SYM-AM-20-067

Proceedings
of the
Seventeenth Annual Acquisition Research Symposium

Acquisition Research: Creating Synergy for Informed Change

May 13–14, 2020

Published: April 15, 2020

Approved for public release; distribution is unlimited.

Prepared for the Naval Postgraduate School, Monterey, CA 93943.

The research presented in this report was supported by the Acquisition Research Program of the Graduate School of Defense Management at the Naval Postgraduate School.

To request defense acquisition research, to become a research sponsor, or to print additional copies of reports, please contact any of the staff listed on the Acquisition Research Program website (www.acquisitionresearch.net).

Acquiring Artificial Intelligence Systems: Development Challenges, Implementation Risks, and Cost/Benefits Opportunities

Johnathan C. Mun—is a research professor at the Naval Postgraduate School (and teaches master's and doctoral courses as well as executive seminars in quantitative risk analysis, decision sciences, real options, simulation, portfolio optimization, and other related concepts. He has also researched and consulted on many Department of Defense and Department of Navy projects and is considered a leading world expert on risk analysis and real options analysis. He has over 23 patents and patents pending and has authored 22 books including *Modeling Risk: Applying Monte Carlo Simulation, Real Options, Optimization, and Forecasting*, First, Second, and Third Editions (Wiley 2006, Wiley 2010, and Thomson-Shore 2015); *Real Options Analysis: Tools and Techniques*, First, Second, and Third Editions (Wiley 2003, Wiley 2005, and Thomson-Shore 2016); *Advanced Analytical Models* (Wiley 2008 and ROV Press 2016); *The Banker's Handbook on Credit and Market Risk* (Elsevier Science 2008); and others. He is the creator of the following software: *Real Options Super Lattice Solver, Risk Simulator, Project Economics Analysis Tool* (PEAT), *Modeling Toolkit, Risk Explorer*, and *ESO Valuation*.

Mun has taught at universities all over the world, from the Naval Postgraduate School (Monterey, CA) and University of Applied Sciences (Switzerland and Germany) as full professor, to Golden Gate University (California) and St. Mary's College (California), and has chaired many graduate research thesis committees. He also teaches public courses in risk analysis, real options analysis, and risk analysis for managers where participants can obtain the Certified in Quantitative Risk Management (CQRM) designation on completion of the week-long program. He also holds the position of the EU President of the American Academy of Financial Management and sits on the Global Board of Standards at the AAFM. He is currently the CEO of Real Options Valuation, Inc., and was formerly the Vice President of Analytics at Crystal Ball/Decisioneering, Inc., where he headed the development of options and financial analytics software products, analytical consulting, training, and technical support. Prior to that, he was a Consulting Manager and Financial Economist in the Valuation Services and Global Financial Services practice of KPMG Consulting and a Manager with the Economic Consulting Services practice at KPMG LLP. He has extensive experience in econometric modeling, financial analysis, real options, economic analysis, and statistics. He has consulted on a variety of real options, risk analysis, financial forecasting, project management, and financial valuation issues for over 100 multinational firms. His experience prior to joining KPMG included being Department Head of Financial Planning and Analysis at Viking Inc. of FedEx, performing financial forecasting, economic analysis, and research.

Mun received his PhD in finance and economics from Lehigh University. He also has an MBA in business administration, an MS in management science, and a BS in biology and physics. He is Certified in Financial Risk Management (FRM), Certified in Financial Consulting (CFC), and Certified in Quantitative Risk Management (CQRM). He is a member of the American Mensa, Phi Beta Kappa Honor Society, and Golden Key Honor Society as well as several other professional organizations, including the Eastern and Southern Finance Associations, American Economic Association, and Global Association of Risk Professionals. [Email: jcmun@nps.edu]

Tom Housel—specializes in valuing intellectual capital, knowledge management, acquisitions research, telecommunications, information technology, value-based business process reengineering, and knowledge value measurement in profit and nonprofit organizations. He is currently a tenured full professor for the Information Sciences (Systems) Department of the Naval Postgraduate School (NPS). He has managed a $5+ million portfolio of field studies, educational initiatives, and industry relationships.

His current research focuses on determining the most promising methodologies for improving the acquisitions life cycle and developing a physics-based model to predict revenue growth in knowledge-intensive companies. He continues to work in the behavioral accounting decision-making area and

has published in a number of international accounting journals, such as ABR and EAR. His research was published twice in the top information systems journal, *MIS Quarterly* (MISQ).

Prior to joining NPS, he was an Associate Professor at the Marshall School of Business at the University of Southern California. His last assignment in the corporate world was as the Chief of Consumer Market Research for Telecom Italia in Venice, Italy, where he developed new methods for predicting the adoption rates for new interactive multimedia broadband applications.

He received his PhD from the University of Utah in 1980. He won the prestigious Society for Information Management award for best paper in the field in 1986 that was subsequently published in the MISQ. His work on measuring the value of intellectual capital has been featured in a *Fortune* cover story (October 3, 1994) and in numerous books, professional periodicals, and academic journals. His books include *Measuring and Managing Knowledge* (English and Chinese versions) and *Global Telecommunications Revolution: The Business Perspective* with McGraw-Hill (both in 2001). [Email: tjhousel@nps.edu]

Abstract

The acquisition of artificial intelligence (AI) systems is a relatively new challenge for the U.S. Department of Defense (DoD). Given the potential for high-risk failures of AI system acquisitions, it is critical for the acquisition community to examine new analytical and decision-making approaches to managing the acquisition of these systems in addition to the existing approaches (i.e., Earned Value Management, or EVM). Also, many of these systems reside in small start-up or relatively immature system development companies, further clouding the acquisition process due to their unique business processes when compared to the large defense contractors. This can lead to limited access to data, information, and processes that are required in the standard DoD acquisition approach (i.e., the 5000 series). The well-known recurring problems in acquiring information technology automation within the DoD will likely be exacerbated in acquiring complex and risky AI systems. Therefore, more robust, agile, and analytically driven acquisition methodologies will be required to help avoid costly disasters in acquiring these kinds of systems. This research identifies, reviews, and proposes advanced quantitative, analytically based methods within the integrated risk management (IRM) and knowledge value added (KVA) methodologies to complement the current EVM approach.

Introduction

Acquisitions of artificial intelligence (AI) systems is a relatively new challenge for the Department of Defense (DoD). Given the high risk of failure for such system acquisitions, it is critical that the acquisition community examines potential new approaches to help manage the AI acquisition life cycle. The well-known recurring problem in acquiring information technology within the DoD will likely be exacerbated in acquiring these complex and risky systems. The identification, review, and recommendation for use of new acquisition methodologies, to supplement or replace existing methodologies, should help avoid costly disasters in AI system acquisitions.

AI has been in use in various commercial and governmental domains to address a variety of decision support problems. However, existing DoD acquisition frameworks may not be adequate to address the unique nature of AI systems life-cycle investments. AI systems are qualitatively different than standard automation systems that focus on routine, repeatable tasks. To develop acquisition frameworks for AI systems, it is first necessary to examine how AI systems will be used to support, or supplant, decision-makers. The purpose of this research project is to provide a set of quantitative and analytically based decision-making methodologies for acquiring AI systems that address the inadequacies of the current investment acquisition life-cycle framework.

Research and Problem Statement

The current problem at the DoD is that the complexity and speed of decision-making is increasing exponentially with the advent of intelligent systems that support or actually make decisions in time-critical, high-impact problem spaces. The current process management and control tools that a program manager (PM) might use to support acquisitions do not provide adequate warning of or sufficient information about the root causes of fiscal budgetary overruns and time schedule delays. This is a problem because PMs are, as a result, unable to respond to issues in a timely manner, delaying the delivery of promised capabilities to the services. Additionally, the money and resources spent in excess of the original budget could be used in other acquisition programs.

Research Objective

The purpose of this study is to better understand the possible causes of and solutions to the AI acquisition problem. This study will examine the strengths and weaknesses of several performance and project management methodologies. These methodologies, Earned Value Management (EVM), Knowledge Value Added (KVA), and Integrated Risk Management (IRM), are used to strategically and tactically plan, monitor in real time, measure, and preemptively forecast the value and progress of AI acquisitions. A review of these recommended project analysis and control methodologies will offer insights into the strengths and weaknesses each approach could offer acquisition professionals within the general phases of the Defense Acquisition System. This research could offer potential solutions to improve early warnings of cost and schedule overruns, and value opportunities foregone in the acquisition process. As such, this research focuses on the review of these methodologies and their applications to the acquisition process as they pertain to AI systems.

The primary objective of the proposed research is to provide a set of recommendations for acquiring organically developed AI systems. This will be accomplished by performing a comparison and contrast analysis of the proposed methodologies. The result will demonstrate when and how each method can be applied to improve the acquisitions life cycle for AI systems, as well as to provide additional insights and examples of how some of these methods can be applied for any system acquisition.

