrastructure including the cell towers. These use cases are sometimes referred to as bring-your-own-network (BYON). BYON would involve a temporary network deployment if a fixed solution is not feasible. The primary motivators for the government to adopt this type of deployment are because it provides faster build-out, better performance, higher security, and no subscription cost to meet the requirements of special missions. For a special mission use case, the government also owns the spectrum over which the network operates, providing more security over the air.

Enterprise Environment. In an enterprise use case environment, there are two procurement models to consider: Government purchased services from commercial providers and mobile provisioned infrastructure.

Mobile Virtual Network Operator. This deployment model is the most commonly used among Department of Defense (DOD) departments for the enterprise service use case. LTE carriers provide RANs and core networks and perform all services.

The benefit of the government purchased services model is that there are no separate costs for infrastructure, maintenance, deployment, and personnel. Unfortunately, in this scenario, costs are determined by carriers based on an SLA, and the government cannot control the technology used, leading to concerns about security and quality of service (QoS).

A mobile virtual network operator (MVNO) does not own any RANs; however, an MVNO may own the core network for applications as well as subscriber identity module (SIM) card provisioning and billing. An MVNO will purchase the bulk of airtime minutes for voice and gigabytes for data at a discounted price. The MVNO then resells the services to consumers. The profit margins for MVNOs are typically low since it is a very competitive space.

There are a number of benefits with this deployment model: the cost of infrastructure is not the government's responsibility, the government manages and determines the data plan, and the ability of an MVNO to cover the whole DOD has the potential of reducing cost compared to each DOD department purchasing their own service plan. On the other hand, the risk with this model is that the government might not use all the data it has purchased in bulk leading to an overpayment for the service. Moreover, the government has no control over the cellular infrastructure, meaning it cannot control the technology used or security features, does not determine when service would be terminated for 3G/4G, cannot determine OCONUS coverage, and needs to have agreements with allies (which is important for the tactical use case).

Tactical Environment: Government-Owned Cell Tower. The tactical use case involves the government deploying a government-owned cell tower to create a private network for tactical operation. These are land-based 4G systems that consist of mobile stations, a point-to-multipoint (PMP) base station, and a point-to-point (PTP) station for backhaul services. LTE backhaul refers to the connection from the core network to the base stations. The base station is referred to as a network in a box (i.e., it has both eNB and the core network). PTP backhaul services are provided with wireless systems such as microwave, satellite communication, or landline systems (e.g., fiber or cable). PTP backhaul services are utilized for connecting 4G systems to the host central management system or connecting to two or more PMP base station towers.

The benefits of this deployment model are that the government controls everything including the schedule, technology, SLA, spectrum, security provisioning, and QoS provisioning, and there are no subscription fees. A disadvantage with this option is that the government increases its threat area, meaning there is no layered defense (i.e., the public cellular wireless network cannot be used as a buffer). Additionally, the overall cost of building out the infrastructures of cellular wireless networks, which include capital expenditure and operating expense, is significantly higher than acquiring services from LTE carriers.

Consideration for Long-Term Evolution Spectrum Strategy. LTE can be utilized by the DOD in many situations, but there are still considerations that need to be made to provide the most efficient deployment of the technology. This section discusses the factors in LTE deployments for CONUS and OCONUS. In the U.S., there can be more licensing issues and less available spectrum to use, whereas overseas, the U.S. must be more cautious to not infringe upon other countries' use of their spectrum. The three main types of spectrum usage that can be leveraged for both OCONUS and CONUS are licensed, unlicensed, and shared.

Licensed Spectrum. There are two types of licensed spectrum: one owned by cellular service providers and the other by the government. Both types would provide exclusive control over the spectrum band, but this would mean that the government should purchase or acquire a lease of the license. In the U.S. and in many other countries, there is little overlap between what is licensed to federal and commercial users. Deploying LTE in the federal bands may require custom base stations and mobile devices because the frequency range may not be a commonly used by a commercial 3rd Generation Partnership Project LTE band. An example in the U.S. lies in 4.4–5.0 GHz. One consequence of this would be the increase of overall deployment cost due to the higher prices of custom base stations and mobile devices.

> *The three main types of spectrum usage that can be leveraged for both OCONUS and CONUS are licensed, unlicensed, and shared.*

Unlicensed Spectrum. The unlicensed spectrum is located around 5.15–5.925 GHz. A user does not need a license to transmit within this spectrum band, but the device transmit power must follow federal and international regulations. There are also a few challenges to operating in unlicensed bands because no users have explicit priority. This means that users who are in that spectrum will have to co-exist with other users. Numerous techniques are being implemented to mitigate this and allow all users to fairly share the spectrum. One way to try to prevent interference is to limit the transmitting power of equipment. The Federal Communication Commission in CONUS may restrain the power limits depending on the region and the band. An operator's handset is typically not affected; however, there can be some limits on the base station. OCONUS would follow similar rules for unlicensed bands.

Shared Spectrum. Shared spectrum is mutually shared licensed or unlicensed spectrum between two or more parties. These parties may or may not have exclusive access to the band, but they must be able to deal with other incumbents within the band. The Citizens Broadcast Radio Service band in 3.5 GHz is utilizing shared spectrum. In the U.S., incumbent users that use this band for military radar will have top priority for the band. Secondary users will have to purchase priority access licenses (PALs) to gain access in a particular region for a temporary period of time. Finally, there are also general authorized access users that can use this spectrum on a non-interfering basis with incumbents. The DOD would require a lease from PAL users to use this band for mobile deployments.

Continental United States Long-Term Evolution Deployments. Considering the discussions in the prior sections, it is possible to discuss the LTE deployments in CONUS and OCONUS and how they differ. In both CONUS and OCONUS deployments, there are public and private LTE networks that operators can leverage.

