NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

ANALYZING THE CHALLENGES AND OBSTACLES TO DEVELOPING AND FIELDING AUTONOMOUS AND SEMI-AUTONOMOUS SYSTEMS

by

Mario D. Williams and Phillyp C. Lawson

June 2020

Thesis Advisor:	Robert F. Mortlock
Co-Advisor:	Raymond D. Jones

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE				*Form Approved OMB No. 0704-0188*

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington, DC 20503.

1. AGENCY USE ONLY ***(Leave blank)***	**2. REPORT DATE** June 2020	**3. REPORT TYPE AND DATES COVERED** Master's thesis

4. TITLE AND SUBTITLE ANALYZING THE CHALLENGES AND OBSTACLES TO DEVELOPING AND FIELDING AUTONOMOUS AND SEMI-AUTONOMOUS SYSTEMS	**5. FUNDING NUMBERS**
6. AUTHOR(S) Mario D. Williams and Phillyp C. Lawson	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000	**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A	**10. SPONSORING / MONITORING AGENCY REPORT NUMBER**

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release. Distribution is unlimited.	**12b. DISTRIBUTION CODE** A

13. ABSTRACT (maximum 200 words)

Autonomous and semi-autonomous systems operate on a system of systems (SoS) framework, utilizing their own ability to sense, perceive, analyze, and execute actions to achieve their goals. Unmanned systems present a significant challenge for the Department of Defense's (DoD) acquisition process that was established to develop and field man-in-the-loop type capabilities. The purpose of our thesis is to perform an analysis on the challenges of developing semi-autonomous and autonomous systems through the military acquisition process in order to determine best practices and trends necessary to increase the likelihood of program success.

The secondary goal of our analysis is to compare and contrast Test & Evaluation methodologies for systems with autonomous capabilities. The purpose of the test and evaluation process is to enable decision makers to manage technical risk and assess how robust and mature capabilities are prior to making a fielding decision. Autonomous systems require a rigorous test/build strategy that for most programs lead to cost overruns and schedule breaches. In addition, trying to keep pace with rapidly changing technologies outpaces the DoD's ability to field mature systems with cutting-edge technology.

14. SUBJECT TERMS T&E of Autonomous Systems, system of systems, SoS	**15. NUMBER OF PAGES** 75
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	**18. SECURITY CLASSIFICATION OF THIS PAGE**	**19. SECURITY CLASSIFICATION OF ABSTRACT**	**20. LIMITATION OF ABSTRACT**
Unclassified	Unclassified	Unclassified	UU

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

ANALYZING THE CHALLENGES AND OBSTACLES TO DEVELOPING AND FIELDING AUTONOMOUS AND SEMI-AUTONOMOUS SYSTEMS

Mario D. Williams
Civilian, Department of the Army
BS, Alabama Agricultural & Mechanical University, 2008

Phillyp C. Lawson
Civilian, Department of the Army
BSME, Drexel University, 2001

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN PROGRAM MANAGEMENT

from the

NAVAL POSTGRADUATE SCHOOL
June 2020

Approved by: Robert F. Mortlock
Advisor

Raymond D. Jones
Co-Advisor

Raymond D. Jones
Academic Associate, Graduate School of Defense Management

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Autonomous and semi-autonomous systems operate on a system of systems (SoS) framework, utilizing their own ability to sense, perceive, analyze, and execute actions to achieve their goals. Unmanned systems present a significant challenge for the Department of Defense's (DoD) acquisition process that was established to develop and field man-in-the-loop type capabilities. The purpose of our thesis is to perform an analysis on the challenges of developing semi-autonomous and autonomous systems through the military acquisition process in order to determine best practices and trends necessary to increase the likelihood of program success.

The secondary goal of our analysis is to compare and contrast Test & Evaluation methodologies for systems with autonomous capabilities. The purpose of the test and evaluation process is to enable decision makers to manage technical risk and assess how robust and mature capabilities are prior to making a fielding decision. Autonomous systems require a rigorous test/build strategy that for most programs lead to cost overruns and schedule breaches. In addition, trying to keep pace with rapidly changing technologies outpaces the DoD's ability to field mature systems with cutting-edge technology.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

AAF	Adaptive Acquisition Framework
AI	Artificial Intelligence
ACAT	Acquisition Category
AMAS	Autonomous Mobility Applique System
APB	Acquisition Program Baseline
ASI	Autonomous Solutions Inc.
ATL	Automatic Take-Off and Landing
BDD	Binary Decision Diagrams
CRL	Carnegie Robotics Limited
COTS	commercial-off- the- shelf
DARPA	Defense Advanced Research Projects Agency
DBB	Defense Business Board
DJI	Dajing Innovations
DMSCO	Defense Modeling & Simulation Coordination Office
DOD	Department of Defense
DOT&E	Defense Operational Test and Evaluation
EMD	Engineering maturation development
EH	Explosive Hazards
FAA	Federal Aviation Administration
HRI	Human-Robot Interaction
IOT&E	Initial Operational Test & Evaluation
ISR	Intelligence Surveillance and Reconnaissance
IPT	Integrated Product Team
JAP	Joint Applied Project
LUT	Limited User Test
MDAP	Major Defense Acquisition Program
MS	Milestone
MTA	Middle Tier Acquisition
MRTFB	Major Range and Test Facility Base
M&S	Modeling and Simulation

NASA	National Aeronautics and Space Administration
NTC	National Training Center
OT	Operational Testing
QRC	Quick Reaction Capability
RDT&E	Research Development Test and Evaluation
SAR	Selected Acquisition Report
SREHD	Stand-off robotic explosive Hazard Detection-Neutralization
SoS	System of Systems
TARDEC	Tank Automotive Research, Development and Engineering Center
T&E	Test and Evaluation
UAS	Unmanned Aircraft System
UAV	Unmanned Aerial Vehicle
USDOT	U.S. Department of Transportation

ACKNOWLEDGMENTS

We would like to acknowledge everyone who has helped us create and evolve this thesis into the final document is has become. It seems like we started this paper with the best intentions to get it completed quickly, and on time, but this Joint Applied Project (JAP) took on a life of its own and quickly grew in scope. We would like to thank our significant others, who provided us with the time to work on this within our busy schedules, and our families, who supported us during the times we struggled to make significant progress. We would also like to thank our advisors, and the Graduate Writing Center coaches, who tirelessly provided input on making this a better product, especially in the weeks prior to our JAP final submission. This report is the summation of everything we have learned during our two years of Naval Postgraduate School (NPS) and everyone that has influenced us in our lives.

Thank you.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

Autonomous systems have gained tremendous interest and traction within the Department of Defense (DOD) over the past five years. The problem is as technology matures, the need to integrate autonomous capabilities and weapon systems becomes more of a requirement for the DOD. We will attempt to show through case studies, how the DOD is looking to develop a next generation autonomous systems capable of dominating near-peer threats. The DOD's acquisition process for autonomous systems or any other technically complex weapon system can be highly complex, and the implementation costly. As the DOD's interest for autonomous systems increases, it must take into consideration future capabilities and operational training techniques of autonomous systems. According to Parson and Ahner (2016) future capabilities of autonomous systems consist of augmenting human decisions, teaming with humans in the execution of missions, teaming with other systems for collaborative autonomous functions, and extending intelligence surveillance and reconnaissance (ISR). In the same report Parson and Ahner (2016) propose that these future capabilities will be operated in hostile environments. DOD will gain a multitude of new capabilities by implementing autonomous systems out in the field, but with the new capabilities come new challenges.

A. THE ROLE OF T&E WITHIN THE DEFENSE ACQUISITION PROCESS

According to the Parson and Ahner study (2016), the early stages of the DOD adaptation of autonomous systems have provided new challenges in conducting a proper developmental and operational test and evaluation (T&E) process. The challenges include development of requirements, the need for test infrastructure and personnel, autonomous systems test adequacy, need for safety/cybersecurity, and post-acceptance testing of autonomous systems. Figure 1 shows how these challenges line-up within the standard defense acquisition management system framework.

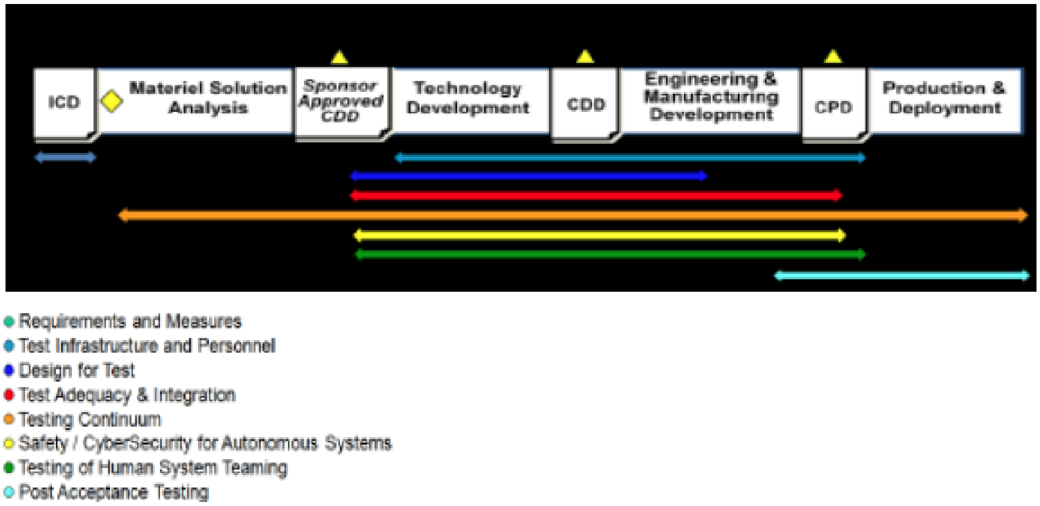

Figure 1. DOD Acquisition Life Cycle Program.
Source: Parson and Ahner (2016).

As more warfighting systems adopt autonomous systems and semi-autonomous systems, the DOD's main goal is to utilize T&E acquisition process to manage technical risks and ensure maturity of requirements are considered before final decision within the acquisition process. As technology matures the DOD and industry communities must keep pace and adapt the changes in technology to the warfighter.