Literature Survey

This section starts with a discussion of the Defense Acquisition System and the system acquisition life-cycle methodologies that might support the acquisition process. Each methodology is then described, providing a basic understanding of its purpose as well as how the method can be applied to the acquisition life cycle. This review provides a basic understanding of these methods, whereas the next section covers each methodology in more detail.

Defense Acquisition System

The DoD oversees the acquisition of new systems through the Defense Acquisition System, which manages national investment in technologies, programs, and product support for the United States Armed Forces (DoD, 2003). Its primary objective is “to acquire quality products that satisfy user needs with measurable improvements to mission capability and operational support, in a timely manner, and at a fair and reasonable price” (DoD, 2003). Within the DoD Decision Support System there are three separate but interrelated processes: JCIDS, PPBE, and the Defense Acquisition System (DoD, 2017b). This research focuses on program management, versus contract management, within the Defense Acquisition System.

Acquisition programs are divided into different ACATs based on the type of program and the dollar amount that is spent or is projected to be spent within the program (DoD, 2015a). Figure 1 shows the various cost-based designations and categories within the Defense Acquisition System. All dollar amounts for ACAT classifications are calculated in Fiscal Year (FY) 2014 dollars (DoD, 2015a). ACAT1 designates major defense acquisition programs with an estimated research, development, and test and evaluation expenditure (RDT&E) of more than $480 million, or more than $2.79 billion for the total procurement (DoD, 2015a). An ACATA designation is for major automated information systems that will exceed $520 million in total life-cycle cost, $165 million in the total program cost, or $40 million for any single year of a program (DoD, 2015a). ACATII programs do not meet the criteria for ACATI and will spend more than $835 million in the total procurement (DoD, 2015a) or more than $185 million in RDT&E. Finally, ACATIII programs are those not meeting the criteria for ACATI or ACATII designation (DoD, 2015a). The various designations allow for decentralized control of a program as each category has different reporting requirements and designated decision-makers (DoD, 2017).

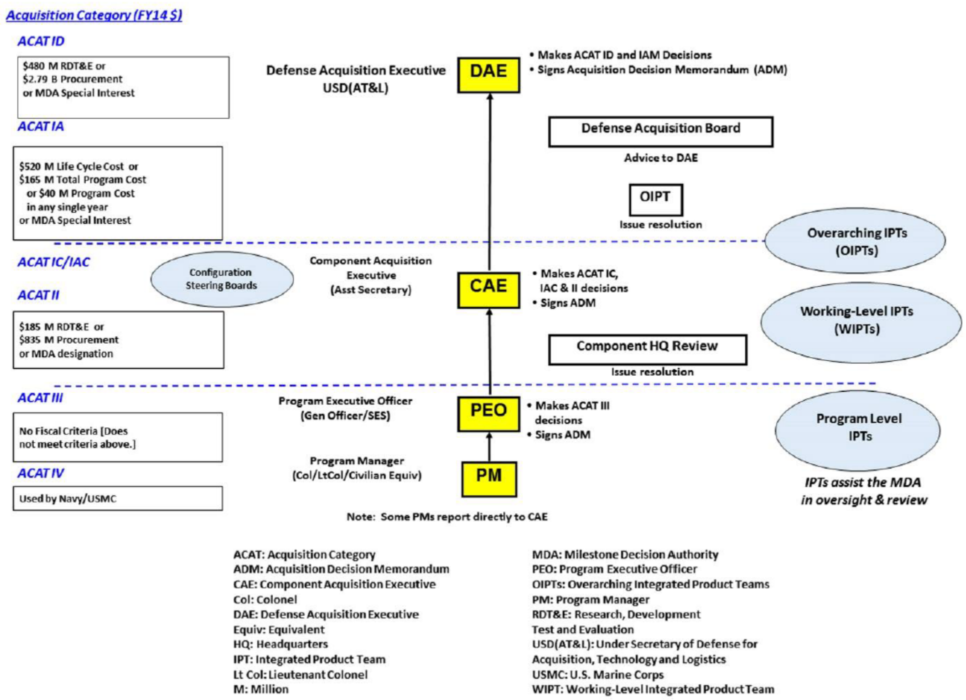

Figure 1. Acquisition Categories
(DoD, 2017)

There are five phases within the Defense Acquisition System:

1. Materiel Solution Analysis (MSA)
2. Technology Maturation and Risk Reduction (TMRR)
3. Engineering and Manufacturing Development (EMD)
4. Production and Deployment (PD)
5. Operations and Support (OS)

Requirements for new or improved capabilities, delivered through JCIDS, drive the acquisition process (DoD, 2015a). Figure 2 illustrates the relationship between the acquisition and capabilities requirement processes and their interaction in the various acquisition phases. This study assumes the capabilities requested from the JCIDS process are accurate and necessary.

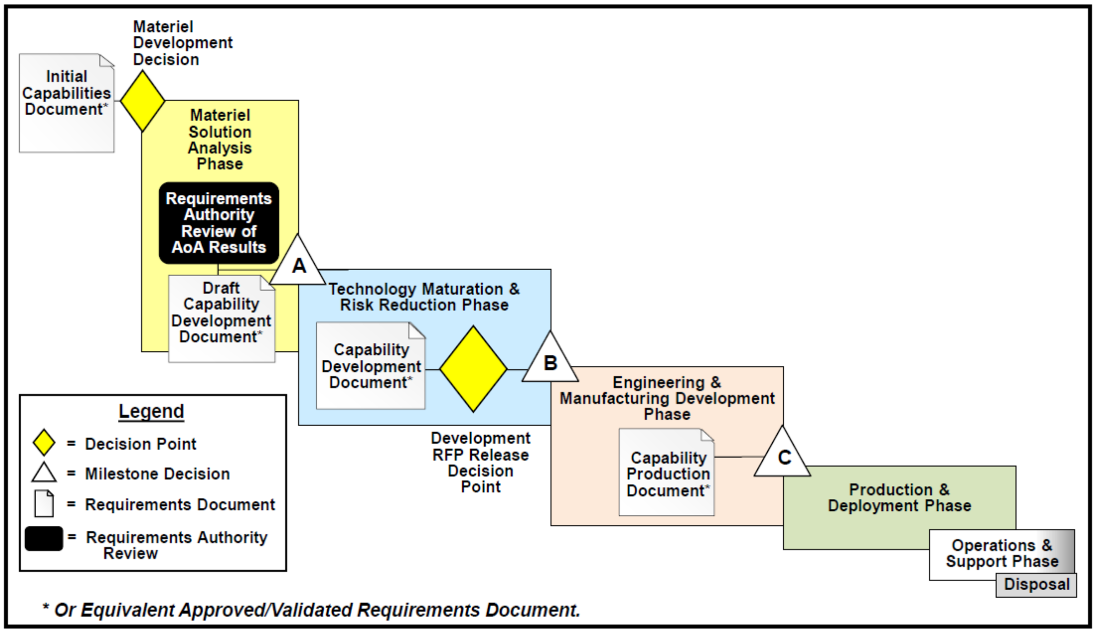

Figure 2. Interaction of Capabilities Requirement and Acquisition Process (DoD, 2015a)

Once an Initial Capabilities Document (ICD) has been validated, the Materiel Development Decision initiates the MSA phase (DoD, 2015a). This decision begins the acquisition process, although an acquisition program is not officially created until milestone B at the completion of the phase (DoD, 2015a). The purpose of the MSA phase is to choose the most promising potential solution for the acquisition process that will fill the needs of the ICD and to establish Key Performance Parameters (KPPs) and Key System Attributes (KSAs) for the system (DoD, 2015a). To accomplish this, an Analysis of Alternatives (AoA) is conducted to determine the suitability of potential acquisitions based on "measures of effectiveness; key trades between cost and capability; total life-cycle cost, including sustainment; schedule; concept of operations; and overall risk" (DoD, 2015a, p. 17). The PM is selected, and the Program Office established during this time (DoD, 2015a). Once the necessary analysis is concluded, the decision authority—usually the Defense Acquisition Executive, head of the DoD component, or Component Acquisition Executive unless otherwise delegated—determines whether the program will continue to the next phase based on the justification for the chosen solution, how affordable and feasible the solution is, how adequate the cost, schedule, and technical risk mitigation plan is, and how effective the acquisition strategy will be (DoD, 2015a). This decision is known as Milestone A (DoD, 2015a). The MSA phase takes a broad look at the potential solutions to a stated need and, as such, may be an appropriate place to consider strategic methodologies like KVA or IRM.

After approval at Milestone A, the program enters the TMRR phase to reduce the risk associated with the technology, engineering, life-cycle cost, and integration of the program to begin the EMD phase (DoD, 2015a). Design and requirement trades occur at this point that are based on the budget, schedule, and likelihood of completion (DoD, 2015a). Guided by the acquisition strategy approved at Milestone A, contractors develop preliminary designs, including competitive prototypes if feasible within the program, to demonstrate the feasibility of their proposed solutions to the Program Office (DoD, 2015a).