Public Long-Term Evolution Networks. Due to the vast number of commercial LTE deployments, the DOD can leverage commercial networks to perform activities such as training and coverage for military bases. The DOD can negotiate with the LTE carriers before deployment to allow access for government use. There are risks to this approach because it is based on a commercial network; this means that the information will be harder to secure. Another consideration is that because the DOD is operating on a commercial network, resources will need to be shared with public users. This may mean that service could be impacted, and reliability cannot be guaranteed. This could cause issues if security and redundancy are the top priorities.

Spectrum purchases will allow government access to bands and therefore access to the network, but it will require more money to acquire this spectrum. The LTE acquisitions should be done at the command level (or higher) rather than per base or per unit. There also cannot be any bias by the government based on carrier. The carriers will need to be chosen based on what spectrum is needed, but the actual deployment should be operated by the government and remain carrier neutral. If there is a need for LTE infrastructure on base or within a command's zone, these carriers should be charged a fee.

Private Long-Term Evolution Networks. If the DOD were to deploy private networks, there would be more security and redundancy in the network. The DOD would have complete control over the network so it could choose the parameters with which it would operate. This option provides a secure network but at an increased cost due to hardware and operations costs. The acquisition would be on a drastically larger timeline due to the initial setup of the network. Base stations and an LTE core would need to be acquired, deployed, operated, and maintained by the government. The steps involved could make this setup lengthy and logistically challenging.

The advantage of this network, along with security, would be that the base stations could be configured to operate on military licensed bands and have no need for commercial spectrum. This would save the government money in negotiating commercial spectrum access. The base stations could also choose to augment these bands with unlicensed spectrum, which would increase data rates and provide another lower-cost spectrum solution.

Outside the Continental United States Long-Term Evolution Deployments. Similar to CONUS deployments, OCONUS deployments can use public and private LTE networks, each with tradeoffs.

Public Long-Term Evolution Networks. This option would pose more security concerns and would require lengthy negotiations with foreign countries. If the U.S. were to use another country's public LTE networks, it could not guarantee the security of those connections. This would cause a problem for any information traversing through the network. Most countries would not willingly allow another country to openly use its public networks for military or government purposes.

Private Long-Term Evolution Networks. OCONUS deployments can be more challenging because the U.S. will be operating on foreign soil.

There are many ways to use the spectrum available, but that requires different deployment strategies. Logistical considerations for private, government-built networks are still applicable because they will need to have the equipment procured and maintained. This will give the DOD more freedom to operate as desired and it can choose the bands to support. It is best to be as diverse as possible with band selections. There will be many different operating constraints and diversity can mitigate them by offering different solutions where applicable. Therefore, deployments should use strategies with a combination of licensed, unlicensed, and shared spectrum. In some countries, spectrum allocation may not cover all of the bands so the DOD could easily adopt the unlicensed and shared spectrum techniques required to operate on these bands. If licensed bands will be used, then the networks will most likely have to be built-out privately by the government and the licenses purchased.

Impact on Data Management. As discussed in this section, there are a variety of deployment models for enabling networks for data transmission CONUS and OCONUS. The considerations, availability, bandwidth, and costs of various deployment models will vary, and the data transmission requirements of the use case are likely to drive the decisions for network support. If the Special Operations Forces (SOF) enterprise truly wants to operate based on a big data foundation, it will have to generate the downrange architecture for transmitting the data in the first place. Otherwise, having all the data scientists in the world will not be able to compensate for the lack of input—once again, the garbage in-garbage out conundrum. What this section hopefully revealed is that a true enterprise-wide big data capability has second- and third-order communication costs associated with the capability. Certainly, operating offline is an option, but then the benefits of operating at the speed of artificial intelligence (AI) become moot. Having a clear sense of the communication options—and the costs associated with them—is a preliminary requirement for building out an enterprise architecture. Without this step, much of the big data capability the SOF enterprise desires will be undermined by a lack of timely data.

Disruptive Technology

Despite the government's lag behind industry in adopting and driving technological change, there are technological advances that are being actively monitored and pro-actively driven by government use cases. Quantum computing promises to be a disruptive change to the model of computing in multiple domains—particularly that of current AI applications. Its application toward breaking existing encryption techniques is also well-publicized.[28] In a variety of domains (and related to the discussion of CONUS and OCONUS bandwidth), data must be exchanged and processed as effectively as possible despite network restrictions. Data exchange, storage, and processing in potentially disconnected environments is a current research effort explored in this section.

Quantum Computing

Quantum mechanics is the physical theory describing nature at its smallest and most fundamental level. Particles such as atoms, electrons, and photons, as described by quantum mechanics, behave very differently than cars, tables, and other objects that people are accustomed to from everyday experience. For example, a quantum system such as an electron can spin in two opposite directions at the same time, a quantum phenomenon known as superposition. Multiple quantum systems can exhibit correlations between them, known as entanglement, that are impossible from the classical view of the world. Over the past three decades, substantial effort has been spent exploring these phenomena to exploit them and improve a range of technologies.

Quantum information is the field dedicated to utilizing quantum mechanics for the improvement of information-related technologies. These technologies are computation, communications/cryptography, and sensing. In each of these fields, significant progress has been made, first in academia and government laboratories and more recently in industry. This section concentrates on quantum computation, the study of how quantum phenomena can improve computation. It discusses what a quantum computer can and cannot do, reviews the current state of the field, and outlines some of the basic concepts.