B. DEFINING AUTONOMY

The *Role of Autonomy in DOD Systems* by the Defense Science Board (2012) (hereafter called *Autonomy Report)* defines an autonomous system as a system that is self-governing, with the capability for the operator to control the system without human interactions. When most people think about autonomous system, they think of software taking the place of a human interaction. Autonomous systems do not replace a human being, but rather extend an operator's effectiveness in a multitude of possibilities. The DOD indicates that the main function of autonomous systems is to reduce the cognitive burden placed on warfighter while maximizing effectiveness of any given capability. By adding autonomous systems, the *Autonomy Report* by the Defense Science Board (2012) states that the warfighter can use a DOD system or sub-system more effectively by integrating different requirements between multiple unmanned systems and software-dependent devices, thus allowing more focus on the complex decision-making process. Per the *Autonomy Report* by the Defense Science Board (2012) report the warfighter would gain added advantages by utilizing autonomous systems capabilities. Almost all the DOD deployed forces have been eager to use autonomous systems, particularly to counter the threat in compromised areas.

According to a 2009 report by Zilberstein, autonomous systems offer a reframing effect on society as they help decrease human workload and achieve productivity and safety. The same 2009 report, Zilberstein highlight that autonomous systems have been in operation in a great variation of forms, from special types of vacuum cleaners to space exploration vehicles. Ironically, there is no specific definition of autonomy in artificial intelligence (AI), but if a system can build and carry out a plan that achieves an assigned goal without human intervention it is considered autonomous.

The DOD sees the need for full autonomy but has incrementally implemented autonomy through semi-autonomous systems as a more feasible goal. The Zilberstein (2009) report defines semi-autonomous systems as a system that can operate autonomously under certain conditions but may require some human interactions in order to achieve its final goal.

For semi-autonomous systems to be successful, three technical approaches reported in the Zilberstein 2009 report must be considered. First, a human operator should have high competence to observe or interact with the current state of the process. Second, a human operator must have the ability to perform actions that are not available to the semi-autonomous systems (e.g., climb stairs). Third, a human operator must demonstrate a different level of competence in performing various tasks depending on certain actions (e.g., removing a stuck light bulb without breaking it) within the DOD T&E process.

C. THE IMPORTANCE OF AUTONOMOUS SYSTEMS ON THE BATTLEFIELD

In the future, autonomous systems will give warfighters numerous enhancements such as: augmenting human decision-making, teaming with humans in the execution of missions, teaming with other systems for collaborative autonomous functions, and extending situational awareness. These enhancements will allow the warfighter to operate a system that has a set of high-level capabilities that allow the warfighter to reply to a set of circumstances that were not pre-programmed prior to system deployment.

In addition, these enhancements will provide systems with the ability to operate in any environment without the need to preprogram software within specific boundaries, due to the artificial intelligence being able to adapt to its environment. By giving the warfighter autonomous systems capability, the number of warfighters on the battlefield will decrease, and the efficacy of each warfighter will be greater. The DOD systems that are autonomous systems or semi-autonomous systems enabled, will allow the warfighter to combat in areas not accessible in the past. As a result, we opine that the more the DOD adopts autonomous systems the less causalities will occur.

The DOD has the operational need to increase autonomous systems across air, sea and ground applications. As DOD adopts autonomy, DOD must maintain and improve a systemic approach for T&E and account for the mission needs for present and future operational environments. For industry to account for DOD mission needs, it first must understand how DOD intends to use autonomous systems.

Our plan for this JAP is to analyze test documentation, engineering reports and scientific studies to answer the research questions to follow, specifically focusing on T&E data extrapolated during the technical maturation and risk reduction phase, the engineering and maturing phase (EMD) and production and deployment phases.

The purpose of this analysis is to: 1) identify the shortcomings within the T&E methodology in comparison with industry best practices; and 2) determine corrective actions that can transition to the organic T&E test agencies in the hopes of increasing the chance of fielding autonomous systems to the warfighter. We use the case studies demonstrated in the later part of this JAP to analyze the challenges these systems face and broadly extend lessons learned from these case studies to the development of all autonomous capabilities.

In addition, we show how best practices and commercial trends can be transitioned from industry to DOD’s T&E methodology. At the end of our JAP we determine if the current DOD T&E process or industry's T&E processes provide a better evaluation of autonomous systems by answering the following primary and secondary questions:

(1) Primary:

- What unique challenges are presented by autonomous systems to the DOD T&E and industry communities?

(2) Secondary:

- Does the DOD have the correct resources to conduct tests and evaluations of autonomous systems?

In this chapter we introduced autonomous systems, explained the role of T&E within the defense acquisition process, defined autonomy, and discussed the future importance of autonomous systems on the battlefield. As the DOD begins to adopt autonomous and semi-autonomous systems, so does industry. The following chapter provides additional information on the background of autonomy based on DOD's vision, several implementations of autonomy in current day programs and a summarization of their T&E phase as compared to the industry standard. The maturity of autonomous systems and semi-autonomous systems will cause DOD and industry communities to evaluate the challenges they will have to meet, and how to best allocate resources to conduct T&E of autonomous systems.

THIS PAGE INTENTIONALLY LEFT BLANK

II. BACKGROUND

A. DOD VISION FOR AUTONOMOUS SYSTEMS IN THE BATTLEFIELD

Beginning in 2011, the DOD published a series of Unmanned Systems Integrated Roadmaps, laying out a vision for future military technology investment, which covered advancements in sensors, communication, power, weapons, propulsion and other key enablers for future autonomous systems. The purpose of these roadmaps was to provide a reference for defense acquisition leadership to pursue science and technology initiatives that align with the overall DOD strategic guidance. The 2011 roadmap presented key technological areas that represented DOD's specific interest in artificial intelligence and machine learning, which are the key enablers for future autonomous innovations and machine human collaboration (see Figure 2).

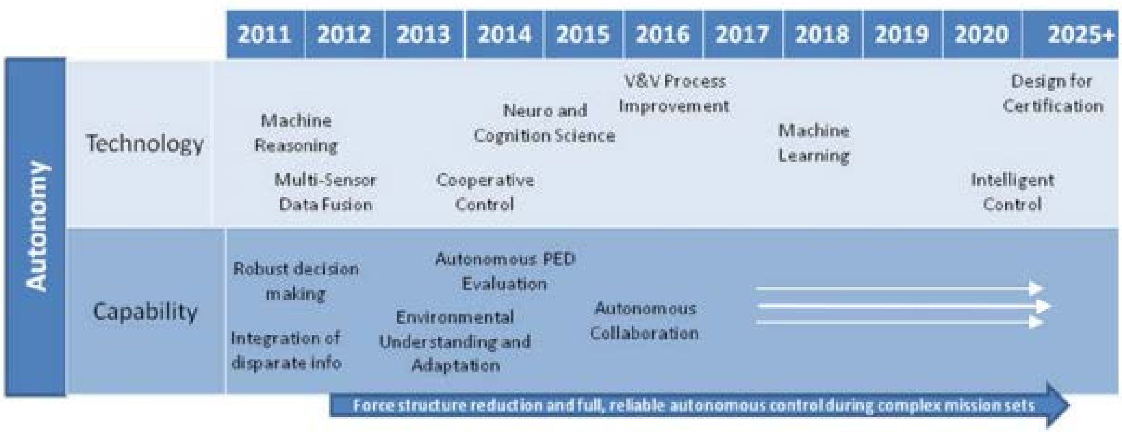

Figure 2. Defense Science Board's 2012 Autonomy Roadmap. Source: Defense Science Board (2012).

The DOD's vision behind the 2011 roadmap provides insight into the military's gamble with autonomy and the heavy reliance on artificial intelligence in future warfare (Defense Science Board, 2012).

> For unmanned systems to fully realize their potential, they must be able to achieve a highly autonomous state of behavior and be able to interact with their surroundings. This advancement will require an ability to understand

> and adapt to their environment, and an ability to collaborate with other autonomous systems. (Scharre, 2018)

This excerpt from (Scharre, 2018) provides context to how the DOD is positioning autonomous systems in the future fight and how the military is planning to achieve overmatch against near-peer threats. Based on the current dominance of drone strikes utilized in the war on terror in the Middle East, future threats will focus on the ability to contest remotely operated systems such as drones and robotic platforms on the battlefield, rendering these capabilities useless. One option the military is planning to utilize to mitigate this is emphasizing higher levels of autonomy for the purpose of developing systems that can operate independently without the need of manned teams facilitating functionality. Additionally, the military is conducting research on drone swarm technology, which is capable of controlling hundreds of “mini-drones” required to overcome enemy threats. DOD is betting on a future of AI and unmanned systems seamlessly coordinated with manned teams influencing and dominating the battle space. AI capable of identifying threats at longer range, providing intel to ground forces and coordinating activities between unmanned and manned forces. These capabilities will enhance the military’s capability in the future by reducing the warfighter’s cognitive and physical load, while providing a means to improve lethality on the battlefield.

B. DOD KEY ENABLING TECHNOLOGIES FOR AUTONOMOUS SYSTEMS

The DOD *Autonomy Report* (2012) identifies six areas in artificial intelligence that have made significant improvements for autonomy: “perception, planning, learning, human-robot interaction [HRI], natural language understanding, and multi-agent coordination” (p. 31).

The report also describes how perception is key to autonomy and enables the autonomous systems to reach mission objectives. Per the DOD *Autonomy Report* (2012) perception includes sensors (hardware) and sensing (software). The hardware (sensors) includes raw inputs to sound, pressure, temperature, and lights. The DOD *Autonomy Report* (2012) explains how “perception for mission sensing is needed for mission planning, offering four significant benefits"(pg. 42). Planning is the steps related to how a sequence is computed or partial order of actions that changes from present circumstances to a desired circumstances. In the same report, the planning process consists of two components. The first component represents human actions, and description of world objectives and resource optimization criteria. The second component represents how the DOD learns to design software algorithms that will compute actions in sequences and assign the appropriate actions to conform the hard constraints of the problem, while optimizing the soft constraints. According to the DOD *Autonomy Report* (2012) machine learning is considered the most critical aspect of advancing intelligent autonomous systems.

The DOD *Autonomy Report* (2012*)* states HRI is new to autonomous systems. The same report defines HRI as an integrative field that interact direct human interaction compared to current usage of robots versus computers or tools. The DOD *Autonomy Report* (2012) further states that,

> HRI is a group of a bigger field that focuses on bi-directional, cognitive interactions in which the robot is a physically situated agent operating at a distance from the user, versus a computer or autopilot, thus leading to significant distinctions. (Defense Science Board, 2012)

In the same report the DOD benefits from HRI by improving performance of existing platforms, which will reduce cost for new situations.