Technology Readiness Levels (TRLs) serve as benchmarks that indicate the level of risk associated with a solution reaching maturation per the schedule (DoD, 2015a). Technology Readiness Assessments (TRAs) are a systemic, metric-based method to evaluate the maturity and risk associated with the critical technology in an acquisition program (DoD, 2011). A TRA will assign a TRL for each critical technology in a program, ranging from 1 to 9 from the lowest to highest readiness level (DoD, 2011). Additional methods to assess the likelihood a program will remain on schedule and on budget may be beneficial at this stage, such as IRM. The Development Request for Proposals (RFP) Release Decision Point authorizes the release of an RFP with firm and clearly stated program requirements for contractors to submit their bids (DoD, 2015a). The Preliminary Design Review occurs prior to the completion of the TMRR phase unless waived by the milestone decision authority (DoD, 2015a). Milestone B approves a program to enter the EMD phase and awards a contract while establishing the Acquisition Program Baseline (APB) (DoD, 2015a). The APB describes the approved program, specifically the cost and schedule for the life of the program, and is a formal commitment to the milestone decision authority (DoD, 2015a).

EMD begins once Milestone B is approved. During EMD, the materiel solution is developed, built, and tested to verify all requirements have been met prior to production (DoD, 2015a). Hardware and software designs are completed, and prototypes are built to identify any deficiencies in the design, which will be discovered during developmental and operational testing (DoD, 2015a). DoD acquisitions programs with a contract value greater than $20 million are required by federal regulation to use EVM to track and report the progress of the program, which begins during this phase (DoD, 2019a). Once a stable design that meets the specified requirements has been verified, the manufacturing or software sustainment processes and production capability must be properly demonstrated (DoD, 2015a). Milestone C confirms these requirements are satisfied and approves entry into the PD phase (DoD, 2015a).

The objective of the PD phase is to deliver a product that fulfills the requirements specified in the earlier stages (DoD, 2015a). Initial operational deployment and testing occurs with Low Rate Initial Production (LRIP) for manufactured systems or limited deployment for more software-intensive programs where the system undergoes Operational Testing and Evaluation to verify stated requirements were met (DoD, 2015a). Once satisfied with the fielded systems, full-rate production begins and the product is deployed to operational units (DoD, 2015a). Design changes are limited at this point, although some changes may still occur based on noted deficiencies (Housel, Mun, Carlton, & Jones, 2019). Contracts typically revert to a fixed price strategy during this phase, reducing the PM's focus on cost and schedule variance (Housel et al., 2019).

OS is designed to maintain support for the product and sustain its performance throughout its life cycle, ending with the disposal of the system (DoD, 2015a). OS overlaps with the PD phase since operational units are using the product while production continues, beginning after the production or deployment decision (DoD, 2015a). PMs will sustain the system using the Life Cycle Sustainment Plan (LCSP) developed during the acquisition process, providing the necessary resources and support to keep the system operational (DoD, 2015a). Sustainment and support may include technological upgrades, changes due to operational needs, process improvements, and other activities that may require updates to the LCSP (DoD, 2015a).

There are six different models, four standard and two hybrid, that PMs use to create their program structure, depending on the type of system being acquired (DoD, 2015a). These standard models are templates for hardware-intensive programs, software-intensive programs that are defense unique, incrementally deployed software-intensive programs, and accelerated acquisition programs (DoD, 2015a). As shown in Figure 3, the hybrid models mix the incremental nature of software development within a hardware-centric program. In this model, software development is organized via a series of testable software builds that will culminate with the fully required capability before reaching the Initial Operating Capability (DoD, 2015a). The incremental builds are synchronized with hardware testing requirements for prototypes and other developmental requirements (DoD, 2015a). Other models, with the exception of the accelerated program, use the same basic framework within the five phases.

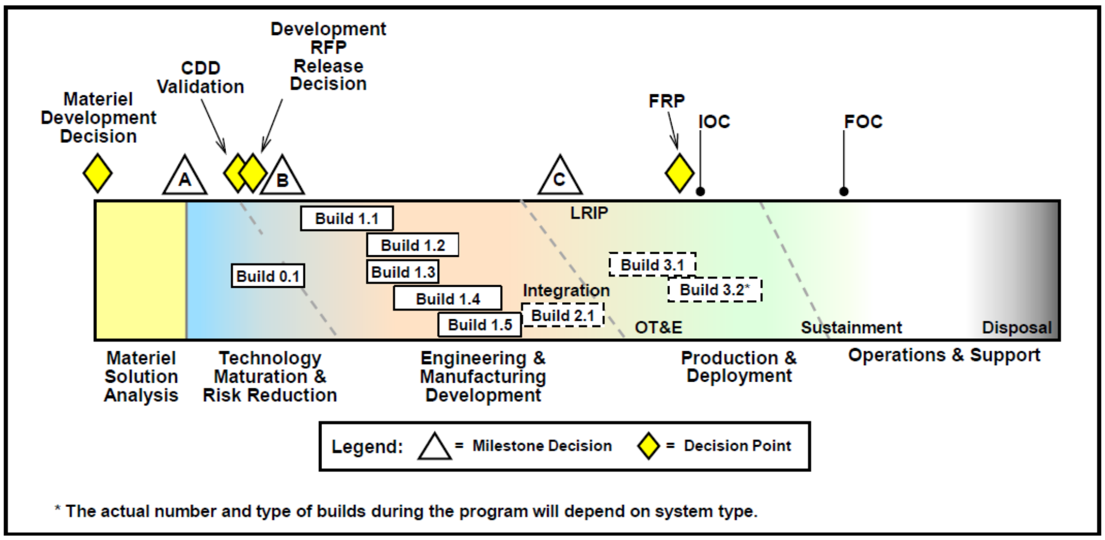

Figure 3. Hardware-Dominant Hybrid Program
(DoD, 2015a)

AI and IT systems are increasingly prevalent throughout the DoD along with their connection to weapon systems, facilities, and Command, Control, Communications, Computers, Intelligence, Surveillance, and Reconnaissance (C4ISR; DoD, 2015b). With the integration comes an increased security risk from adversaries, elevating the importance of effective cybersecurity capabilities and practices (DoD, 2015b). The DoD manages cybersecurity policy through the Risk Management Framework (RMF) by applying security controls founded on risk assessments throughout the life cycle of a system (DoD, 2015b). RMF applies to “all DoD IT that receive, process, store, display, or transmit DoD information” (DoD, 2014, p. 2). Cybersecurity within RMF is more than simply information security, including items such as stable and secure engineering designs; training and awareness for all users, maintainers, and operators of a program; and the response, recovery, and restoration of a system following an internal or external failure or attack (DoD, 2015b). The RMF occurs throughout the acquisition process, and Figure 4 illustrates the six steps of its process.

Artificial Intelligence

Artificial intelligence (AI) does not refer to a specific thing. It is a broad nomenclature for a collection of related inorganic computer science methods used to simulate intelligence.

The term *AI* typically conjures up the nebulous concept of machine learning, which, in reality, is a subset of AI where a computer system is programmed to identify and categorize external real-world stimuli. The DoD's AI strategy defines AI as "the ability of machines to perform tasks that normally require human intelligence—for example, recognizing patterns, learning from experience, drawing conclusions, making predictions, or taking action—whether digitally or as the smart software behind autonomous physical systems" (DoD, 2019). This capability of enhanced automation is of great interest to the DoD as potential future near-peer adversaries, such as Russia and China, are investing heavily in this field for military purposes (DoD, 2019).

Given the large AI field of study, the current research study focuses on the AI processes that are deemed most appropriate for procurement, which may include applications of Machine Learning (ML), Natural Language Processing (NLP), and Robotic Process Automation (RPA). Figure 5 shows that AI itself is a combination of AI sciences, such as ML and NLP, and while RPA benefits from AI application, RPA does not simulate human intelligence, but rather, it just mimics capabilities.