Why Quantum Computing? The fundamental component of a quantum computer is a qubit (also known as a two-state quantum system parallel to the bit in classical computers). However, unlike classical bits, qubits exhibit quantum phenomena: they can exist in superpositions and become entangled, resulting in massively parallel computing. These phenomena allow a quantum computer to solve certain computation problems more efficiently than is possible for classical computers. The term efficiency is used here with respect to the number of resources (number of gates or time) needed to implement an algorithm to perform a certain task. This section later describes some of the tasks for which algorithms implemented on a quantum computer provide an advantage.

More on Qubits. The fundamental component of a classical computer is a bit, a system that can be in one of two possible states: off or on, or alternatively, 0 and 1. In modern computers, a bit may be comprised of a capacitor that will be in the state charged or uncharged. In order to perform an algorithm, a series of single- and multi-bit gates are applied to the bits. The fundamental unit of a quantum computer is a qubit, a two-state quantum system. A possible qubit is a superconducting Josephson junction,[29] a small loop of superconducting material broken by a thin slab of non-superconducting material. Current in this loop can rotate clockwise and counterclockwise. Another possible qubit is an atomic ion in which an electron can inhabit the ground or first excited state.[30] Again, the two states can be referred to as 0 and 1. However, unlike classical bits, qubits can exist in both states simultaneously. In such a situation, the qubit is said to be in a "superposition" of states. In fact, there are an infinite number of superpositions as the qubit can exist in a state anywhere between 0 and 1. The state of the qubit can be described as a point on the surface of a sphere whose poles represent the 0 and 1 state. Mathematically (and presented for clarity and completeness), the state of the qubit can be described as $\alpha|0\rangle+\beta|1\rangle$ where α and β are complex amplitudes.[31]

While one qubit can exist in a superposition of two states, two qubits can exist in a superposition of up to four states: 00, 01, 10, and 11. Generalizing from the one-qubit case, mathematically the general two-qubit state can be described as $\alpha|00\rangle+\beta|01\rangle+\gamma|10+\delta|11$.[32] If the two qubits are known to be in this state and it is possible to then measure one of the two qubits, it can be known with certainty the state of the other qubit. Another way to look at this is to realize that the state of one of the qubits cannot be properly described without stating the state of the second qubit. This high degree of correlation, which is impossible for classical systems, is called entanglement and is a fundamental resource for many quantum protocols in quantum communication and computation.

Breaking Public Key Encryption. A quantum computer could break public key encryption (including that used in blockchain ledgers).[33] Current public key encryption schemes are secure because they rely on the inefficiency[34] of a computer in solving mathematical problems such as factorization, discrete logarithms, and the elliptic-curve discrete logarithms. Shor's algorithm,[35] perhaps the most well-known of the quantum algorithms, allows a quantum computer to efficiently solve any of these problems, thus rendering practically all known public key encryption protocols (including PKI and RSA) insecure.

Searching Unsorted Databases and Function Inversion. Grover's algorithm[36] allows a quantum computer to search an unsorted database utilizing only a square root of the number of gates that would be necessary for a classical computer. The same concept can be used for function inversion. Thus, this algorithm is used as a sub-routine in other algorithms such as quantum machine learning (ML).

Solving Linear Systems of Equations. The Harrow-Hassidim-Lloyd algorithm[37] solves systems of equations (or, equivalently, inverts a matrix) using exponentially fewer gates than is possible for a classical computer. While there are certain restrictions on the complexity of the equations, a version of this algorithm has been shown to efficiently solve for radar cross-sections.

Quantum Simulations. Computational material science is a field that attempts to simulate materials at the most basic level. However, the use of a classical computer to simulate what is inherently a quantum system is highly inefficient. Quantum computers, on the other hand, may properly and efficiently simulate such systems.[38] Such simulation could provide vital insight into the nature of certain materials. For example, the phenomenon of high temperature super conduction, which currently suffers from a lack of basic theoretical understanding, may be aided by quantum simulations, perhaps allowing for the design of material that is superconducting at even higher temperatures.

Quantum Machine Learning

Given the advent of ML, it was only a matter of time before the question arose as to the utility of quantum computers for these tasks. The answer is yes, quantum computers can utilize less samples in sampling algorithms, train on smaller training sets, efficiently determine quality of least-squares fit for exponentially large datasets, and speed up for support vector machines, among other ML protocols.[39]

Not every computational task can be solved more efficiently on a quantum computer. In fact, the entire field of post-quantum cryptography seeks to build public key encryption protocols based on computation problems that are not breakable by a quantum computer. Nonetheless, the above potential has galvanized researchers and industry throughout the world to try to build quantum computers.

Not every computational task can be solved more efficiently on a quantum computer.

Why Are Quantum Computers So Hard to Build?

While nature is fundamentally quantum, people do not experience quantum phenomenon in their daily lives. This is because quantum phenomenon such as superposition and entanglement get suppressed when the systems exhibiting these traits interact with their environment (an interaction known as decoherence). Thus, a quantum computer must be built in as complete isolation as possible. Simultaneously, the different qubits must interact with some sort of external system and with each other to implement computational gates that make up an algorithm. This balance between isolation and interaction creates quite the tightrope for a quantum computer designer. Generally, it is assumed that a quantum computer will require the utilization of strong magnetic fields or need to operate at extremely low temperatures.

Another reason why quantum computers are so hard to build is because of errors in the implementation of basic gates. As previously seen, quantum systems can evolve in ways impossible for classical systems. This also means that quantum systems can go wrong in ways that are impossible for classical systems. While a classical bit can be in error only via a bit flip, there are an infinite number of possible ways a qubit can be in error. Identifying and fixing such errors is complicated by the fact that measurement of a quantum system tends to change its state. The solution to this is called quantum error correction (QEC), a framework allowing quantum information to be encoded in multiple qubits. However, QEC requires a myriad of additional qubits and gates, further complicating the task of building a quantum computer.