Natural language is defined in the DOD *Autonomy Report* as the most inherent way for the user to control autonomous systems. Natural language understanding grants for a more diverse, superior operation rather than and more detailed teleoperation. As revealed in the DOD *Autonomy Report*, multi-agent coordination is applied to accomplishing a designed task for software or humans. The DOD *Autonomy Report* states that,

> Each multi-agent component is considered to have some degree of individual autonomy and the coordination may either emerge from the agents interacting or negotiating with each other directly (distributed coordination) or be explicitly directed by a planner (centralized coordination). (Defense Science Board, 2012)

In the next section, we present a brief overview of the four DOD programs of record (PORs) that we chose to analyze in this JAP. Each of the programs demonstrates one or multiple areas of autonomy designated by the Defense Science Board and were fully funded efforts that required a formal T&E process by the Army Evaluation Center. Each of the four PORs have a semi-autonomous systems capability that required an integrated project team (IPT) to develop non-traditional T&E methodology to verify and validate the autonomous functionality. In the next section we provide a brief synopsis of each program's development history and the implementation of their autonomous functionality.

C. DOD PROGRAMS OF RECORD WITH AUTONOMOUS FUNCTIONALITY

1. Gray Eagle

According to the *Selected Acquisition Report* (SAR) (2010) Gray Eagle or MQ-1C Unmanned Aircraft System (UAS) program is the Army's primary aerial system capable of providing reconnaissance, surveillance, security, attack and command/control missions' types for the joint force units. Gray Eagle Quick Reaction Capability (QRC) 2 Units are capable of autonomous takeoff and landing, employing hellfire missiles and tactical landings (SAR 2010). All other functionality is coordinated by an operator within a ground control station.

The SAR (2010) states that the Gray Eagle was initiated in April 2005 at milestone (MS) B and began as an acquisition category (ACAT) II program planned to replace the Hunter UAS, in response to secretary of defense requested to increase support in Iraq and Afghanistan. The SAR (2010) reports that the initial program was approved for procurement of four systems and associated equipment, but was increased to 12 systems post MS B.

The SAR (2010) states by the spring of 2008, the secretary of defense managed the deployment of prototype Gray Eagle UAS equipment to support the war in Iraq. QRC Gray Eagle systems were procured and deployed to support Operation Iraqi Freedom and Operation Enduring Freedom. Based on operational mission duration taken from both accounts, the deployed systems were capable in excess of 10,000 flight hours with zero accidents. In May of 2008, Gray Eagle was re-designated as an ACAT 1D and by March of 2010 approved for a MS C. The MS C was later rescinded and changed to March 2011 based upon recommendation from the vice chief of staff of the Army convening a configuration steering board. Adjustments increased the allowable Gray Eagles within a platoon to 3 per company, which was a total increase from 12 to 17 Gray Eagle UAS companies. According to the SAR report PM-UAS performed a successful limited user test (LUT) in the third quarter FY 2010 and initial operational test and evaluation (IOT&E) was completed in the first Quarter FY 2012. Figure 3 is an image of a MQ-1C UAS Gray Eagle.

Figure 3. Selected Acquisition Report: MQ-1C UAS Gray Eagle. Source: SAR (2010).

2. Autonomous Mobility Applique System (AMAS)

According to the AMAS system overview, the AMAS is an Army initiative providing a low-cost/low-risk solution to retrofit an autonomous driving capability onto existing military tactical wheeled vehicles. AMAS offers driver warning/driver assist performance, leader-follower vehicles for protection operations, and waypoint following efficiency. The AMAS kit is composed of a three-part system of environment sensor, actuators that move the vehicle and pump the breaks and central computer that processes the sensor data and commands the drive system.

Based on the AMAS system overview, in the summer of 2014, the engineers at Tank Automotive Research, Development and Engineering Center (TARDEC) demonstrated the AMAS system at Savannah River site, in Aiken, South Carolina. The demo consisted of several types of vehicles being autonomously driven in operationally realistic scenarios. The assessment resulted in a positive recommendation to continue development. Additional progress was made in 2015 again when the driver warning/assist capability was successfully tested with soldiers. The demonstration was set up to test the collision mitigation braking, adaptive cruise control and lane keeping capability of vehicles equipped with the AMAS technology. The primary and secondary safety systems of the AMAS were successfully demonstrated, establishing AMAS’s capability to reduce equipment loss and mitigate potential accidents.

In addition, the AMAS system was successfully demoed during the U.S. Army Extended Warfighter Experiment at Fort Leonard Wood, Missouri, and Fort Bliss, Texas. The EWE is a U.S. army training and doctrine command sponsored event, managed by the U.S. Army TARDEC. The event was significant for both Lockheed Martin (the initial system developer) and TARDEC as they successfully logged 55,000 miles on the system and received a recommendation to move to widespread fielding across multiple military applications. The Figure 4 is an image of an AMAS system.

Figure 4. Autonomous Mobility Applique System Overview. Source: Lockheed Martin (2018).

3. Project Perdix

According to the Strategic Capabilities Office (SCO) (2011), Project Perdix is the name of a MIT Lincoln Laboratory initiative to use a group of micro-drones (mini unmanned aerial vehicles) for ISR missions. Project Perdix was successfully demonstrated by the SCO, in September 2014, for the DOD at the Air Force Test Pilot School, Edwards Air Force Base. The test utilized an F-16 flare canister as a transportation/ejection method and was successfully capable of launching a swarm of autonomous drones without damaging any of the drones. Over the next two years the program would transition to the SCO, Naval Air Systems command, where the military alongside MIT, demonstrated a swarm of 103 Perdix drones being deployed from an F/A-18 Super Hornet. The SCO revealed the demonstration is considered a milestone in drone swarm technology and demonstrated advanced behavior "such as collective decision-making, adaptive formation, flying, and self-healing"(pg.1). The Perdix drones can share a single collective mind for decision making and adapting to each other and the different environmental factors. The

Perdix drones will provide the military the capability to allow a swarm of inexpensive drones to perform missions, disrupt communications and track targets of interest. The Figure 5 is an image of a single Perdix drone.

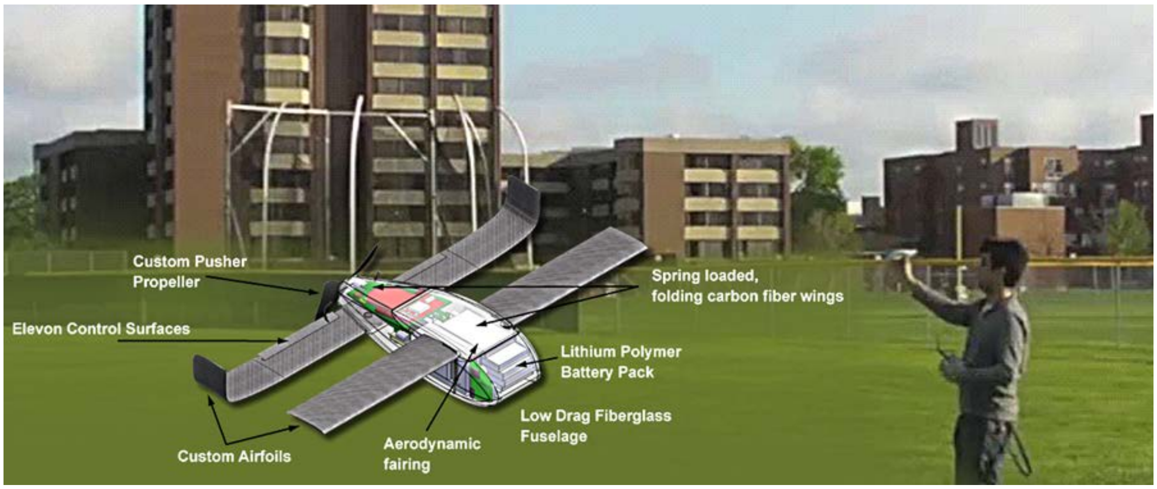

Figure 5. Project Perdix. Source: Hansman (2011).

4. Stand-off Robotic Explosive Hazard Detection-Neutralization (SREHD)

SREHD is an Acquisition Category III (ACAT III) program developed as a capability to detect, mark, and neutralize explosive hazards (EH) (e.g., landmines, improvised explosive devices, booby-traps, and unexploded explosive ordnance) in either complicated or underground environments. SREHD is capable of being mounted on unmanned ground vehicles platforms organic engineering unit.

In 2002 the Army Technology Objective initiated an effort to develop a concept to autonomously detect explosive hazards on the battlefield. Concept Phase 1 was conducted as a market survey for technology approaches to autonomous mine detection of anti-personnel and anti-tank landmines from a small surrogate robotic platform. Concept Phase 2 was a prototype design evaluation to determine the probability to detect and coverage rate autonomous capabilities against realistic threats in varying environmental conditions. SREHD formerly known as the Autonomous Mine Detection System was developed by two competitive, state-of-the-art sensor packages developed for handheld, man portable detection systems. CyTerra Corporation leveraged technology from the Army/Navy Portable Special Search 14 Mine Detecting Set and NIITEK Inc. was able to leverage ground penetration radar to meet the required objective. Both contractors were awarded contracts to explore technology approaches capable of autonomously detecting landmine. Concept Phase 3 was developed and assessed prototypes against mission scenarios in relevant environments. At the culmination of Phase 3, the Program Management office for Close Combat Systems awarded a contract to Carnegie Robotics Limited (CRL) for their Dismounted Standoff Explosive Hazard Detection, Marking and Neutralization approach. On 30 April 2018, PEO Ammo as the SREHD Milestone Decision Authority successfully approved a MS C decision and CRL was award an additional option to produce 16 units in support of low rate initial production (LRIP) (Test and Evaluation Master Plan, SREHD, 2018). The Figure 6 is an image of a SREHD system.

Figure 6. SREHD Shreds the handheld competition. Source: James and Chang (2017).

THIS PAGE INTENTIONALLY LEFT BLANK

III. ANALYSIS OF TEST & EVALUATION METHODOLOGIES

A. THE IMPORTANCE T&E FOR AUTONOMOUS SYSTEMS

The DOD has not developed any specific guidance for T&E of autonomous systems. Currently, autonomous systems are tested and evaluated under the same methodology as any other complex weapon system. The process starts by decomposing requirements into their basic operational intentions. Test plans are created and approved by system stakeholders, test leads evaluate system performance by executing a series of repeatable system tasks for the purpose of evaluating and verifying predictable system operational requirements.