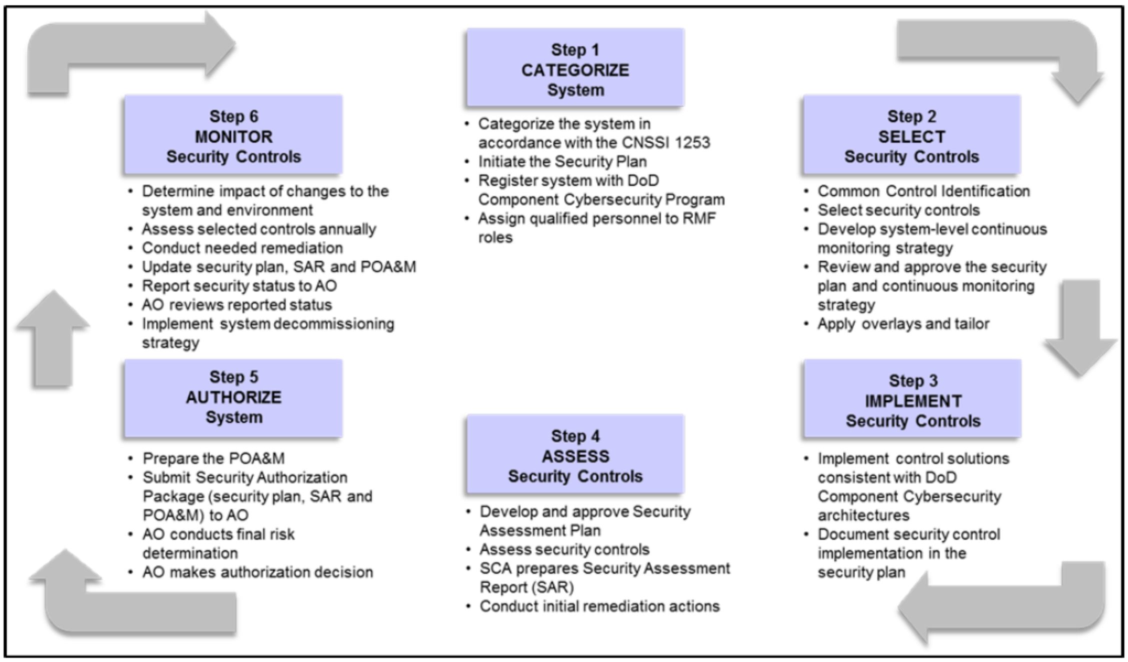

Figure 4. Risk Management Framework Process
(DoD, 2014)

A Brief History of AI

Formally founded in 1956, the science of AI was created to determine if inorganic machines could perform human-level intelligent functions (Denning, 2019). It went through several hype cycles, due primarily to sensationalizing of what it might be able to do, with frequent disappointments. Significant enthusiasm for AI reemerged at the same time that Big Data computing power became more accessible to researchers and companies, which, in turn, could apply the science to multiple tangible applications (Haenlein & Kaplan, 2019). Currently, examples of commercially viable applications of AI exist in manufacturing robots, smart assistants, proactive healthcare management, disease mapping, automated financial investing, virtual travel booking agents, social media monitoring, conversational marketing bots, NLP tools, and contract management (Daley, 2019).

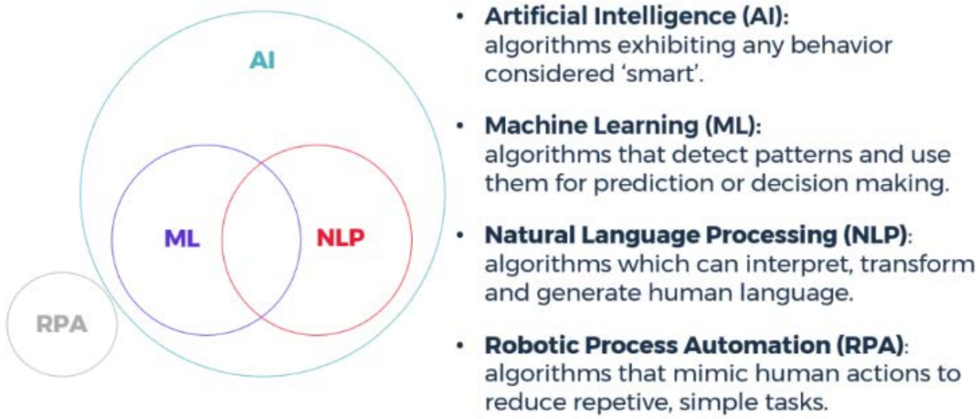

Figure 5. AI Terms and Relationships
(Sievo, 2019)

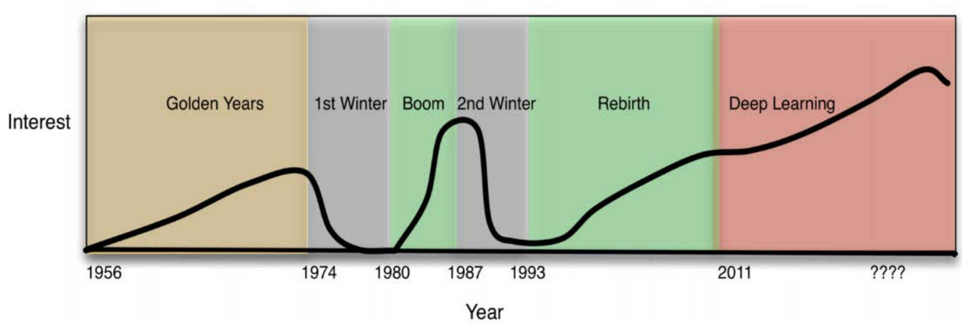

Figure 6. The Timeline of Interest in AI During Different Phases of its Development
(Denning, 2019)

Machine Learning

Intelligence is the capacity to process a certain type of information, allowing a processor to solve problems that are of consequence (Gardner, 1993). Multiple types of intelligence have been proposed by psychologists beyond the classic understanding of a person's intelligence quotient (IQ), which can sometimes only measure how well someone takes an IQ test instead of one's innate abilities. Howard Gardner proposed a theory of multiple intelligence, which suggests that the traditional psychometric views of intelligence are too narrow and should include more categories where certain processors, people in this instance, are stronger than others at making sense of different stimuli. These categories of intelligence include visual-spatial, linguistic-verbal, interpersonal, intrapersonal, logical-mathematical, musical, body-kinesthetic, and naturalistic (Gardner, 1993). An argument against this proposition would be that these categories merely represent learned and disciplined behaviors that someone developed in their life as a result of personality and circumstances. Regardless, both concepts of intelligence (classic and multiple) play well into what steps are taken to make a machine with artificial intelligence.

A computer is capable of performing calculations based on the data provided and ultimately returns a result. It can be programed and coded to follow certain steps or algorithms repeatedly, and even to alter its findings based on its own previously calculated results through some error correction algorithms. A combination of these two steps is the basic concept of machine learning. A computer system is first fed data structured in a way that the algorithm is programmed to recognize, derive patterns from the data, and make assumptions about any unstructured data provided subsequently (Greenfield, 2019). Figure 7 shows how this works in an x-ray learning algorithm.

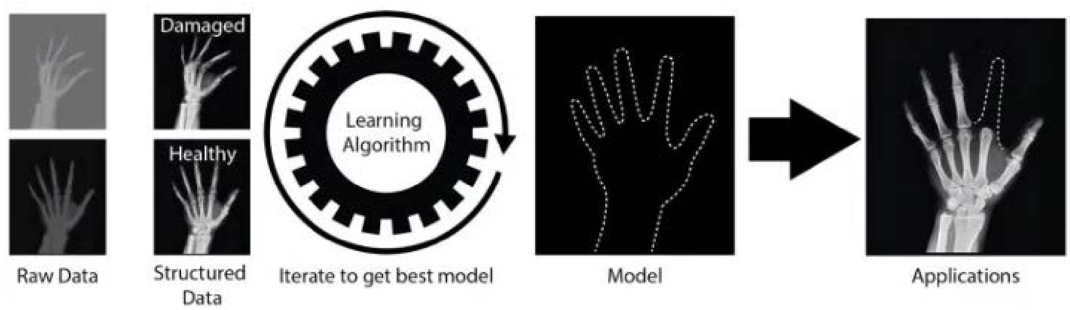

Figure 7. AI Training Algorithm
(Greenfield, 2019)

The image shows the steps an AI algorithm goes through in order to make a recommendation to a physician on where a missing body part should be. It takes in structured data and develops its understanding of what "right" looks like. When given unstructured data, it compares the image against previously trained models and identifies the abnormality with a recommendation on where to apply a fix, such as a prosthetic.

Supervised Learning

An algorithm is taught the patterns using past data and then detects them automatically in new data. Supervision comes in the form of correct answers that humans provide to train the algorithm to seek out patterns in data. This is commonly used within procurement areas such as spend classifications (Sievo, 2019).

Unsupervised Learning

The algorithm is programmed to detect new and interesting patterns in completely new data. Without supervision, the algorithm is not expected to surface specific correct answers; instead it looks for logical patterns within raw data. This is rarely used within critical procurement functions (Sievo, 2019).

Reinforcement Learning

The algorithm decides how to act in certain situations, and the behavior is rewarded or punished depending on the consequences. This is largely theoretical in the procurement context (Sievo, 2019).

Deep Learning

An advanced class of machine learning inspired by the human brain where artificial neural networks progressively improve their ability to perform a task. This is an emerging opportunity in procurement functions (Sievo, 2019).