Current State of Quantum Computing

There is a plethora of quantum computing startups[40] and a number of established industries[41] attempting to construct quantum computers. Currently, the most popular and advanced form of qubit are superconducting qubits, Josephson junctions in which current can rotate clockwise and counterclockwise. Google has fabricated a chip with 72 qubits,[42] IBM has one with 50 qubits,[43] and Intel has 49 such qubits.[44] While these advances have rightfully been given much press, the number of qubits alone is an insufficient metric in determining how close the world is to a fully capable quantum computer. Demonstrating control and accuracy of basic gates is necessary to show a path forward. A number of early quantum computers are available to the general public. Startup Rigetti Computing[45] is providing access to its 19-qubit system online. IBM was the first to put such as system online (one with five qubits) and has made a 20-qubit system available as a cloud service.[46]

Academia and other startups are exploring different types of qubits. IonQ[47] is investigating trapped-ion quantum computers, and startups Xanadu and Sparrow Quantum are looking to build a quantum photonic processor. In China, Alibaba has launched an 11-qubit quantum computing cloud service[48] and Baidu[49] has promised a large investment as well.

Various governments are investing heavily as well. More recently, the European Union (1 billion Euros over 10 years) and China ($10 billion) have increased funding for quantum information.

Quantum Annealers. Another type of system that utilizes quantum phenomenon for quantum computation, but which is not a full quantum computer, is a quantum annealer. A quantum annealer, as manufactured by the Canadian company D-Wave,[50] can solve what is known as an Ising problem. Such problems can be mapped to certain optimization problems such as the traveling salesman problem. The traveling salesman problem is computationally expensive to solve given current, classical algorithmic solutions. It is a problem that—given a graph of nodes (cities) and edges connecting the edges of varying weights (distance between cities)—seeks the optimal (shortest) route that visits all notes in the graph and returns to the starting node. The most advanced D-Wave system currently consists of 2,000 qubits (though not with the complete control necessary for quantum computing) and utilizes a phenomenon known as quantum tunneling to detour through different possible solutions in an effort to find a global minimum (the most optimal solution). Recent investigations of the D-Wave quantum annealer have concentrated on determining whether the system can outperform all classical optimization routines and towards simulating quantum systems.

When will there be a quantum computer capable of breaking public key encryption protocols and what can be done about it? There is a divergence of views as to when quantum computers with sufficient resources and accuracy to break public key encryption will come online. The estimates range from ten years to never, with most experts assuming somewhere between twenty to thirty years. While even ten years may seem like a long way off, the overhaul of public key encryption to utilize a protocol that is quantum resistant is a major task likely to take at least 20 years. Hence, the need is to start addressing the challenge now.

In 2017, the American Innovation and Competitiveness Act tasked NIST to "research and identify, or if necessary, develop cryptography standards and guidelines for future cybersecurity needs, including quantum-resistant cryptography standards."[51] There are two general approaches to cryptography that can counter the power of a quantum computer.[52] The first is to utilize a public key encryption protocol based on a classical algorithm which cannot be more efficiently implemented on a quantum computer. One possible example is cryptographic systems which rely on a symmetric key like Advanced Encryption Standard. However, this may be unwieldy for public key encryption purposes. Other examples include lattice-based cryptography and supersingular elliptic curve isogeny cryptography. The second approach is to utilize quantum cryptography. Quantum cryptography (or more accurately, quantum key distribution [QKD]) is a secure method of sharing a cryptographic key via line of sight. Quantum phenomenon are invoked to determine the presence of an eavesdropper during the key sharing process. The key can then be used for a one-time pad. There are several commercial QKD systems available. However, the use of one-time pads is generally regarded as expensive (one bit of key is needed to encode one bit of message) and overkill since security based on algorithm inefficiencies is usually sufficient. In addition, this technology has not undergone a proper vetting process to ensure robustness and lack of vulnerabilities.

The Future of Tactical Data Processing

While this chapter has touched on traditional cloud architectures and next generation quantum computing, users in environments with unreliable connections between computing resources cannot make use of these resources. Traditional cloud computing architectures—according to official NIST definitions—must offer several features: on demand self-service, broad network access, resource pooling, rapid elasticity, and measured service. They also offer IaaS, PaaS, and SaaS options for operation within a cloud.

Tactical clouds are being investigated to enable cloud services to be implemented in potentially challenged environments.[53] Tactical clouds operate in contested and often disconnected environments using smaller-form devices than their traditional cloud counterparts. These devices are often physical components that form the cloud rather than the virtual components that are provisioned on demand in a traditional cloud. As such, what are considered tactical clouds often do not meet the NIST definition of a cloud because of their inability to satisfy and offer the required features; often, tactical clouds are more specialized and limited (e.g., offering potentially only infrastructure and distributed computing/storage) rather than the full spectrum of cloud features. While specific definitions of a tactical cloud will exist and likely vary, several elements of a tactical cloud are common across implementations. Primarily, there exists a need to connect separate computational elements within potentially disconnected or austere environments to one another to share data, computational responsibilities, or knowledge. In most cases, a larger traditional cloud service exists in a non-tactical environment to which the tactical components send, receive, and synchronize data.