The DOD *Autonomy Report* (2012) describes the current state of DOD's test and evaluation challenges and recommendations for T&E of autonomous systems by stating,

> Traditional test programs have been focused on repeatable performance testing, measuring a response and comparing that response to a documented performance specification. Testing perspectives need to shift to a view that is more broadly mission based and assesses the ability of the autonomous system to meet mission goals. The framework employed for testing must provide leeway to the system to adapt plans to achieve mission goals in a variety of ways that cannot necessarily be predetermined. The fact that the system's software reacts to external stimuli and makes non-scripted, but bounded, decision is particularly challenging to the test community that is more accustomed to executing testing in a fully scripted sense. The ability of the T&E community to react to this changing paradigm is limited by the understanding of how autonomous systems truly make decisions. Additionally, the DOD T&E workforce must be enhanced with new skills for robotics, artificial intelligence, networking and systems engineering for autonomous systems. (Defense Science Board, 2012)

This excerpt clearly defines the critical need for a T&E community tailored to the nuances of measuring the effectiveness and suitability of autonomous systems.

B. THE COST OF TEST & EVALUATION FOR AUTONOMOUS SYSTEMS

Autonomous systems operate on a system of systems (SoS) basis. Fundamentally, this means that a collection of resources and capabilities work together to create a more complex and enhanced capability, which offers more functionality and performance than the sum of its individual components. The increased amount of functionality can lead to emergent behavior that can be a challenge to the T&E community to replicate, verify and validate. The government's current verification and validation methodology ensures user requirements are evaluated based upon a predictable and expected performance envelope. The T&E community then collects, assesses and judges whether the performance measures fulfill requirements. This data collection period begins during the technical maturation and risk reduction phase and typically ends during the end of the EMD phase. The purpose of T&E within the Defense Acquisition process is to help decision makers assess the maturity of a capability and reduce the risk of failure to the system owner and users. The systems approach to test and evaluation is the principal means of demonstrating a program's readiness for deployment.

SoS T&E requires broad-scale resources based upon the complexity of the systems, missions and conditions the warfighter expects to utilize the capability. System engineers work with the T&E community to break down requirements, systems and missions into smaller pieces, which are translated into quantifiable measures. Large, exhaustive and costly experiments are conducted to inform development integrated product teams (IPT) and other system stakeholders of the overall maturity of a system and its relative timeframe to fielding.

As an example, Gray Eagle is a major defense acquisition program (MDAP) that was required to undergo IOT&E to validate its autonomous functionality. Gray Eagle's autonomous functionality includes a system capable of automatic take-off and landing (ATL) as well as a system capable of employing Hellfire missiles on a verified threat.

The information that was gathered in Table 1 and Table2, includes Gray Eagle's acquisition program baseline (APB) and the project management office (PMO) budget to fund autonomous requirements in IOT&E from September 2011 and August 2012. Using information gathered in Table 1 and Table 2, which includes Gray Eagle's acquisition program baseline (APB) and the project management office (PMO) budget to fund IOT&E from September 2011 to August 2012. We can use the established timeline in Table 1 with Gray Eagle's annual funding in Table 2 and surmise how much money was spent to validate the autonomous functionality.

Table 1. MQ-1C Gray Eagle's Schedule. Adapted from SAR (2010) and SAR (2007).

Schedule Events				
Events	**SAR Baseline Production Estimate**	**Current APB Production Objective/Threshold**		**Current Estimate**
Milestone B	Apr 2005	Apr 2005	Apr 2005	Apr 2005
SDD (EMD) Contract Award	Apr 2005	Apr 2005	Apr 2005	Apr 2005
Critical Design Review	Feb 2006	Feb 2006	Feb 2006	Feb 2006
Milestone C	Mar 2011	Mar 2011	Mar 2011	Mar 2011
IOT&E				
IOT&E Start	Sep 2011	Jul 2012	Jul 2012	Jul 2012
IOT&E Complete	Oct 2011	Aug 2012	Aug 2012	Aug 2012
IOC	Jun 2012	Dec 2012	Dec 2012	Dec 2012
FRP Decision	Apr 2012	Jul 2013	Jul 2013	Jun 2013
FOT&E I	Aug 2012	May 2015	Nov 2015	Jun 2015
FOT&E II	May 2013	N/A	N/A	N/A

Prior to the ATL being implemented on production units, General Atomics was required to verify Gray Eagle's ability to autonomously take-off and land to ensure the safety of military test assets and a successful IOT&E. Additionally, information was gathered through a press release by General Atomics, establishing the system demonstrated 20,000 successful take-off and landings on various runways around the world (General Atomics, 2012). The (PMO's) schedule of events published in the MQ-1C Gray Eagle 2016 SAR (Table 1), confirmed Gray Eagle's IOTE was scheduled to begin in September of 2011 and ended in August of 2012. Execution of all these events required prior year planning to include the acquisition of test hardware, verification of system performance, maintenance/logistic documentation and trained operators for system demonstration before the beginning of testing in September of 2011. This schedule of events coincided with a peak operating test budget between the years of 2010 to 2012 which reflects the highest development cost during Gray Eagle's engineering maturation phase as shown in Table 2 SAR (2007).

Table 2. MQ-1C Annual Funding Report. Adapted from SAR (2010) and SAR (2007).

Annual Funding 2040 | RDT&E | Research, Development, Test, and Evaluation, Army							
		BY 2010 $M					
Fiscal Year	Quantity	End Item Recurring Flyaway	Non End Item Recurring Flyaway	Non Recurring Flyaway	Total Flyaway	Total Support	Total Program
2005	--	--	—	—	--	--	58.8
2006	--	--	—	—	--	--	95.5
2007	--	--	—	—	--	--	127.3
2008	--	--	—	—	--	--	104.4
2009	--	--	—	—	--	--	61.6
2010	--	--	—	—	--	--	132.7
2011	--	--	—	—	--	--	114.8
2012	--	--	—	—	--	--	115.6
2013	--	--	—	—	--	--	64.1
2014	--	--	—	—	--	--	12.0
2015	--	--	—	—	--	--	41.8
2016	--	--	—	—	--	--	—
2017	--	--	—	—	--	--	11.8
Subtotal	2	--	—	—	--	--	940.4

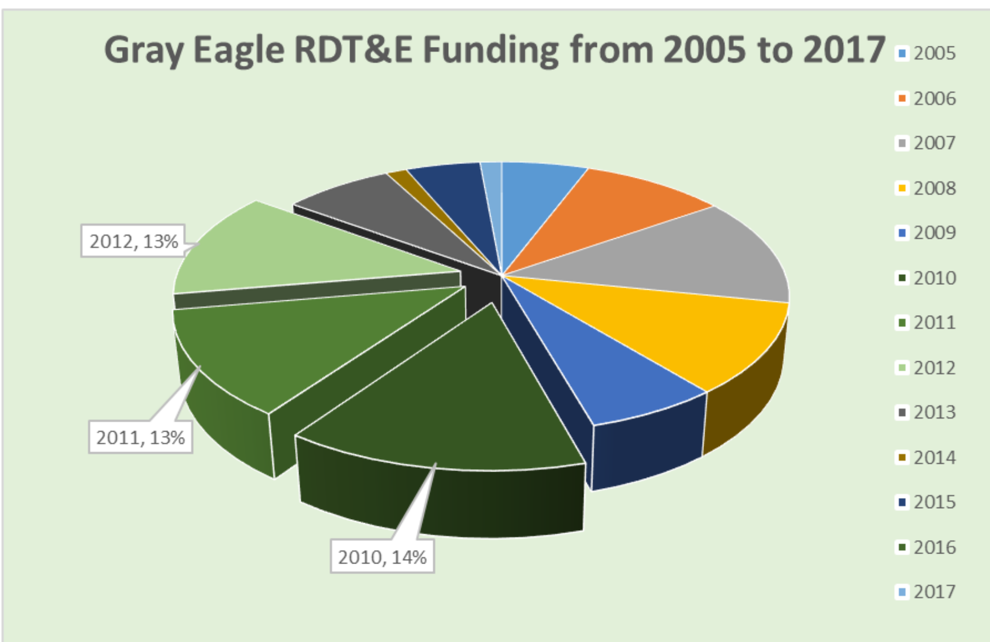

Figure 7. Gray Eagle RDT&E Funding. Adapted from SAR (2010).

Based on the data provided, the Gray Eagle program spent the largest amount of RDT&E funds during the same period of the time in which they were verifying the system's automatic takeoff and landing capability. A one-to-one correlation to Gray Eagle's budget and IOT&E cannot be discerned from just the information researched for this JAP. There are most likely several other major factors that affect the overall increase in cost. We are assuming that since the majority of testing between 2010 to 2012 was contractor testing of the ATL semi-autonomous system, then the overall cost to verify Gray Eagle's autonomous capability in an operational environment was the cost driver during the end of the EMD phase, as referenced by both the SAR and General Atomics System AUSA announcement, (General Atomics, 2012).

C. DOD TEST & EVALUATION METHODOLOGIES

Throughout this next section, we present the research was conducted on our four (4) DOD programs of record to analyze the T&E techniques utilized throughout their system development.

1. SREHD T&E Test Methodology

SREHD's autonomous capability enables the system to independently detect and mark a broad spectrum of surface-laid and buried explosive hazards and triggering mechanisms in support of route clearance and breaching operations. SREHD is unable to neutralize explosive hazards autonomously, rather the operator is always required to provide the primary and secondary confirmations prior to a neutralization. This is to prevent accidental functioning of the munition during a route clearance mission.

The testing methodology utilized by the SREHD IPT involved various test sites that specialized in specific areas of testing in order to assure the proper standards were followed providing the evaluators the necessary information to assess autonomous performance. Both Yuma Proving Ground in Arizona and AP Hill in Virginia were used to simulate likely threat environments in which a clearance brigade would operate. During the technology demonstration phase, the contractor utilized experimental prototype systems in order to assess SREHD's autonomous surface detection capability and ability to independently maneuver a detection system at precise head height to maintain accuracy of probability of detection model.

Modeling and simulation (M&S) of SREHD's detection algorithm software was also utilized throughout initial development to assess how different external environmental factors could potentially impact the system's performance. This allowed the contractor to quickly evaluate changes and train the system without requiring weeks of probability detection testing.

During the EMD phase, a combination of component, subsystem, and system level testing was utilized to stress SREHD's performance over several hundred hours of testing. Operators were required to simulate a route clearance mission with the entire SREHD operational load and clear a 1.5-meter-wide lane for troops to utilize. Both surface-laid and deep-buried targets were hidden within the "lane" and SREHD was required to autonomously travel the length of the path marking and detecting possible UXOs. Test periods ended with data collectors scoring system aborts and essential function failures and required the engineering team to work with the contractor to develop a corrective action plan that would mitigate the failure and reduce future risk of failure.