Human–Machine Partnership

DARPA believes AI integration is critical as a human–machine symbiosis because sensor, information, and communication systems generate data at rates beyond which humans can assimilate, understand, and act on (DARPA, 2019). As was the case in the industrial revolution, machines are better at certain activities and using machines for those activities frees humans to become productive in other areas. Humans and machines excel in separate areas of processing. Consider these comparisons between computers and humans: calculate versus decide; compare versus make judgments; apply logic versus empathize; unaffected by tedious monotony versus having preferences; deals with large data versus intuitional focus on what is most important (Darken, 2019). And while AI performs well in some tasks, it works better with a human partner. Without proper controls, AI is a gullible learning system and can be vulnerable to being deceived by bad actors. Some studies show that AI can be fooled in a way that humans would not be due to human intuition. Other research has been able to fool a self-driving car into thinking a benignly tampered with stop sign was a speed limit sign (see Figure 8), which would undoubtedly lead to collisions if the car were left unsupervised (Eykholt et al., 2018).

Many people are familiar with current intelligent machine partnerships that they, unknowingly, may experience on a daily basis. As discussed with its other applications, Google is the most popular search engine on the Internet because it provides better satisfaction to users than its competitors (Shaw, 2019). Google is so common as the preferred search engine that when someone talks about searching for something online, they refer to it as “Googling.” This is a good example of people interacting naturally with an AI system that uses Bidirectional Encoder Representations (BERT; Nayak, 2019). This is a technique that teaches a machine to not answer the user’s question based on individual words but rather, on their meaning in the context of the question. For example, when asking what time it is right before lunch, the user is not asking for the actual time, but is really asking when they can eat; the outright answer would give the actual time, whereby the asker will deduce eating time, which was the underlying meaning of the question. So-called self-driving vehicles provide another example of human interaction with intelligent machines. For the most part, the user sits in a supervisory role while the car takes over one of the most dangerous events in that person’s life and autonomously conducts all road tasks to drive (Darken, 2019).

If contractors relied on an AI system to make all of the decisions for them, they would be susceptible to purposive misdirection by enemies introducing adversarial information for either competitive advantages or disruptions. Fraudsters can learn how to exploit computer algorithms, but only humans can judge the actions that result. What AI software can do, however, is easily extract data and clarify the content of contracts. It can quickly pull and organize the renewal dates and renegotiation terms from any number of contracts. It can let companies review contracts more rapidly, organize and locate large amounts of contract data more easily, decrease the potential for contract disputes and antagonistic contract negotiations, and increase the volume of contracts companies are able to negotiate and execute (Rich, 2018).

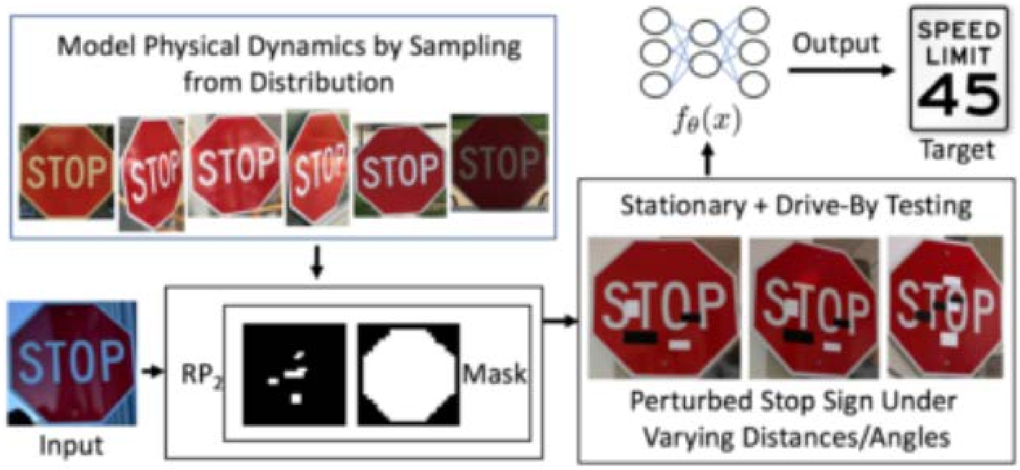

Figure 8. AI System Interpreting a Stop Sign
(Eykholt et al., 2018)

Cloud-Based AI

To understand how AI can be propagated throughout a system and update regulations and learn from multiple human teachers instantly, we look at the concept of cloud computing. The speed of relevance is a popular term in discussing DoD technology adoption. In the 2018 DoD Cloud Strategy, the term *cloud* refers to an offsite physical IT infrastructure. This external infrastructure communicates with a user's computer through the Internet to access data servers that store information and run operating systems, such as Microsoft Windows, which are centrally maintained. This means that every user has the same software computing power and access to the most up-to-date software at all times and is not limited by their organization's IT professional talent or budget for new software. Organizations can get as much, or as little, access to what they need for projects and are unaffected by times of surging need or times of idleness, which currently add excess cost to DoD systems (Shanahan, 2018). The objective for DoD is to have AI-augmented rapid decision-making, in an environment where data is secure and visible for enhanced operational efficiency.

> Data stored in an enterprise DoD cloud will be highly available, well-governed, and secure. Data will be the fuel that powers those advanced technologies, such as ML and AI. This critical decision-making data will be made available through modem cloud networking, access control, and cross domain solutions to those who require access. Common data standards will be a key part of the Department's methodology for tagging, storing, accessing, and processing information. Ensuring an enterprise cloud environment will increase the transparency of this data, and drive the velocity of data analysis, processing, and decision making. Leveraging advances in commercial cloud security technologies will ensure the Department's information is protected at the appropriate level. (Shanahan, 2018, pp. 5–6)

Private Sector AI Application to Contracting

To compare solutions to the DoD acquisition problem, we examine similar circumstances in the U.S. private sector. Lawgeex is an example of a company that is applying the AI integration process in the private industry procurement world. It demonstrated that its AI software could outperform U.S. trained lawyers on an example contract aspect, the Non-Disclosure Agreement (NDA), with an average accuracy of 94% as compared to 85% for humans (Lawgeex, 2018). The study was conducted to respond to a common business problem in large companies that rely on contracts to engage with partners, suppliers, and vendors having an 83% dissatisfaction rate with their organization's contracting processes (Lawgeex, 2018). Another example is Icertis, a company that services large and commonly familiar companies such as 3M, Johnson & Johnson, and Microsoft, to list a few (Icertis, 2019). Icertis provides its customers with a cloud-based AI platform that learns from contracts provided by the client, along with control measures, to create and assist in contract setup; contract operations; governance, risk, and compliance; and reporting (Icertis, 2019).

What makes this possible now, instead of when it was first theorized decades ago, is that industry is more accustomed to storing professional documents on a digitally accessible storage platform, whether local hard drives or the cloud (Betts & Jaep, 2017). Currently, the major hurdles that prevent a fully automated contract review and analysis process are nontechnical, such as the collection of contract performance data; publication of private contracts and their corresponding performance data; and changes in ethical constraints on computer usage in legal practice (Betts & Jaep, 2017). The authors of these obstacles also offer possible policy solutions to address them: start using contract management software that will be a forcing function to create data in an AI teachable format; expand copyright protection for vendors to protect their intellectual property; and create new rules to help mitigate AI risks to enable its ability to work (Betts & Jaep, 2017).

Methodologies

Earned Value Management

EVM offers numerous advantages to PMs using the methodology to track the progress of a project. When properly implemented, it provides an early warning of potential issues within a program and a forecast of the total cost and schedule requirements (Fleming & Koppelman, 2010). Research has shown that the CPI (cost performance index) stabilizes within a 10% range after a project reaches the 20% completion point (Christensen, 1998). The 20% stabilization point holds true across various contract types, programs, and services (Fleming & Koppelman, 2010). This early indication provides PMs a reliable prediction of the project's final costs (Fleming & Koppelman, 2010). The low range of the cost overrun is the current cost plus the remaining scheduled cost (Fleming & Koppelman, 2010). The high end of the costs is the budgeted cost divided by the CPI, which is a more accurate estimate unless extenuating circumstances caused the overrun (Fleming & Koppelman, 2010). SPI (schedule performance index) also provides an indication of future cost increases. An unfavorable SPI indicates spending will grow larger than initially planned to reduce the schedule variance (Christensen, 1998).

Each component has a cost and schedule associated with its completion that feed into the overall completion of the project. Since the baselines are established prior to work beginning on any subcomponents, management can create well in advance a detailed timeline with the expected cost necessary to achieve the tasks. Having such a thorough plan allows the PM to focus on the areas that are reporting discrepancies in their CPI and SPI, without the benefits of a quantitative value metric, rather than concentrating on all areas of the project. They can interact when there are issues with a component, trusting the aspects of a project that are on schedule and on budget to remain performing well, that is, producing an acceptable level of value. Instead of actively controlling all parts of the project, they can spend their limited time correcting issues with portions that have unfavorable SV and CV. The key metrics allow PMs to manage by exception.