To operate at the tactical edge, the tactical components of the architecture frequently leverage techniques such as shared local caches, message routing, and caching and queuing data transfer to operate when disconnected. Further, data prioritization and shared computational responsibilities are implemented in the tactical application layer. Due to the potentially unreliable networked connections between cloud members, nodes, or other elements, applications must offer techniques that provide the appearance of reliable connectivity to the user and mitigate the impact of unreliable connections on the cloud's operation. For example, queueing requests or accessing local peer caches for responses to information requests are common tactical cloud techniques. Further, data is often not transmitted in its entirety but rather signatures are computed, summarizing information is derived, or high priority data is transmitted at the expense of raw data.

Again, while other interpretations of tactical cloud computing may vary, they share the need to operate, access, and maintain a computation capability in a tactical environment.

Coupled with the challenges facing bandwidth CONUS and OCONUS, the ability to process data effectively independent of global networking is beneficial for operators in disconnected, interrupted, and low-bandwidth environments. Localized networks such as Bluetooth, radio frequency, or localized Wi-Fi can help exchange data locally to enable the local distributed computation similar to that in a fully capable cloud. This model also assumes trust can be established between participating nodes, a problem not present in traditional cloud models.[54]

Access to information derived from data at the tactical edge may be essential to mission success. As such, access to computation resources or cloud architecture to process data at the point of collection can help increase mission effectiveness.

Planning Appropriately

Federal government agencies often establish plans and strategies for adopting future disruptive technologies along with currently in-development technologies that are not yet mature enough for government adoption. Understanding the implications of design decisions and the way in which solutions are considered, evaluated, and adopted have an impact on the level of effort and potential successes of technology adoption programs. This section discusses two aspects of program planning. The first is the selection of software and the impact this has on a project. The second is a set of recommended best practices for engaging with emerging technology domains.

Access to information derived from data at the tactical edge may be essential to mission success.

Software Selection

There are a variety of software types that can be considered during project and program planning. Selecting a type of software to adopt should be driven by the business and technological needs of the project. Due to the differences, trade-offs, and benefits of various software types, the utility and suitability of a type of software will vary depending on the intended implementation and use case. Prior to selecting a type of software to procure, stakeholders should work to understand the full software lifecycle to include the ability to support, develop, and adopt the software over time. Market research and alternatives analysis can help drive the decision-making as well.

Open-source software is typically community supported and provides open access to the software and associated source code. Open-source software is often community supported with many contributors that help maintain the code base or even adapt the software to add functionality. Shareware is proprietary software that is provided to users with no cost but is still profitable to the software owners. Commercial off-the-shelf (COTS) and government off-the-shelf (GOTS) software is proprietary software that is available for purchase and is often fully supported by a commercial or government entity. COTS and shareware software are centrally controlled and developed as opposed to the community supported nature of open-source software. This section explores examples and varieties of each software type and discusses the general trade-offs using each in a project.

Open Source and Free and Open Source. Open-source (often referred to as free and open-source software or FOSS) project usage is gaining popularity and attention in the federal government. The quality of available tools and the ability to customize existing applications drive the appeal. However, utilizing open-source software often means it must be maintained and supported in-house utilizing a software team. Open source projects may require customization for effective utilization as well. This is the trade-off between the low (i.e., free) up-front software costs, the customizability, and the need to support in-house.

Due to these considerations, government adopters must consider the intended use and ability to operate without third party support for the software (and the resulting potentially elevated level of effort). Open-source software may not be approved for secure environments, creating a longer process of adoption for classified uses. Examples of open-source software that might be used for data analysis in the government includes D3.js[55] for browser-based data visualization or NumPy[56] for scientific computing in Python.

Commercial Off-the-Shelf and Government Off-the-Shelf Software. COTS software is widely understood and more typically acquired for government projects. Similarly, GOTS projects are available and often have the same characteristics of COTS.[57] Government users of COTS typically purchase licenses for software that is fully supported by the vendor. This software is more frequently approved and vetted for sensitive environments and uses than open-source software. Because of these features, COTS software is typically easier to procure and implement as compared to open source. Examples of COTS products include Microsoft products.[58] While the cost for licenses is higher than FOSS, the robustness and support of the software is typically higher and often requires less in-house expertise.

Shareware has features of COTS and open source. In general, shareware source code is not exposed (as it is in open source) and it is available for use without a license (as opposed to the licensing models in COTS). Shareware software is provided to users for no initial cost, but there is a model for profit by the software maintainers. There are a variety of models for software vendors to profit from shareware.

In the freemium model, basic or bounded functionality is available for no cost but with increased features or enhanced usage available for a fee. Examples of broadly available freemium services include DropBox[59] (small, bounded storage available for free, and advanced features and increased storage available for purchase) and Amazon Web Services[60] (which is free for a small amount of computation but expanded usage must be purchased). Adware applications are free to use but inject advertisements into the software, either as barriers to unlocking content or on-screen real estate. In other cases, shareware software is provided with limited functionality. Demoware and trialware are both abbreviated versions of a full production package. These are related to crippleware, which is a software package that degrades over time until a license is purchased. Donationware solicits monetary donations from users. This model is typical with non-profit organizations that provide software (e.g., The Apache Foundation[61]). Finally, freeware is a software package that is offered for free and often has a governing or guiding board that may be sponsored by participating organizations. The R Project[62] is an example of freeware. In current markets, shareware (and even COTS) is often delivered as SaaS to users via a web browser for free with premium upgrades available (e.g., Slack[63]).

Preparing for the Future

Planning for disruptive technology adoption five to ten years in the future is difficult without a foundational set of variables regarding the indicators and specific markets being disrupted as inputs to an effective prediction model.[64] The ability to predict and assess the effect of disruptive technologies is a debate among academic researchers. Government has the ability to monitor technical evolutions to understand their impact on government missions through agile innovation practices (e.g., focusing on agile principles to understand, re-vector, and identify direction for innovative solutions to gaps and needs). This can be performed using rapid prototyping on emergent technologies to better understand the current needs, solutions, and requirements moving forward.