2. Gray Eagle T&E Test Methodology

According to the (DOT&E,2010), test team generated thousands of hours of operational flight testing, autonomous takeoff and landing in addition to threat acquisition and threat neutralization test flights (DOT&E, 2010). Testing was conducted at Edwards Air Force Base, California, and the National Training Center (NTC), Fort Irwin, California. To date, Gray Eagle utilized computer-based simulation training for Gray Eagle operators and actual test flight in both the continental United States and outside continental United States environments.

The computer-simulated test environment is primarily used to train Gray Eagle operators with monitoring flight systems, air traffic and communication systems with other forces. In the rare instance a target requires neutralization, operators are required to authorize the use of force prior to threat engagement.

During Gray Eagle operational testing, autonomous systems are verified and validated using continuous software checks, flight instrumentation data collection and continuous hour's flight reliability data. Since Gray Eagle is an ACAT 1D program, the low rate initial production (LRIP) quantity was requested and authorized for employment in OCONOS missions for reconnaissance. The system was able to be tested in a constrained and conflicted environment. Operators were able to fully test Gray Eagle's sensors and autonomous capabilities against foreign threats.

3. Project Perdix T&E Methodology

Project Perdix is still pre-MS A as of completion of this JAP, so there hasn't been a formal T&E master plan drafted or approved. Based upon the limited data collected on Project Perdix development, the test methodology has been to evaluate the swarm intelligence of Project Perdix, the swarm's ability to autonomously avoid obstacles and its ability to be launched from an F-16 in flight. Since the primary autonomous capability of the Perdix system is its ability to autonomously navigate low altitudes and independently control individual drones within the swarm, government experiment and testing has primarily focused on increasing the size of the swarm and increasing the complexity of the test environments.

On October 2016, the Navy successfully deployed 103 Perdix drones from three F/A-18 Super Hornets over China Lake, California. Perdix demonstrated quick maneuverability, adaptive flying formations and self-healing.

Typical testing for military-employed small UAV's consists of autonomous target tracking, flight duration and payload delivery. DOD uses White Sands Missile Range, New Mexico and Fort Bliss, Texas as its primary test facilities for new UAV's in development.

4. AMAS T&E Test Methodology

The Army's AMAS or "convoy system" is primarily tested either at Fort Leonard Wood, Missouri, Aberdeen, Maryland and Fort Bliss, Texas. The applique kit has been tested on the M915 line-haul truck, medium tactical vehicle replacement, palletized load system and the heavy equipment transport vehicles. Lockheed Martin has successfully logged more than 55,000 miles using the AMAS capability.

The test methodology for AMAS is primarily roadside simulations of its autonomous adaptive cruise control, assisted breaking and parking, collision avoidance and pedestrian detection. The automotive technology evaluation facility at Aberdeen Proving Ground (APG) can use plastic vehicle targets, Styrofoam mannequin pedestrians and other robotic devices to represent various scenarios to gather system characteristic data and replicate possible safety hazards. Testing is done with little to no human intervention to accurately simulate realistic operational scenarios of the AMAS system and mitigate any possible safety risk to personnel. During test events, data collectors are monitoring the AMAS capability to navigate dense urban environment in a convoy formation, braking and avoiding obstacles and its ability to warn other vehicles in the convoy of sudden changes in terrain. Data collectors measure the accuracy of the AMAS capability to correctly navigate the course based on a predetermined path pre-programed into the system versus how well it navigates the course based on real-life obstacles moving in and out of lanes. The data is then used to create models that eventually will be used to simulate the AMAS capability across different types of courses and reduce future test cost.

The goal of the AMAS test program was to transition into other programs of record for route clearance and interrogation systems. The future of the AMAS system will be fully autonomous vehicles as a counter to roadside hazards explosive operations mitigating the need of a soldier directly in harm’s way.

5. DOD T&E Facilities for Autonomous Systems

Currently the military does not have dedicated test facilities positioned for unmanned autonomous systems. The major range and test facility bases (MRTFBs) are the designated core set of T&E infrastructure and associated workforce for the DOD and are preserved as national assets. The common methodology across DOD for testing unmanned systems is to utilize existing facilities and infrastructure used for manned systems.

Typically unmanned aerial systems are tested in conjunction with the existing Army facilities. The Army's Gray Eagle utilizes Edwards Air Force Base in California and the National Training Center (NTC), Fort Irwin, California.

The SREHD payload is currently tested on route clearance lanes at Yuma Proving Ground in Arizona or in the rougher terrain of Fort AP Hill, Virginia. Test facilities are reconfigured from either existing foot trails or vehicle lanes and repurposed for SREHD payload test events.

The AMAS capability went through verification testing at Fort Leonard Wood, Aberdeen Proving Ground and Fort Bliss during its development. These test sites were originally intended to be vehicle test ranges that again were repurposed as test facilities for an autonomous capability.

6. DOD T&E Software, Simulation and Modeling Strategy for Autonomous Systems

All unmanned systems utilize advance software algorithms as the backbone their functionality and architecture. The DOD has a long running history of software-intensive programs that result in long delays in fielding, significant and cost overruns, as seen in high profile programs like the F-35 joint strike fighter (Maurer, 2019). The system engineering process within the defense acquisition environment has a solid foundation in mechanical-based designs versus software development, even though both contribute an equal measure to system functionality. This has resulted in challenges in both software development and software T&E within acquisition programs. Typical government procurement of software-intensive systems requires heavy industry investment and government oversight and requirement decomposition. Commercial off-the-shelf (COTS) software solutions remodified for government weapon system purposes are the predominant acquisition scenario for DOD software-intensive systems. COTS software solutions reduce the overall government investment cost, and quickly provide a solution mature enough to field.

DOD methodology for T&E of software-intensive systems during development is a controlled, statistically significant and iterative process. Tests are usually designed to verify performance and ensure quality attributes have been satisfied. During development of software-intensive systems, software is the significant driver of system maturity, to include system performance measures. System integration teams are established to iteratively validate and verify software builds prior to release.

The DOD T&E acquisition community defines software maintenance as the, “process of modifying a software system after delivery to correct faults, improve performance or adapt it to a changed environment” (Hamilton, 2017). Testing of software-intensive systems during the sustainment phase of life cycle acquisition involves verifying software systems continue to perform in accordance with requirements.

In addition to software testing, DOD also uses M&S as an additional means to test, analyze or train real world capabilities in controllable environments. The M&S Coordination Office is the DOD’s central repository for providing policy to Army, Navy and Air Force Agencies (Defense Modeling & Simulation Coordination Office [DMSCO], 2019). DOD supports using M&S across the entire life cycle of its programs, but it’s often used to verify metrics during the EMD life cycle phase. The data gained from simulation is used to provide confidence in design decisions prior to integration and recommendation on a critical path forward. It is DOD policy that only programs that have had their models verified and validated throughout their life cycle by an accredited organization can be used as substitutes for actual test data (Department of the Army, 2014).

D. TEST & EVALUATION METHODOLOGY FOR COMMERCIAL AUTONOMOUS SYSTEMS

Currently there is no overarching policy or industry standard for the commercial sale of autonomous systems. Unlike either the Food and Drug Administration which was stood up to protect and control food safety, tobacco products and over-the counter pharmaceutical drugs or the Federal Communications Commission, which is an agency within the government that regulates interstate communications by radio, television satellite or phone, there currently no federal agency to create, promote or oversee the safety standards for autonomous systems.

Commercial autonomous systems have been in the public sector since the 2004 defense advanced research projects agency (DARPA) challenge, where autonomous vehicles were tested in a series of challenges by the DARPA to self-navigate a 150-mile course in the Mojave Desert region of the United States. This event was key in pushing autonomous vehicle technology forward because the major car manufactures started investing more and more funding into research and design solutions for autonomous capabilities.

This thesis analyzes the autonomous test capability of various industries with commercial autonomous systems in order to compare similar capabilities to the four identified DOD programs. The autonomous vehicle industry, autonomous agriculture systems, and autonomous aerial systems each have their own set of separate standards and testing capabilities that sets them apart from one another.

1. Agriculture System Capability T&E Methodology

The future of agriculture is heavily dependent on producing healthier food, quicker and more efficiently than existing farming techniques of today. Farmers are more reliant on highly skilled personnel to work fields and drive machinery necessary to keep crops thriving, which can be a challenge during peak seasons. Tomorrow's farmers will be driven to utilize more autonomous machinery that will be just as safe if not safer than humans. Autonomous agriculture systems enable farmers to extend their workday to a 24-hour, non-stop operation (Edan et al, 2009, pp.1095-1128).

According to the same report Autonomous Solutions Inc. (ASI) and Case and New Holland (CNH) Industrial are the current market leaders in providing automation technology for the agriculture industry. ASI and CNH provide autonomous tractors that can autonomously seed, plant and till for broad acre and row crop farms. Per the automation agriculture report (Edan et al., 2009), state operators can collect real time data on both the health of crops and maintenance of the machinery. ASI provides the hardware and software solutions that are capable of being retrofitted into existing agricultural systems or integrated into new machines. CNH industries designs, produces and sells agriculture machinery all over the world.

ASI's test methodology, verification and validation of their autonomous agriculture equipment is done on a private proving ground built within their engineering facility (ASI

2016). ASI can rapidly prototype, develop and test experimental autonomous equipment on a test farm environment. The test farm is pre-equipped with sensors, cameras and digital measuring equipment that enables ASI to remotely monitor their new equipment and adjust systems in real time to overcome issues and failures. Systems can be run 24/7 within a simulated environment to test maintainability and sustainability of systems prior to commercial release. After equipment is commercially released, new features can be added based on customers' specific needs and problems can be replicated to reduce customer down time. ASI's proving ground also enables them to monitor the stress of high-risk components over time and make design decisions at a higher confidence rate (Scott Garvey 2018).

During early development of CNH's autonomous tractor, ASI oversaw the intensive test program for both the tractor and planter interfaces as the system transitioned unto "real world" testing across four locations in the U.S. and Canada. CNH has partnered with Bolthouse Farms, a major carrot grower and food wholesaler to utilize the autonomous tractor alongside existing equipment. Testing was conducted to evaluate how well autonomous tractors till versus existing manned tractors. Testing has taken place across a variety of tillage applications, soil types, weather conditions and sensing and perception activities. The goal of the "real world" test was to verify that the autonomous tractor can efficiently function in any environment and provide traditional farmers an experience farming in the future with autonomous systems. The Figure 8 is an robotic tractor of the future.

Figure 8. Forage 2017. Robotic Tractor of the Future. Source: Nielsen (2016).