Using a WBS to assign, track, and complete tasks facilitates the accurate and timely reporting of a project's cost and schedule performance, but not its value to cost performance. The use of the existing common EVM approach facilitates communication between the PM and the contractor with regard to cost and schedule. Since the individual components of a project are broken down into small subsections that must be completed, management can more easily communicate with contractors concerning discrepancies with subcomponents. This breakdown allows a PM to assign resources and facilitates subsequent tracking of a system's cost and schedule progress without creating additional reporting requirements. Given that reporting on the status of a program through EVM is required by law, managing with the same system reduces administrative requirements (Christensen, 1998). Managing from one methodology while reporting from a different system is a more difficult and costly way to conduct business (Christensen, 1998). EVM gives managers a single system to track cost, completion, and project schedule performance, centralizing the control system while allowing lower level managers to oversee their sections with the same limited metrics.

Simplicity is a key attribute of the EVM methodology. The process is broken down into three main variables: money, time, and work completed. The key metrics managers use to assess a project are simple ratios derived from these cost and schedule parameters. Determining the performance of the project is as easy as determining how much money has been spent to complete the current amount of work at a given time. The methodology readily scales from the overall project to individual components without changing reporting requirements; managers simply compile the component reports into an aggregate report. However, in AI programs, it can be more difficult to determine if the work is completed satisfactorily (i.e., did it produce expected value?) and in accordance with the stated requirements, as seen in the JTRS acquisition.

One of the most important attributes of EVM is the combination of schedule and cost performance. Traditional cost management approaches reflect a project's funding performance and not the true cost of the program (Fleming & Koppelman, 2010). Flexible budgets also give management the option to vary the budget for work based on the actual valued output level (Christensen, 1998). EVM differs from a flexible budget by including the time dimension in its calculations. It puts the schedule variance into a cost dollar amount, quantifying delays. The amount of money spent completing a project is important, but it does not show the complete picture. If an initiative is on or under budget yet is delivered seven years later than promised, it did not perform well despite the potential cost savings. Cost monetizing the scheduled time gives the PM the ability to determine how much a delay will cost, providing greater flexibility within management decisions. However, this approach does not guarantee that value has been provided at an acceptable returns level.

EVM has proven to be an effective management system in traditional manufacturing processes. However, in FY2008, the U.S. federal IT portfolio contained 346 major IT programs worth approximately $27 billion that received a rating of "unacceptable" or are on the "Management Watch List" (Kwak & Anbari, 2012). This suggests there is significant room for improvement within the current AI acquisition process. AI projects are not developed in the same manner as projects that do not process variable information. When building a warehouse, a contractor uses the architect's blueprints to determine exactly how many bolts will be needed for each beam and how long it will take to install each item. Combining each subprocess, which consist of known cost/schedule parameters, the manager can establish a baseline with a high degree of accuracy. However, when a program involves writing code, the process is more complicated due to the relative creativity of the programmers to generate elegant code. Keeping with EVM principles, the desired outputs for a highly structured program are known before work begins and the plan to accomplish those steps can be created and mapped out prior to construction of the code. However, the time required to write, test, debug, and retest the computer code can vary significantly between projects and individuals, leading to significant variability. Using the EVM system that measures the overall progress of a program based on the completion of subcomponents compared to a baseline established well before work began is not an accurate assessment for projects such as AI programs that do not progress linearly, but rather, progress iteratively with highly variable production processes.

The requirements in large AI projects are often not well defined, leading to cost and schedule overruns. For instance, JTRS (joint tactical radio system) used LOCs (lines of communication) to create the baseline. However, LOC can vary drastically from one programmer to another or between programming languages. While there are industry standards for estimating the number of LOCs, there is still variance between each individual creating a software program. In a traditional manufacturing application of EVM, a WBS containing the task “turn on the light” would have schematics and plans associated with the task so that any qualified worker could complete the work. In an AI program, turning on the light via a software program could be completed in any number of ways and still meet the specifications. This simplistic example illustrates issues that will develop when using EVM to manage a more complicated IT-based project, such as creating a software defined waveform.

One of the core principles of EVM is maintaining the baseline of a program throughout the life of the project, allowing for consistent and accurate measurement of a project’s cost/schedule progress. Adding requirements or changing the specifications of certain components can alter the trajectory of a program entirely. While any program will need to make adjustments when scope creep occurs, EVM is particularly ill-suited to make these modifications midstream for AI-type projects. If new requirements are added to a program or existing requirements changed after the baseline is established, especially after work commences, the baseline is no longer useful as a point of comparison. More often than not, there will be additional costs and longer timelines required to fulfill the new specifications. This will change the ACWP while leaving the BCWS the same. As the actual costs increase and schedule expands compared to the baseline, the main performance metrics (CV, CPI, SV, and SPI) are negatively impacted, even if the newly added requirements add value or are completed on time and on budget. EVM does not function well in projects with changing scope or requirements.

When developing JTRS, much of the software had never been written or created, calling on developers to estimate the cost and schedule for the various components. To account for the possibility it may not be completed in time, the engineer or PM may add one additional week to something originally thought to take three weeks, accounting for some risk of the unknown. Should overruns occur in other areas, PMs often remove this risk factor from their calculations to improve their performance compared to the baseline. This omission eliminates much of the risk mitigation the program originally established, eventually leading to further discrepancies to the original CV and SV estimates. However, when creating a complex product based on the assumption of future technological advancements, as in JTRS, there is a high degree of risk. Presuming all tasks will be completed on schedule without additional schedule risk mitigation is a misguided assumption to make after previous components of the project have already fallen behind schedule.

EVM functions well as a tool to monitor the cost and schedule of programs developed in waterfall or parallel design methods. In alternative design methodologies, such as agile or iterative designs, EVM does not provide the same level of flexibility to project managers. EVM does not require a particular development approach and can be used in any system that uses a baseline plan specifying schedule and cost (DoD, 2018). Techniques such as agile EVM and scrum attempt to use a more iterative approach typical of the software development process but are still using the same EVM concepts with tasks broken into various work packages. By assigning a budget and timeline to specific features within an AI initiative, PMs may use EVM to oversee progress. As each feature is completed, value is earned and the EVM metrics are updated. However, within the specific features and components of a program, EVM is not as useful. As previously mentioned, the specifics of writing computer code are not as cut and dry as a typical well-structured physical product project. If two capabilities written in parallel both work individually, it is still possible there will be issues when combining the features in the final project output. To reduce this risk, PMs feel forced to add time in their schedule. This can be done by scheduling multiple increments that are planned, designed, coded, tested, and demonstrated (DoD, 2018). While this is a viable way to use EVM in agile design, it is not as accurate or precise as an EVM program in a brick-and-mortar, physical product type project.

As discussed above, the metrics within EVM are simple to understand and use to make management decisions regarding cost and schedule. However, for the metrics to be accurate, the cost and schedule must be accurate to a great level of detail. WBSs for multibillion-dollar programs often cover all items from the strategic overview down to small tasks within the project. Creating such a detailed list of requirements for each component within a system is both timely and costly. While the result may be a simple schedule with easily discernible metrics, the initial setup process to establish the baseline can be immense.

Knowledge Value Added (KVA)

KVA is an objective, quantifiable method to measure the value produced by a system and the subprocesses within the system. The value measurements of each process are ratio scale numbers, allowing analysts to compare them with the values from other subprocesses to determine their relative effectiveness. KVA converts the outputs of all processes into common value units allowing a standard productivity performance ratio, that is, output/input across all processes. PMs can determine the value generated from the human component against the value added by IT processes. Because of the scales, PMs can use these measurements to develop useful ratios in their analysis of the program's performance. Productivity ratios include return on knowledge (ROK), where the output of a process is divided by the process cost required to produce the output, and return on investment (ROI), which is the monetized output minus cost divided by cost. The ROKs and ROIs, which are always 100% correlated, give managers information about the amount of value a process generates compared to the amount of money spent to create the value. Unlike any other methodology, KVA assigns these figures to both the process and subprocesses rather than only the firm as a whole (as is done in standard generally accepted accounting practice).

Conducting an analysis of a program using KVA will give a PM a clearer picture of the value of the operational components of the program. While organizations likely have cost/schedule metrics used to determine the performance of a project or operation, ROK will give them additional value-based information to improve their management decisions. PMs can determine the relative value of the components that comprise the program. Knowing a particular job or subprocess gives the same output value as a different process but at a different cost may provide context for understanding the variations in the performance of the overall system. This, in turn, gives experienced managers the information needed to allocate resources to specific components of a program that need improvement or should be utilized more frequently, resulting in greater value added.

While a KVA analysis can provide information that will change the course of a program or project, it does not require significant changes to organizational structure or reporting processes to do so. The evaluation can be conducted during normal operating conditions without introducing complicated new metrics into the system. The learning time, process description, or the binary query method are all based on information that should be available within the organization. A small amount of hands-on measurement may be required to verify the accuracy of the given data. As such, the analysis can be done quicker than the other methodologies, giving PMs access to actionable information more rapidly.