The U.S. Federal Government has invested in organizations and physical and virtual spaces in which innovation is fostered. These innovation entities often encourage the collaboration and collective effort of academics, industry representatives, and other partners (e.g., hobbyists and startups). Through rapid prototyping, collaboration, evaluation, testing, and development, the participants and government sponsors are able to refine approaches, better understand gaps, and foster ideas for solutions—both conventional and non-conventional—to government challenges. Ultimately, government innovation requires a culture that supports an open environment for the exchange of ideas.

By incorporating agile practices into technology outreach or evaluation, government adopters can stay more informed on the latest aspects of potential technical solutions for challenges. Engaging with innovation cells[65] allows government to engage and create relationships with organizations and practitioners that may not typically participate in government innovation activities. In the process of this innovation, several best practices and features of optimal innovation environments exist:

- Opportunities for engagement with diverse, non-traditional partnerships to leverage the best in breed from across the domain
- Freedom to fail and fail fast to receive immediate feedback on what does and does not work well with regards to a gap, which will also allow gaps and needs to be refined by the government based on lessons learned during the investigation
- Feedback loops to apply the state-of-the-possible technologies to the government gap
- A test and evaluation environment with a direct path to adoption by government users

With an environment and culture that fosters innovation, government adopters can maintain awareness of current state of the art technologies. While this does not directly help government adopters and practitioners forecast disruption in a domain, the engagements with industry, academia, and peer government can help raise awareness of current gaps and solutions in other domains and monitor the progress of the respective technologies.

Assume Five Years to Field

Whatever choices are made in developing an enterprise architecture, managers should remember that purchasing, fielding, and training all unfold on different, often protracted timelines. While there is no single way to anticipate the length of time necessary to make a big data capability fully functional for SOF, simply recognizing the different command, component, and unit purchasing authorities and cultures brings into perspective the scale of the undertaking. Given the rate of technological change, it is unlikely that making decisions based on today's operational and administrative needs will be sufficient for delivering an enterprise capability with all the utility necessary for when it actually achieves full operational capability. As a rule of thumb, leaders should do their best to imagine what requirements might look like five years into the future and ensure whatever solution is adopted allows for adaptation to the unknown. With disruptive technologies, amplified requirements for bandwidth, structured systems for ingesting and digesting data from diverse units, and a flexible system are essential. While many SOF leaders want to get after the current mission with Big Data, an enterprise architecture must ultimately fall back on the attitudes in chapter five to ensure that the system chosen can accommodate the factors described above.↑

Endnotes

1. Nathan Strout, "Making the Hyper-Enabled Operator a Reality," *C4ISRnet*, 12 June 2020, https://www.c4isrnet.com/battlefield-tech/2020/06/12/making-the-hyper-enabled-operator-a-reality/.
2. Strout, "Making the Hyper-Enabled Operator a Reality."
3. "About," Gmail.com, Google, accessed 27 October 2021, https://www.google.com/gmail/about/.
4. "About," Google Docs, Google, access 27 October 2021, https://www.google.com/docs/about/.
5. Peter Mell and Tim Grance, "The NIST Definition of Cloud Computing," *National Institute of Standards and Technology Special Publication* 800, no. 145 (September 2011): 2, https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-145.pdf.
6. Mell and Grance, "The NIST Definition of Cloud Computing," 2.
7. Mell and Grance, "The NIST Definition of Cloud Computing," 2-3.
8. Mell and Grance, "The NIST Definition of Cloud Computing," 3.

9. Justin F. Brunelle et al., *July 2018 Federal Cloud & Data Center Summit Report* (Washington, D.C.: The Mitre Corporation, 2018), https://www.mitre.org/sites/default/files/publications/PRS18-2725-1_june2018_federal_cloud__data_center_summit_report.pdf.
10. "Home," The Federal Risk and Authorization Management Program, accessed 20 October 2021, https://www.fedramp.gov/.
11. Cloud Migration options are provided in the last section of this chapter.
12. Jeffrey Dean and Sanjay Ghemawat, "MapReduce: Simplified Data Processing on Large Clusters," *OSDI '04: Sixth Symposium on Operating System Design and Implementation* (San Francisco: Operating Systems Design and Implementation, 2004), 137-150.
13. The "cloud-first" policy refers to the Office of Management and Budget (OMB) mandate that U.S. Federal Government agencies adopt cloud policies when possible and suitable for their needs. (https://cloud.cio.gov/).
14. Brunelle et al., *July 2018 Federal Cloud & Data Center Summit Report*.
15. Kevin Buck, Diane P. Hanf, and Daniel J. Harper, "Cloud SLA Considerations for the Government Consumer" in *Systems Engineering Cloud Computing Series* (Bedford: The Mitre Corporation, 2015), https://www.mitre.org/publications/technical-papers/cloud-sla-considerations-for-the-government-consumer-0; Kevin S. Buck, Anthony E. Dziepak, and Daniel Harper, "Modeling the Influence of System and Application Complexity on the Cost of Cloud Hosting," *June 2017 International Cost Estimating And Analysis Training Workshop* (Bedford, The Mitre Corporation, 2017), https://www.iceaaonline.com/ready/wp-content/uploads/2017/07/MM14-PPT-Harper-Modeling-Influence-System-and-Application.pdf.
16. Justin F. Brunelle et al., *July 2016 Federal Cloud Computing Summit Summary*, 16-3496 (Washington, D.C.: The MITRE Corporation, 2016), https://www.mitre.org/sites/default/files/publications/16-3496-federal-cloud-computing-summit-report.pdf.
17. Brunelle et al., *July 2018 Federal Cloud & Data Center Summit Report*.
18. Don Faatz and Mari Spina, "Cybersecurity in the Cloud: The Federal Landscape for Secure Cloud Services, Systems, and Solutions," *The MITRE Corporation*, January 2017, https://apps.dtic.mil/sti/pdfs/AD1107989.pdf.
19. "Home," bitcoin.org, accessed 21 October 2021, https://bitcoin.org/en/; "The Trust Machine," *The Economist*, 31 October 2015, https://www.economist.com/leaders/2015/10/31/the-trust-machine.
20. Dave Bryson, Dave Penny, David C. Goldenberg, and Gloria Serrao, "Blockchain Technology for Government," *The MITRE Corporation*, December 2017, http://www.mitre.org/publications/technical-papers/blockchain-technology-for-government.
21. "Home," Ethereum, accessed 21 October 2021, https://ethereum.org.