(1) Drone Manufacturer T&E Methodology

Commercial drone manufacturer’s T&E methodology of autonomous aerial systems is centered on assuring effective and safe operational performance in response to sudden changes in external environments (Intertek, 2016). Autonomous aerial systems are required to safely fly in fair and extreme weather conditions around the world and make independent decisions regarding the safest course of action.

Based on the severity and expense associated with testing autonomous aerial systems, private industry spends most of its resources utilizing simulation and modeling to verify and validate software systems responsible for independent behavior early in the development phase of new products. Even before metal is cut in the manufacturing of a UAV, software engineers have already verified and validated software models that simulate and correct system behavior. Industry has the capability to heavily rely on reusing much of their existing software code from prior products as software assurance for future products.

This method of software development minimizes the time needed to validate new features while reducing the total amount of resources needed for test and evaluation newer products. This incremental approach to T&E of autonomous software systems has turned into a reliable way of introducing new products to the commercial sector faster with minimal amount of testing required and spending. New products are scaled to work with limited features that have been thoroughly tested and as newer models are released, additional features and increased levels of autonomous behavior are added to all of the existing product lines, adding new features and capabilities. Companies such as Dajing Innovations (DJI) and 3D Robotics have used a single software platform to iterate on various products to the market within a short time frame.

Although the civilian drone market has become a booming sector for new startup businesses, BI Intelligence, a media outlet that specializes in data driven research, predicts that consumer spending will be a significantly smaller portion of the global aerial drone market in the near future. UAV systems used for military purposes will quickly dominate the market and require smaller, niche drone manufacturers to expand their portfolio. The difference in the civilian drone market as compared to the Defense industry is an order of magnitude greater. Autonomous drones manufactured for defense usually have much more complicated functionality and require substantially more development time. The Figure 9 is the amount of money being spent on development of UAVs for military defense purposes will continue to rise in relation to the amount of money being spent on commercial UAVs Business Insider Intelligence [BII] (2016).

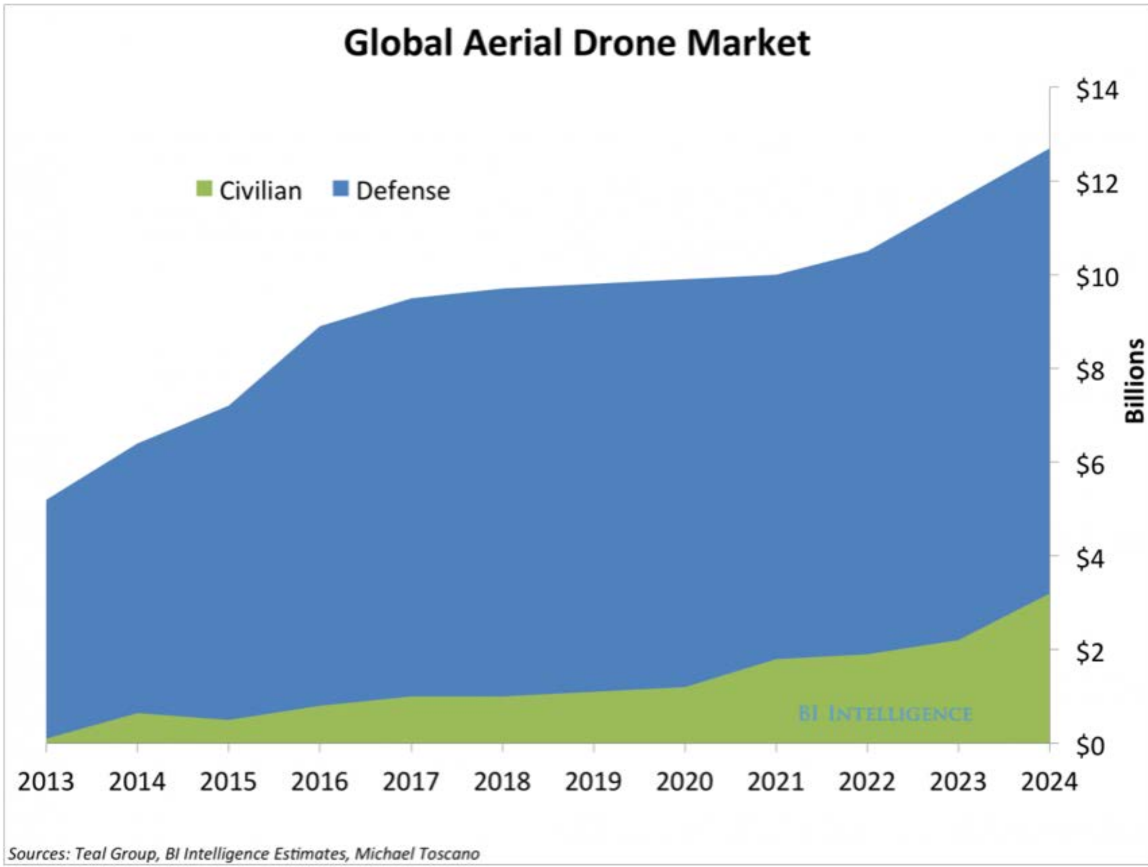

Figure 9. The Drone Report: Market, Forecasts, Regulatory, Barriers, Top Vendors, And Leading Commercial Applications.. Source: Business Insider Intelligence [BII] (2016).

(2) Industry's Autonomous Driving T&E Methodology

Comparatively, the autonomous vehicle industry has the most mature T&E methodology of all the current autonomous systems in use. Autonomous vehicles have been in use since the 2004 DARPA challenge. There are various methodologies currently in use and facilities such as the University of Michigan's Mcity facility that are dedicated to validating vehicles' autonomous software systems. The accelerated rate at which autonomous vehicles are driving on our roads is specifically due to the collaboration of the auto-industry, government, academia and technology companies who have come together in a short period of time to resolve consumer needs.

The most popular T&E method for verifying autonomous systems on vehicles is a form of accelerated testing developed by engineers (Hue Peng, director of Mcity and the

Roger L. McCarthy Professor of Mechanical Engineering at U-M*)* at the Mcity facility requires autonomous vehicles to be rigorously tested in repeatable components based on extremely difficult real-world driving situations. According to Zhao and Peng (2017) the test method can reduce 300,000 miles of real world driving to only 1,000 miles of test data required to yield equivalent information. In the same report the goal of this test method is to analyze each vehicles' ability to make difficult decision that might result in a fatal accident. Per Zhao and Peng(2017), researchers need to be able to identify the most likely scenarios in which autonomous vehicles will make an incorrect decision and thoroughly vet software until errors can be evaluated and resolved. In the same article, Zhao and Peng (2017) noted that researchers are hoping to provide consumers up to 80 percent confidence that they are 90 percent safer in autonomous vehicles than self-driving ones.

Alongside a focus to reduce the amount of testing for autonomous vehicles, T&E engineers have also developed new methods of validating autonomous vehicle performance. Researcher Ding Zhao has stated that "Test methods for traditionally driven cars are something like having a doctor take a patient's blood pressure or heart rate, while testing for automated vehicles is more like giving someone an IQ test" (Nicole Moore, 2017). Engineers can not only determine the effects of what happens to the passenger and vehicle during an accident but also determine the best way to prevent an accident of occurring.

Verifying and validating how the autonomous systems is triggered and if the software enables the intended preventative action reliably are all part of the evaluation process during development. Autonomous vehicle proving grounds can capture real world data and creating a simulation where variables can be tweaked and changed, reducing the overall amount or real-world test data required to make design decision early on.

More recently, Toyota has adopted a similar test methodology, creating a 60-acre proving ground at Michigan Technical Resource Park in Ottawa Lake to test "Edge Cases" or extreme driving scenarios. Toyota plans to have a facility capable of simulating congested urban environments, slick surfaces and four-lane divided highways with high-speed exit and entrance ramps.

(3) Commercial Industry Autonomous T&E Facilities

In 2013, UAV began being regulated by the U.S. Department of Transportation (USDOT) and the Federal Aviation Administration (FAA). Both agencies have partnered with private industry to designate six (6) locations across the United States designated for the T&E of UAS and providing a space where industry and private stakeholder community can conduct advance research and share information. The data being collected will be used to support future FAA policy on critical UAV safety regulations. (Federal Aviation Administration, 2013) The Figure 10 is an image of the six locations across the U.S. designated for T&E of UAS.

Figure 10. Federal Aviation Administration Report. Source: FAA (2013).

The auto industry has several T&E facilities across the U.S. designated for autonomous vehicles, the primary T&E facility is the Mcity Test Facility, located within the University of Michigan’s Norther Campus. Mcity is a 16-acre controlled environment that is capable of simulating both urban and suburban environments prior to deploying experimental vehicles on public roads. Mcity was a joint venture between the Michigan Department of Transportation and the U. M. to create a space where new technologies can be safely tested. Since the development of Mcity, the U.S. Department of Transportation has designated an additional 10 proving grounds dedicated for the test and evaluation of autonomous vehicles. Locations range from the Northern Pittsburg Thomas D Larson transportation center to the southern Texas AV proving grounds and all the way to the Iowa City Area Development Group. The Department of transportation has identified these proving grounds as central hubs for the auto industry and academia to share best practices in the hopes of accelerating autonomous technologies and deploying safer cars on our roads (Mcity Test Facility, 2019). The Figure 11 is an image of Mcity test facility map.

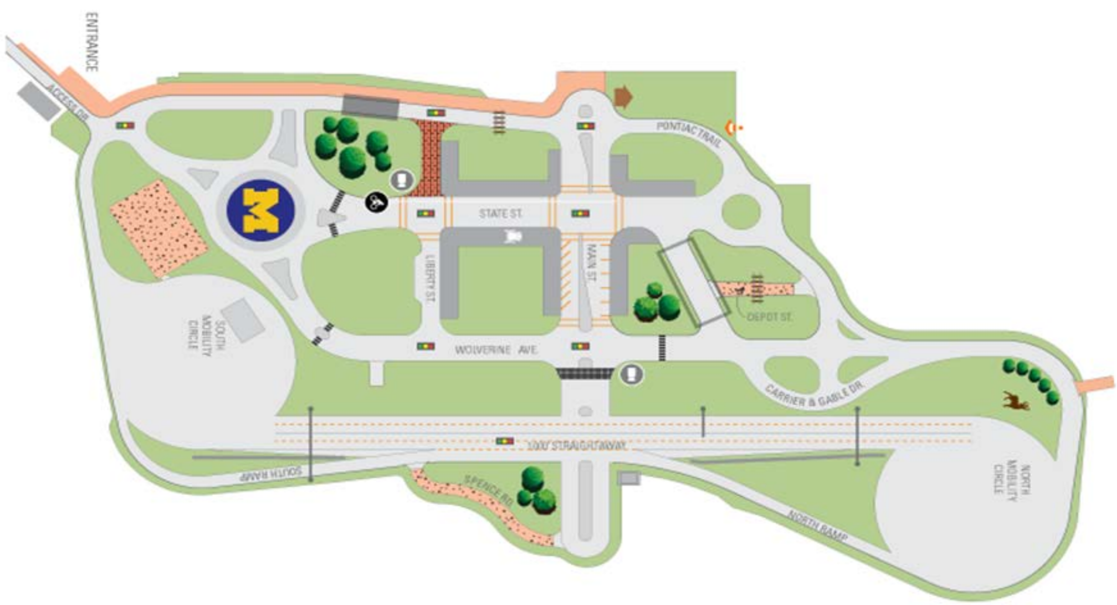

Figure 11. New Software Allows Users to Control Mcity Test Facility.
Source: University of Michigan (2019).