KVA will give analysts a quantifiable, ratio-scale number for the value of the subprocesses. However, it does this only with processes that consist of known a priori outputs. The intangible items, such as creativity and imagination, that occur within the human brain cannot be quantified with this method, or any other method for that matter. In fact, no current system is able to accurately quantify these types of intangibles within a process because there is no algorithm for creativity. These factors are not common to the average user and as such, cannot be defined via any of the KVA methods—learning time, binary query, or process description—because the creativity process cannot be learned or described. However, this was only possible after the system was completed and described. KVA will assign the value of the process but it cannot predict the value of potential outputs, only those that are specified a priori.

Although KVA will provide ratio-scale numbers to aid in evaluating processes within a program, the ratios are often only valid for comparisons within the same analysis. Benchmarking the raw numbers with other organizations or with different divisions in the same organization will provide a usable comparative performance assessment depending on the techniques used when determining the ROK. Regardless of the language of description for outputs, the resulting ROK and ROI measures will be comparable just as they are among industries that produce different products or services. Because these descriptions of outputs are in common units, they can be treated as constants across all processes. The final results of any properly conducted analysis will return comparable ROK and ROI estimates, which is the ultimate goal of KVA.

Integrated Risk Management

All organizations depend heavily on project planning tools to forecast when various projects will complete. Completing projects within specified times and budgets and a given value is critical to facilitate smooth organizational operations. In our high-technology environment, many things can impact schedule. Technical capabilities can often fall short of expectations. Requirements may be insufficient in many cases and need further definition. Tests can bring surprising results—good or bad. A whole host of other reasons can lead to cost increases, schedule slips, and value variability. On rare occasions, we may run into good fortune and the schedule can be accelerated without harm to the productivity of the project outcome.

Project schedules are inherently uncertain, and change is normal. Therefore, we should expect changes and find the best way to deal with them. So why do projects always take longer than anticipated? One reason is inaccurate schedule estimating. The following discussion presents a description on shortcomings in the traditional methods of schedule estimation and how simulation and advanced analytics can be applied to address these shortcomings.

It is important to first understand the Integrated Risk Management (IRM) process and how the techniques involved are related in a risk analysis and risk management context. This framework comprises eight distinct phases of a successful and comprehensive risk analysis implementation, going from a qualitative management screening process to creating clear and concise reports for management. The process was developed by the author (Dr. Mun) based on previous successful implementations of risk analysis, forecasting, real options, valuation, and optimization projects both in the consulting arena and in industry-specific problems. These phases can be performed either in isolation or together in sequence for a more robust integrated analysis.

We can segregate the process into the following eight simple steps (Mun, 2016):

1. Qualitative Management Screening
2. Forecast Predictive Modeling
3. Base Case Static Model
4. Monte Carlo Risk Simulation
5. Real Options Problem Framing
6. Real Options Valuation and Modeling
7. Portfolio and Resource Optimization
8. Reporting, Presentation, and Update Analysis

Comparison of Key Attributes

Choosing a methodology should depend on the nature of the project under consideration, specifically, the commitment needed from the organization, the organization's desire to align strategic goals with the project, the predictive capability of the methodology, the flexibility required, and the time available. While others in the organization need to understand the concepts to comprehend status reports, EVM only needs the management team to track the cost and schedule of the project compared to the baseline as there is no goal alignment with the organization. While the CPI and SPI can help estimate the final cost and schedule, there is no true predictive ability associated with EVM since the assumption is that the schedule will proceed according to the baseline, regardless of previous performance. Adherence to the baseline is essential in EVM, and changing requirements can drastically alter a baseline, reducing the effectiveness of the methodology. Setting up, monitoring, and reporting the cost/schedule performance of each work package within the WBS can be a time-consuming and expensive task. KVA needs only the KVA analyst and the process owner, as the subject matter expert, to determine the value of a process's output, supporting the need to align the project with an organization's productivity goals. Using this analysis, they can model the current baseline as-is process ROK and compare it with the proposed to-be process model ROK, thus offering a simple prediction of the improvement between the models. Since KVA can be used with any language of description to define the process outputs, analysts can choose whichever method is most beneficial for the particular system in question, providing flexibility. This analysis can be completed quickly, potentially providing a rough-cut assessment within a few days. IRM requires the organizational leadership, portfolio and project managers, and the analyst to determine how a project fits within an organization's portfolio, the present value (PV) of the project, and potential real options. By analyzing and simulating various scenarios, IRM provides a prediction of a project's likely performance, which allows managers to build in flexibility via real options at the appropriate locations within the project. Assuming the data necessary for the analysis is available, the process can be completed in a relatively quick manner.

Methodologies in AI Acquisition

As previously discussed, the methodologies all have strengths and weaknesses, making them more suitable in certain applications than others. The biggest challenge in using EVM when acquiring AI is the iterative nature of software development. EVM needs clearly stated, detailed requirements for intermediate steps to be most effective. While the outputs of software programs are defined well, the steps required to build the software are not, leading to issues when developing cost and schedule estimates. If the software is not complex or consists of known processes, EVM can sufficiently monitor the progress. Integrating software and hardware is also complicated with EVM since there are numerous pieces of the program that must be combined to meet the goals, resulting in additional debugging and recoding. EVM is more efficient when used to manage the physical creation of systems or infrastructure. It can monitor the cost/schedule progress of software work packages but is not as useful at estimating the value of those programs until the requirements have been delivered. KVA can provide an objective, ratio-scale measure of value and cost for each core process and its subprocesses within any IS system. Using the two measurements, managers can then analyze productivity ratios information, such as ROK and ROI, to determine the efficiency of a process compared to the resources used to achieve the output. This can help the manager decide how to use resources to update systems or estimate the future value of a system being acquired. Combining the KVA results with IRM allows managers to iterate the value of system real options analysis through simulation and other techniques. IRM can also quantify risks and forecast performance probabilities for measures of the potential success for programs and components of programs using historical data. It is a tool to assist with the investment strategy, making it useful when acquiring all types of AI. However, it is not designed to help manage the actual acquisition of an AI program or to determine how to meet its detailed requirements.

Summary

Examining the benefits and challenges of the proposed methodologies demonstrates the scope, capabilities, and limitations of various AI systems. It also helps inform in which areas and phases of the Defense Acquisition System life cycle it may be appropriate to include the methodologies or components of the methodologies within the system. The next section provides recommendations based on these findings.

Conclusion

The main question of this research was, simply, how can certain advanced analytical decision-making methodologies be used in the acquisition life cycle to complement existing methods to ensure a successful acquisition of AI technologies?

As discussed, EVM remains the only program management methodology required by the U.S. government for all DoD acquisition programs with a contract value exceeding $20 million. Regardless of this requirement, EVM is a methodology that provides a structured approach to the acquisition of IT via program management processes that can help ensure an acquisition program stays on schedule and within budgeted cost estimates. However, there are significant limitations when using EVM for AI acquisitions, the major weakness being that it was not designed for managing AI acquisitions that follow a very iterative pathway. Organic AI acquisitions require a given level of flexibility to deal with the unknowns that arise during the development process. In addition, EVM does not provide a common unit of value metric to enable standard productivity metrics, such as ROI. When value is inferred by how consistent a program is with original baseline cost and schedule estimates, the performance of the program may be sacrificed in terms of the quality of the outputs when planned program activities become iterative, as in the development of many AI programs. If an AI acquisition program is trending toward cost and schedule overruns, but the resulting value added of the modifications to the original requirements provides disproportionate increases in value, EVM is not designed to recognize this increase in value.

To remedy these shortcomings of EVM in AI acquisitions, the methodology should be combined with KVA and IRM, which can be useful during the requirements phase of EVM by ensuring that a given AI acquisition is aligned with organizational strategy and that a baseline process model has been developed for establishing current performance before acquisition of the supporting AI. A future process model that estimates the value added of the incorporation of the AI can also set expectations that can be measured against the baseline model after the AI has been acquired. IRM can be used to forecast the value of strategic real options flexibility that an acquired AI may provide so that leadership can select the options that best fit their desired goals for the AI in defense core processes.

Because it provides an objective, quantifiable measure of value in common units, KVA should also be used in AI acquisitions to allow decision-makers to better understand and compare different strategic options based on their value and the cost. Obtaining a return on investment of AI systems can only be done when using KVA to determine the value embedded in the system. This information provides insight to PMs as well as a more complete perspective regarding the performance of both the current and the to-be systems.

Likewise, using IRM is recommended when acquiring AI through the Defense Acquisition System. Applying dynamic and stochastic uncertainty and risk-based modeling techniques to predict likely and probabilistic outcomes can improve the risk estimates associated with the components and subcomponents of a program, in terms of their potential cost overruns, value variabilities, and schedule delays. Analyzing various real options within the context of the models' outputs will help PMs make the most advantageous choices when determining a program's future.