22. Brian Curran, “What is Game Theory? And How Does It Relate to Cryptocurrency?” *Blockonomi*, 13 August 2020, https://blockonomi.com/game-theory/.

23. Dave Bryson, David R. Penny, David C. Goldenberg, and Gloria J. Serrao, “Blockchain Technology for Government,” *The MITRE Corporation*, December 2017, https://www.mitre.org/sites/default/files/publications/blockchain-technology-for-government-18-1069.pdf.

24. “Home,” Etherscan, accessed 21 October 2021, https://etherscan.io/.

25. “Home,” Tendermint, accessed 21 October 2021, https://tendermint.com/.

26. LTE is a wireless communication standard often used for data transmission for mobile devices.

27. More information can be found at https://www.firstnet.gov/.

28. Meredith Rutland Bauer, “Quantum Computing Is Coming for Your Data,” *Wired*, 19 July 2017, https://www.wired.com/story/quantum-computing-is-coming-for-your-data/; Paul Teich and Tirias Research, “Quantum Computing Will Not Break Your Encryption, Yet,” *Forbes*, 23 October 2017, https://www.forbes.com/sites/tiriasresearch/2017/10/23/quantum-will-not-break-encryption-yet/#6d5c40bd7319.

29. Yuriy Makhlin, Gerd Schön, and Alexander Shnirman, “Quantum-State Engineering with Josephson-Junction Devices,” *Review of Modern Physics* 73, no. 2 (2001): 357-400.

30. J. I. Cirac and P. Zoller, “Quantum Computations with Cold Trapped Ions,” *Physical Review Letters* 74, no. 20 (1995): 4091-4094.

31. It is imperative to note however that upon measurement, the state of the qubit will be changed to either 0 or 1 with probability $\alpha|$^2 or $|\beta|$^2.

32. Now let us examine the two qubit state in which $\beta,\gamma=0$ and $\alpha=\delta={}^{1}/\sqrt{2}$. The two qubits are thus in the state ${}^{1}/\sqrt{2}(|00\rangle+|11\rangle)$.

33. Brandon V. Rodenburg and Stephen P. Pappas, “Blockchain and Quantum Computing,” *The MITRE Corporation*, June 2017, https://www.mitre.org/sites/default/files/publications/17-4039-blockchain-and-quantum-computing.pdf.

34. An algorithm is inefficient if the number of resources needed for its implementation grows exponentially with the size of the problem. For example, the number of gates needed to factor a number grows exponentially as the number increases. An algorithm is efficient if that rate of growth is polynomial (or slower).

35. P.W. Shor, “Algorithms for Quantum Computation: Discrete Logarithms and Factoring,” in *Proceedings 35th Annual Symposium on Foundations of Computer Science* (Santa Fe: IEEE, 1994), 124–134, https://ieeexplore.ieee.org/document/365700.

36. L.K. Grover, “A Fast Quantum Mechanical Algorithm for Database Search,” in *Proceedings 28th Annual ACM Symposium on the Theory of Computing* (Philadelphia: ACM, 1996), 212, https://arxiv.org/pdf/quant-ph/9605043.pdf.

37. Aram W. Harrow, Avinatan Hassidim, and Seth Lloyd, "Quantum Algorithm for Linear Systems of Equations," *Physical Review Letters* 103, no. 15 (October 2009): 150, 502.

38. Tom Simonite, "Chemists Are First in Line for Quantum Computing's Benefits," *MIT Technology Review*, 17 March 2017, https://www.technologyreview.com/s/603794/chemists-are-first-in-line-for-quantum-computings-benefits/.

39. Jacob Biamonte et al., "Quantum Machine Learning," *Nature* 549 (September 2017): 195–202.

40. "Private/Startup Companies," Quantum Computing Report, accessed 4 December 2020, https://quantumcomputingreport.com/players/privatestartup/.

41. "Public Companies," *Quantum Computing Report*, accessed 4 December 2020, https://quantumcomputingreport.com/players/public-companies/.

42. Julian Kelly, "A Preview of Bristlecone, Google's New Quantum Processor," *Google AI Blog*, 5 March 2018, https://ai.googleblog.com/2018/03/a-preview-of-bristlecone-googles-new.html.

43. Samuel K. Moore, "IBM Edges Closer to Quantum Supremacy with 50-Qubit Processor," *IEEE Spectrum*, 15 November 2017, https://spectrum.ieee.org/tech-talk/computing/hardware/ibm-edges-closer-to-quantum-supremacy-with-50qubit-processor.

44. Jeremy Hsu, "CES 2018: Intel's 49-Qubit Chip Shoots for Quantum Supremacy," *IEEE Spectrum*, 9 January 2018, https://spectrum.ieee.org/tech-talk/computing/hardware/intels-49qubit-chip-aims-for-quantum-supremacy.