(4) Commercial Industry T&E Software, Simulation and Modeling Strategy for Autonomous Systems

According to Pecheur (2000) the forefront of M&S, the National Aeronautics and Space Administration (NASA) utilizes a suite of autonomous systems to support deep space exploration, from autonomous systems that work independently to keep astronauts on their trajectory towards their destinations, to advanced medical systems that alert ground control when astronauts' bodily vitals drastically change, or system functionality diagnoses a failure. NASA is at the forefront of utilizing autonomous software (Pecheur, 2000).

Based on the amount of unknown external factors and situations NASA systems encounter during space exploration, their systems and software engineers are required to assess a large range of situations when developing autonomous capabilities with limited information accessible. M&S is heavily relied upon as verification of autonomous applications. NASA software engineers utilize model checking, a technique that checks a program for violations or errors in every conceivable way early during the software development process. Software engineers contextualize a sub-system through a mathematic logic and analyze the model's properties during simulated situations to prove its functionality. Once a sub-system has been verified it can then be combined with multiple other structures called Binary Decision Diagrams (BDDs) and represent highly complex models that simulate entire systems. This complex model can then be used in a variety of different simulations and improved incrementally until satisfactory results are obtained.

Once a model has been validated, NASA uses analytical testing to run autonomous applications within unknown environments simulations. These unknown scenarios are simulated environment with ever-changing algorithms used to test the autonomous applications in a broad range of scenarios that have never been witnessed, a systematic exploration to identify errors within the software virtually. Analytical testing can provide a limitless amount of information prior to physical testing early in development of the hardware and at a fraction of the time and cost. Testing is done within the software in virtual simulations, parts, systems and materials can be altered reducing the overall cost of late development changes. NASA believes that analytic testing is the overall best path of increasing reliability of autonomous systems. Through the use virtual testing, NASA test evaluators can virtually simulate failure analysis and corrective action in parallel with development process.

E. OPERATIONAL IMPACT OF AUTONOMOUS SYSTEMS

The IOT&E is the final test event for defense acquisitions systems in development and a critical milestone in a program's life cycle. Required by the DOD and by law, it's conducted near a MS C decision to substantiate the operational effectiveness and suitability of a capability in its relative environment and used as the final data point in support of a full rate production decision.

Most technically complex capabilities being fielded by the DOD face a substantial number of hurdles during an IOT&E test. Defense acquisition regulations mandate that soldiers are required to have the proper training prior to test, to operate all systems' subsystems and trouble-shooting procedures prior to the event. Technical manuals are required to be detailed enough to walk soldiers through problems during operations, reliability failures and/or inadvertent user experiences. The IOT&E test is the final exam of the developmental phase of defense acquisition. Autonomous systems must contend with those issues as well as "system" trust and "operational" trust so that operators accept autonomous systems during critical situations. System trust is the operator's trust in the system to independently perform as required and operational trust is the operator's acceptance that the autonomous system is providing a benefit to perform the assigned task over traditional un-autonomous capabilities (Zwillinger, Palmer, & Selwyn, 2014). The U.S. Air Force chief scientist Werner Dahm believes that "Human mistrust of Automation/ Autonomy" is the "major barrier that prevents the USAF from gaining more capability from autonomous systems" (Helle, Strobel, & Schamai, 2016)

In the following chapter, we continue to compare T&E activities for the purpose of identifying specific challenges and future resolutions to DOD's methodology.

IV. ANALYSIS OF T&E METHODOLOGIES FOR AUTONOMOUS SYSTEMS

A. DOD’S T&E CHALLENGES OF AUTONOMOUS SYSTEMS AS COMPARED TO INDUSTRY BEST PRACTICES

The purpose of this JAP is to examine the current test methodologies of government programs with autonomous functionality as compared to the T&E methodologies of the private sector in order to identify successful practices that can be used as actionable changes to enhance the qualification and fielding of future DOD autonomous capabilities. This JAP utilizes the test and evaluation methodologies of four programs of records (SREHD, Gray Eagle, Project Perdix and the AMAS system) alongside similar industry systems within the automobile, agriculture, space and drone industry as a basis for comparison. The analysis section of this JAP outlines four significant weaknesses within the DOD autonomous systems T&E methodology as compared to current industry best practices. The data supporting these 4 (four) deficiencies is based upon research identified in the previous section comparing industry best practices alongside the T&E methodology of DOD programs of record.

Below we explore the following DOD T&E weaknesses

- DOD’s fragmented T&E process diminishes chances of fielding autonomous capabilities
- DOD’s lack of test facilities that can accommodate an autonomous capability concurrently alongside the development process
- DOD’s minimal use of M&S during T&E of new systems
- DOD’s inefficient Operational Testing of autonomous systems methodology

Major defense acquisition programs follow a standard development cycle; once a requirement has been established and approved by the user, those requirements are then taken by the program office and developed into a request for proposal to industry partners and awarded to the overall best proposal that meets the government's selection criteria. A contract is established, and contractors begin developing the system in conjunction with government oversight to ensure cost, schedule and performance metrics established within the contract are being met. Systems and subsystems are then tested and evaluated by the government T&E team to ensure performance requirements are being satisfied prior to full-rate production and fielding. The T&E process continues until the government is satisfied with the system's level of performance and/or the user accepts limitations against requirements. Operational Test and Evaluation (OT&E) activities, typically take place near the end of the engineering maturation phase and are both schedule and cost drivers. Technical issues identified during OT&E can be catastrophic during development, especially for autonomous functionality that is linked to human safety.

As in the case of SREHD's operational test event, in which operators were evaluated on how successful they could utilize all SREHD's capabilities to navigate lanes with hidden explosive devices. This test was conducted at the end of the DT&E phase due to the need to utilizing mature systems and software together. Thousands of hours were spent on hardware and software changes to ensure the system could meet the user requirements, but very little time was spent with the system in actual soldiers' hands. Per (James & Chang, 2017) the test was ultimately a failure because the consensus from the user representatives was that the system found to be non-operationally effective in comparison to the already fielded hand-held systems which were quicker and more, more versatile an in conjunction with the operator's awareness, more efficient at identifying threats.

In contrast, the autonomous systems industry has overcome these issues by establishing a continuous process of experimentation, testing and evaluation. Since there are no gaps in the development team structure, "real-world" data is continuously utilized to identify system problems and assist with better decision making. Statistics and human factor analysis are used throughout the system's life cycle to recognize system limitations early on and assist program managers with adjusting resources to better fit the needs of its customers. Commercial companies like autonomous electric vehicle manufacturer, Tesla and drone manufacturer DJI, both utilize a process of continuous real-world data collection throughout the entire life cycle of their autonomous systems. The continuous data collection enables quicker iterations of products to market and retroactively provide them with upgrades that increase the performance of their autonomous functionality. Continuous experimentation and data collection throughout development reduces the risk of operational test failures.

B. AVAILABILITY OF TEST RESOURCES THAT CAN ACCOMMODATE AUTONOMOUS SYSTEMS

In the DOD *Autonomy Report*, (2012) the Secretary of Defense, stated that "DOD's testing infrastructure will not be able to support future testing of new technologies such as hypersonic and autonomous systems without increased funding." This was also acknowledged by the DOT&E who rebutted major sections of DBB's report but affirmed that the current infrastructure of the DOT&E was not fully postured for the future and requested a follow-on study to assess the state-of-the-art technology to close the capability gap. Additionally, infrastructure was also reported to be an issue by the STAT T&E Center of Excellence which stated, "Many of the necessary processes, systems, test infrastructure and other capabilities simply do not exist, and may not have even been conceived at this point" (Parson & Ahner, 2016). The DOD needs to invest in test facilities and test evaluators that can accommodate an autonomous system in order to be successful in the future.

Based on the data collected for this JAP, all four case studies used to examine DOD's T&E methodology for autonomous systems utilized traditional man-in-the-loop type facilities and equipment as a means of qualification. In none of the four case studies were test facilities and/or equipment developed alongside the development of the autonomous capability to ensure that a better more efficient method of testing was available. Logically, autonomous systems are still an insignificant percentage of capability within the military and, with limited budgets, DOD has not prioritized funds for specialized facilities and equipment to test autonomous systems. However, until program managers, alongside the test and evaluation community, identify gaps in the T&E methodology of AS, programs will continue to be hampered by inefficient, expensive and unadaptable T&E methodology.

As an example, the SREHD T&E methodology required pre-existing lanes created for testing hand-held detectors. The AMAS's test facility is the Army's primary test venue for qualifying medium/heavy military vehicles; Gray Eagle is tested in airspace used to qualify the military's fleet of fighter jets; and Project Perdix is tested over China Lake, a naval weapon test site. Based on our research, currently none of these facilities use adaptive instrumentation needed to evaluate the system's ability to learn and adapt to its environment. Inadequate instrumentation becomes a critical issue when trying to identify the root cause of a failure and leaves test operators and data collectors vulnerable to "false positives" or test failures that are due to systems trying to adapt to environmental changes.

Alternatively, industries like the autonomous driving industry work together to advance test facilities for autonomous vehicles with the vision of modernizing vehicle proving ground facilities for the future. The University of Michigan and industry partners have invested $20 million in the Mcity test facility specifically to evaluate autonomous vehicles through a broad range of urban and suburban environments.