PMs should use EVM only in the EMD phase, as is currently done. That said, EVM will work best in hardware manufacturing solutions with technology that is fully mature prior to the program starting. Since many AI acquisition programs consist of advancing the current technology and developing new software solutions to meet requirements, EVM is not perfectly suited for AI development. Nevertheless, PMs can use various agile EVM techniques to complete projects on cost/schedule/value baselines provided the appropriate steps are taken when establishing the baseline. Requirements must be broken into small, easily definable tasks with suitable risk and uncertainty factors accounted for within the schedule. Other methodologies should be used with EVM to ensure these factors are based on defendable metrics rather than simply guessing how much additional time, money, and value may be necessary to complete complex tasks.

During the MSA phase, KVA will help determine the value of the different options considered in the analysis of alternatives (AoA) process. KVA can objectively measure the value of the current, as-is system and the potential to-be systems under consideration. Using other factors such as cost, value, complexity, timeline, and so forth, the PM can then select an appropriate alternative. As the chosen solutions mature during the TMRR phase, an updated KVA analysis will reassess initial estimates and provide a projected ROI for the AI solution prior to entering the EMD phase. In the OS phase, KVA will help decision-makers establish how a program is performing and use that information to make any adjustments or corrections that may be needed. KVA has limited prediction capabilities, so it should be used in conjunction with other methodologies, particularly IRM, to obtain the most benefit.

IRM techniques should be implemented during most of the acquisition phases. Ideally, portfolio management decisions were made during the requirements development process, although they should also be considered during MSA. Financial and value analysis derived from KVA, as well as simulation of possible outcomes should occur during the MSA, TMRR, and EMD phases. The results of these simulations should be fed into the EVM baselines to account for risk across the program. Real options should be developed during the TMRR phase prior to awarding contracts, and the real options should be executed during the EMD and PD phases as appropriate.

Limitations and Future Research

This research examined whether the various methodologies—EVM, KVA, and IRM—could be used within the Defense Acquisition System to improve the acquisition of AI. Future research should examine how these methodologies may interact with or improve other components of the acquisition system. This includes the JCIDS and PPBE components as individual processes and the interaction of JCIDS, PPBE, and the Defense Acquisition System as a whole. Certain methodologies, specifically IRM, may be more beneficial when used throughout the entire acquisition process instead of within a portion of the system. Additionally, future research could examine how these different methods may be used in the acquisition of products outside the AI or IT realm.

The research conducted looked at AI as a whole and not specific types of AI. Future studies should examine if acquisition methods, strategies, and methodologies should change based on the category of AI being acquired. This is of specific interest when considering artificial intelligence and its subsets. Machine learning, intelligence with a specific focus or field of expertise, and general or universal intelligence would likely have different methods used in the acquisition process based on their complexity, complicated nature, undeveloped technology, and level of risk.

The applicability of these methodologies within commercial acquisition of AI is another area of potential research. This research focused exclusively on the application of the respective techniques within the DoD acquisition process. However, commercial entities also struggle when acquiring complex or complicated AI and IT systems, particularly when the systems operate at the enterprise level. Further research may indicate if these same methodologies could provide value to decision-makers in the private sector during the creation, adoption, or customization of commercial AI.

Finally, this research only examined the most promising methodologies out of numerous different possibilities. Future research could examine other program management tools, management philosophies, analytic tools, or other methodologies and their benefit when acquiring AI. While the examined methodologies were chosen because they would likely benefit the process and support improvements in EVM, other systems may be more appropriate in certain phases or may offer additional benefits not seen in this research.

References

Betts, K. D., & Jaep, K. R. (2017). The dawn of fully automated contract drafting: Machine learning breathes new life into a decades-old promise. *Duke Law and Technology Review*, *15*(1), 216–233.

Christensen, D. S. (1998). The costs and benefits of the Earned Value Management process. *Acquisition Research Quarterly*, *5*(4), 373–386. Retrieved from https://www.webharvest.gov/peth04/20041016215255/http://www.dau.mil/pubs/arq/arq98.asp

Daley, S. (2019, September 24). 19 examples of artificial intelligence shaking up business as usual. Retrieved from https://builtin.com/artificial-intelligence/examples-ai-in-industry

Darken, R. (2019, October 21). *Human-machine teaming AI*. Monterey, CA: Naval Postgraduate School.

DARPA. (2019). AI next campaign. Retrieved from https://www.darpa.mil/work-with-us/ai-next-campaign

Denning, P. (2019, September). *Harnessing artificial intelligence*. Monterey, CA: Naval Postgraduate School.

DoD. (2003, May 12). *The defense acquisition system (change 2)* (DoD Directive 5000.01). Washington, DC: Author. Retrieved from https://www.esd.whs.mil/Directives/issuances/DoDd/

DoD. (2011, April). *Technology Readiness Assessment (TRA) guidance*. Washington, DC: Assistant Secretary of Defense for Research and Engineering.

DoD. (2014, March 12). *Risk management framework (RMF) for DoD information technology (IT) (change 2)* (DoD Instruction 8510.01). Washington, DC: Author. Retrieved from https://www.esd.whs.mil/Directives/issuances/DoDi/

DoD. (2015a, January 7). *Operation of the defense acquisition system (change 4)* (DoD Instruction 5000.02). Washington, DC: Author. Retrieved from https://www.esd.whs.mil/Directives/issuances/DoDi/

DoD. (2015b, September). *DoD program manager's guide for integrating the cybersecurity risk management framework (RMF) into the system acquisition life cycle*. Washington, DC: Author.

DoD. (2017, February 26). *Defense acquisition guidebook*. Retrieved from https://www.dau.mil/tools/dag

DoD. (2019, January 18). *Earned Value Management Implementation Guide (EVMIG)*. Washington, DC: JAuthor.

DoD. (2019, February 12). Summary of the 2018 Department of Defense artificial intelligence strategy: Harnessing AI to advance our security and prosperity. Retrieved from https://media.defense.gov/2019/Feb/12/2002088963/-1/-1/1/SUMMARY-OF-DOD-AI-STRATEGY.PDF

Eykholt, K., Evitimov, I., Fernandes, E., Li, B., Rahmati, A., Xia, C., & Song, D. (2018). *Robust physical-world attacks on deep learning visual classification*. Ithaca, NY: Cornell University.

Fleming, Q. W., & Koppelman, J. M. (2010). *Earned Value project management* (4th ed.). Project Management Institute. Retrieved from https://app.knovel.com/hotlink/toc/id:kpEVPME001/earned-value-project/earned-value-project

Gardner, H. (1993). *Multiple intelligences.* New York, NY: Basic Books.

Greenfield, D. (2019, June 19). Artificial intelligence in medicine: Applications, implications and limitations. Retrieved from Harvard University at http://sitn.hms.harvard.edu/flash/2019/artificial-intelligence-in-medicine-applications-implications-and-limitations/

Haenlein, M., & Kaplan, A. (2019). A brief history of artificial intelligence: On past, present, and future of AI. *California Management Review, 61*(4), 5–14.

Housel, T. J., Mun J., Jones, R., & Carlton, B. (2019, May). *A comparative analysis of advanced methodologies to improve the acquisition of information technology in the Department of Defense for optimal risk mitigation and decision support systems to avoid cost and schedule overruns*. Paper presented at Acquisition Research Symposium, Monterey, CA.

Icertis. (2019, April 2). Microsoft streamlined its contract management. Retrieved from Icertis at https://www.icertis.com/customers/microsoft-information-exchange-agreements-case-study/

Kwak, Y., & Anbari, F. (2011). History, practices, and future of earned value management in government: Perspectives from NASA. *Project Management Journal, 43*(1), 77–90. Retrieved from https://doi:10.1002/pmj.20272

Lawgeex. (2018, February). Comparing the performance of artificial intelligence to human lawyers in the review of standard business contracts. Retrieved from Law.com at https://images.law.com/contrib/content/uploads/documents/397/5408/lawgeex.pdf

Mun, J. (2016). *Real Options Analysis* (3rd ed.). Dublin, CA: Thomson-Shore and ROV Press.

Mun, J., & Housel, T. (2010). *A primer on applying Monte Carlo simulation, real options analysis, knowledge value added, forecasting, and portfolio optimization.* Monterey, CA: Naval Postgraduate School.

Nayak, P. (2019, October 25). Understanding searches better than ever before. Retrieved from https://blog.google/products/search/search-language-understanding-bert

Rich, B. (2018, February 12). How AI is changing contracts. *Harvard Business Review.* Retrieved from https://hbr.org/2018/02/how-ai-is-changing-contracts

Shanahan, P. (2018). *DOD cloud strategy*. Washington, DC: DoD.

Shaw, M. (2019, October 15). Why Google is the best search engine (and why businesses should care). Tower Marketing. Retrieved from https://www.towermarketing.net/blog/google-best-search-engine/

Sievo (2019, November 16). AI in procurement. Retrieved from https://sievo.com/resources/ai-in-procurement

Acquisition Research Program
Graduate School of Defense Management
Naval Postgraduate School
555 Dyer Road, Ingersoll Hall
Monterey, CA 93943

www.acquisitionresearch.net