45. "Home," Rigetti, accessed 21 October 2021, https://rigetti.com.

46. "Bluemix Is Now IBM Cloud," ibm.com, accessed 27 October 2021, https://www.ibm.com/cloud/bluemix?utm_content=SRCWW&p1=Search&p4=43700051838803772&p5=b&gclid=EAIaIQobChMIxo6oztTt8wIVaU1yCh15kgEoEAAYASAAEgIIa_D_BwE&gclsrc=awds.

47. "Home," IONQ, accessed 21 October 2021, https://ionq.co/.

48. Flann Gao and Luica Mak, "Alibaba Cloud and CAS Launch One of the World's Most Powerful Public Quantum Computing Services," *Alibaba Cloud*, 1 March 2018, https://www.alibabacloud.com/press-room/alibaba-cloud-and-cas-launch-one-of-the-worlds-most.

49. Yiting Sun, "Baidu Has Entered the Race to Build Quantum Computers," *MIT Technology Review*, 8 March 2018, https://www.technologyreview.com/the-download/610449/baidu-has-entered-the-race-to-build-quantum-computers/.

50. "Quantum Computing: How D-Wave Systems Work," *DWavesys*, accessed 4 December 2020, https://www.dwavesys.com/quantum-computing.

51. American Innovation and Competitiveness Act, Pub. L. No. 114-329, 130 Stat. 2969, Sec. 104(b)(2)(B)(i) (2017), https://www.congress.gov/114/plaws/publ329/PLAW-114publ329.pdf.

52. Dustin Moody, Larry Feldman, and Gregory Witte, eds., “Securing Tomorrow’s Information through Post-Quantum Cryptography,” Information Technology Laboratory, February 2018, https://csrc.nist.gov/CSRC/media/Publications/Shared/documents/itl-bulletin/itlbul2018-02.pdf.

53. Grace Lewis, Sebastián Echeverría, Soumya Simanta, Ben Bradshaw, and James Root, “Tactical Cloudlets: Moving Cloud Computing to the Edge,” *2014 IEEE Military Communications Conference* (Baltimore: IEEE, 2014), 1440–1446, https://ieeexplore.ieee.org/document/6956959; Mahadev Satyanarayanan, Grace A. Lewis, Edwin Morris, and Soumya Simanta, “The Role of Cloudlets in Hostile Environments,” *IEEE Pervasive Computing* 12, no. 4 (October 2013): 40–49.

54. Sebastián Echeverría, Dan Klinedinst, Keegan Williams, and Grace A. Lewis, “Establishing Trusted Identities in Disconnected Edge Environments,” *2016 IEEE/ACM Symposium on Edge Computing* (Washington, D.C.: IEEE, 2016), 51–63, doi: 10.1109/SEC.2016.27.

55. “Home,” d3js.org, accessed 21 October 2021, https://d3js.org.

56. “Home,” NumPy, accessed 21 October 2021, http://www.numpy.org.

57. For the purposes of this discussion, we refer only to COTS with an understanding that the same considerations of COTS usage typically apply to GOTS.

58. “Compare All Microsoft 365 Products,” Microsoft, accessed 21 October 2021, https://www.microsoft.com/en-us/store/b/office.

59. “Home,” Dropbox, accessed 21 October 2021, https://www.dropbox.com/.

60. “AWS Free Tier,” AWS, accessed 21 October 2021, https://aws.amazon.com/free/.

61. “Home,” The Apache Software Foundation, accessed 21 October 2021, http://www.apache.org.

62. “Home,” The R Project for Statistical Computing, accessed 21 October 2021, https://www.r-project.org.

63. “Home,” Slack, accessed 21 October 2021, https://slack.com.

64. Jonathan Linton, “Forecasting the Market Diffusion of Disruptive and Discontinuous Innovation,” *IEEE Transactions on Engineering Management* 49, no. 4 (November 2002): 365–374.

65. Examples of innovation cells in the Government include a wide variety of organizations such as SOFWERX, Defense Innovation Unit, MD5, In-Q-Tel, and Veterans’ Affairs Center for Innovation (VACI).

Acronyms

AI	artificial intelligence
BYON	bring-your-own-network
CONUS	continental United States
COTS	commercial off-the-shelf
DDS	Defense Digital Service
DL	deep learning
DOD	Department of Defense
F3EAD	find, fix, finish, exploit, analyze, disseminate
FOSS	free and open-source software
GFT	Google Flu Trends
GOTS	government off-the-shelf
GS	general schedule
HEO	hyper-enabled operator
HQ	headquarters
IaaS	infrastructure as a service
IC	intelligence community
IED	improvised explosive device
IoT	internet of things
ISR	intelligence, surveillance, and reconnaissance
IT	information technology
JSOU	Joint Special Operations University
LTE	long-term evolution
ML	machine learning
MVNO	mobile virtual network operator

NGA	National Geospatial Agency
NIST	National Institute of Standards and Technology
NLP	natural language processing
NLU	natural language understanding
OCONUS	outside the continental United States
OPORD	operations order
OS	operating system
OSINT	open-source intelligence
PaaS	platform as a service
PAL	priority access license
PMP	point-to-multipoint
PTP	point-to-point
QEC	quantum error correction
QKD	quantum key distribution
QoS	quality of service
RAN	radio access network
SaaS	software as a service
SIM	subscriber identity module
SLA	service-level agreement
SME	subject matter expert
SOF	Special Operations Forces
TLP	troop leading procedure
UAV	unmanned aerial vehicle
USG	United States Government
USSOCOM	United States Special Operations Command
WARNORD	warning order