The Mcity proving ground can simulate several types of road surface conditions and structures while remotely monitoring a vehicle and passenger conditions. Mcity is also one of the only vehicles proving ground facilities to provide infrastructure capable of dedicated short-range communication, which is the future of autonomous vehicle communication enabling advanced coordination and cooperation that will reduce congestion and improve traffic flow in the future. The test facilities provide the autonomous vehicle industry a test bed to compare its simulated "virtual" performance with actual roadside data that can be evaluated in real time to help engineers identify changes and deficiencies quickly and turn around corrective alternatives.

Theoretically, if the AMAS IPT could duplicate the type of test facility that industry is using to qualify autonomous vehicles, DOD would have the capability of replicating multiple terrain types, environments, and threats at a single test facility versus having to fund and transport equipment, test hardware, and people to multiple test sites. The time and cost involved with transporting and maintaining multiple types of military vehicles to test and evaluate the AMAS system could be used more efficiently to expedite development of the overall test program.

Another significant limitation facing DOD's T&E community is the lack of expertise working with autonomous systems. A report issued by the Scientific Test & Analysis Techniques Center of Excellence, "Test and Evaluation of Autonomous Systems," identified that DOD's current T&E workforce lack personnel who can apply new methodologies for the future autonomous system testing.

Leading autonomous systems technology firms imbed their T&E teams within their design teams to shrink development times and increase system reliability. Companies such as Autonomous Solutions Inc. (ASI), the current market leader in autonomous farming technology, utilize a single facility to both design and validate their autonomous systems. Test engineers work in tandem with the designers, programmers, and mechanical engineers to expedite the validation process of subsystems in simulated "operational environments" to verify functionality early in the development process. (Nielsen, 2016) The synergy created by having onsite test facilities available early in the development process assists with the system education process and allows evaluators to understand the full range of capabilities of autonomous systems.

Based on the four provided examples, the T&E methodologies for the government autonomous programs are typically repurposed man-in-the-loop test methods and facilities that are utilized to verify and validate functionality. Most of these tests take place on MRTFB with personnel who do not specialize with autonomous technology and were not part of the technology demonstration phase of the program. In addition, these tests are typically large and exhaustive efforts to ensure the maximum amount of conditions are evaluated for stakeholders to determine a capabilities readiness.

Conversely, industry has identified the profit to be made by the growing "smart" commercial sector and have heavily invested in specialized personnel and facilities to rapidly provide autonomous products into the marketplace. Newer products are based off designs of older products and do not require an extensive amount of testing, only incremental validation of new functionality. Overall, industry embraces partnerships with agencies who specialize in the T&E of autonomous capabilities, which can rapidly iterate on test results and provide the latest technology to assist with the verification and validation. The differences between the two methodologies drastically impairs the DOD's ability to design, test and field autonomous capabilities.

The DOD, as a non-profit agent, must provide test resources for several thousand programs of record across the country, whereas emergent startup companies can profit off the autonomous systems boom and create niche T&E facilities. Autonomous programs of record are still in the infancy phase and DOD does not have the resources to convert the current T&E facilities to specialized autonomous systems test facilities. Program offices developing new autonomous systems, in conjunction with their prime contractors, will leverage private test facilities and personnel to level the playing field. PMO's will utilize the changes made to Army Regulation 73-1 (AR 73-1), which requires ATEC to minimize redundant testing and utilize both contractor data and test facilities to accelerate development and get capability into the warfighter's hands. In the future, as the need for more autonomous capabilities grow, the DOD will have to assess if there is still a need for government owned, government operated test facilities specifically for autonomous systems.

1. DOD's Minimal Utilization of M&S during T&E of New Systems

Based on the research for the four programs within this JAP, M&S data was only utilized as a tool to prototype and analyze maturing technology. This JAP was unable to identify any M&S being used as a T&E methodology alternative to real-world, physical hardware testing. In most test cases we identified within this JAP, the four systems used M&S to corroborate existing data and verify performance metrics prior to physically testing hardware. A primary reason the DOD doesn't use M&S more liberally throughout system development is because the T&E community lacks a robust M&S test capability that can stress validated models within an integrated virtual environment (J. Brabbs, S. Lohrer, P. Kwashnak, P. Bounker, M. Brudnak, 2019).

During the development of the AMAS system, M&S was utilized to simulate the AMAS architecture within a virtual testbed. Engineers were able to evaluate what the AMAS system was capable of prior to physically running the system on a vehicle. Unfortunately, because of how the AMAS system uses several sensors to capture and relay information to the vehicle, the T&E community was unable to use the information to pre-plan test scenarios or capture any valid performance metrics. Vehicles were still required to be equipped with the AMAS system for testers to complete their analysis of the capability (D. Pirozzo, J.P. Hecker, A. Dickinson, T. Schulteis, J. Ratowski, and B. Theisen, 2019).

Gray Eagle was similarly developed with the same methodology. Government test and evaluators only utilized a virtual model of the Gray Eagle after the capability was complete and real-world data could be acquired to verify performance. Currently, the only government utilization of virtual simulation is to train Gray Eagle operators, but virtual testing could have been utilized to mitigate reliability issues that plagued the system early in development.

Project Perdix hasn't undergone formal developmental testing but there is no indication that DOD test and evaluators are utilizing virtual simulation software to evaluate performance. Only test data that was obtainable was taken from real world testing, evaluated for the purpose of increasing the size of the drone swarm, and evaluating new test environments. Simulation and modeling could be utilized during initial test events to mitigate user interface issues and determine ideal conditions for future test events.

Finally, during the development of the SREHD capability, the government did not utilize SREHD M&S data to validate changes made to the system. Every single design change required additional tests that resulted in additional time and cost against the acquisition program baseline. The government IPT did not request SREHD software models be kept up to date with changes made later in development and was unable to utilize virtual simulation to validate/verify critical reliability engineering change proposals (ECPs) efficiently. The only means of verification was to physically test and evaluate system performance by running the system over test lanes at Yuma Proving Grounds and AP Hill test facilities. If the government development team had contractually obligated the SREHD contractor to maintain virtual models of the SREHD system, months of test and evaluation could have been reduced without requiring a 12-month period of performance (PoP) extension and a substantial cost overrun.

Alternatively, the federally funded National Aeronautics and Space Administration, (NASA) does utilize simulation and modeling at the beginning of the development process to test and validate concepts prior to design completion. Once models have been validated, NASA engineers can run virtual software simulations of entire complex systems within new environments to identify errors and increase reliability. Additionally, since testing takes place within a virtual environment, parts, materials and subsystems can be altered quickly, reducing the overall cost of late development changes.

Programs like SREHD, utilized M&S early in development, as a means for engineering teams to plan budgets, resources and judge the complexity of technical requirements. Once initial prototypes were completed, resources were focused on getting hardware and software mature enough to meet requirements but without regard of maintaining those changes to the virtual simulation. Developing complex software structures such as NASA's binary decision diagrams (BDDs) to simulate autonomous functionality before and throughout the development of autonomous systems, is a proven method of reducing both the cost and time required to physically validate autonomous functionality in complex systems.

2. DOD's Inefficient Operational Testing Of Autonomous System Methodology

Based on the information being presented, it is more likely that industry will be more capable of reducing the amount of resources required during T&E to bring autonomous systems to the market. The primary reason for our hypothesis is that the commercial sector for autonomous products is smaller, more niche and capable of focusing on data derived from experimentation during development testing which can be used to seize the scale of relations and reduce the need for redundant and inefficient testing. This enables industry to focus future tests on only the most critical components of an autonomous capability. The autonomous automotive industry already utilizes test facilities that incorporate the "Hue Peng" methodology of testing the I.Q. of autonomous vehicles rather than its ability to make identical decisions given different types of situations. This methodology not only increases the level of awareness of the autonomous capability but also reduces the amount of testing required to validate complex systems (Peng & Zhao, 2017). The Hue Peng methodology is in direct conflict with DOD’s T&E methodology that requires testers to verify and validate all system parameters. This methodology adapted to the Gray Eagle’s autonomous liftoff and landing capability could have reduced the time needed to qualify this unmanned system.

The DOD uses information gained during the OT&E period to assess a system’s operational effectiveness and operational suitability during a realistic simulated combat scenario. The *Role of Autonomy in DOD Systems* by the Defense Science Board (2012 summed up the issue of testing autonomous systems best in their report, “Testing all System of System requirements of all the systems is impossible.” With the inception of autonomous systems on the battlefield, DOD needs to dynamically change its T&E methodology for operational testing to a more efficient and effective means of verifying and validating operational requirements that will assist in accelerate development and fielding.

Operational test events planned for the end of the EMD phase need to evolve into operational experimentation initiated in the technical demonstration phase The DOD has acknowledged the issues of the standard FAR based acquisition process and provided new guidance called, middle tier acquisition (MTA) authority issued by Congress in the FY16 National Defense Authorized Act (NDAA) Section 804. The authorization establishes an interim adaptive acquisition process that enables the PMO the opportunity to accelerate capability development and field prototype hardware within 5 years. MTA provides the PMO the authority to experiment with mature technologies in an operational environment from the beginning of the acquisition and enter into production within 6 months (Office of the (Under Secretary of Defense for Acquisition and Sustainment [USDAS], 2020).

Each PMO needs utilize tools and authorities like MTA to incorporate an operational experimentation strategy that gets autonomous systems in the hand of the actual warfighter as soon as feasibly possible. The benefits of early feedback by an actual user representative in a relevant environment will assist in both, shaping future test events while accelerating development. Typically, autonomy is incorporated into military systems to provide functionality that reduces burdens for safety critical tasks, such as maintaining head height of a sensor on a mine detection system or constantly monitoring communication systems for the purpose of safely navigating the skies. The T&E of autonomous systems require the evaluation of these system's behavior in both defined and undefined environments. System autonomy is derived from machine learning algorithms which inherently behave in varies ways to the same inputs overtime, since their knowledge increases over multiple operations over time. Fundamentally this means "that one successful test does not guarantee that the system will pass the same test on the next run" (Art & Philip & Helle, 2016). Non-deterministic behavior is a key challenge facing the DOD T&E community, requiring investment into advanced test simulation techniques, that significantly reduce the amount of test scenarios needed, in favor of fewer, more effective test. In the case of SREHD, early experimentation would have assisted the PMO in directing more development with the user interface and mitigated negative user feedback late in the development cycle.

The DOD's issues with aging test facilities, constrained resources for T&E and strict adherence to processes designed around testing "manned" systems is a self-imposed problem and unique to DOD's acquisition process. The commercial sector for autonomous systems created new and adaptive T&E methodologies specific to the needs of its products. As in the case for the automotive industry utilizing the "Hue Peng" methodology, which reduces the time spent, testing vehicles on the road to specifically what is required to verify the autonomous systems decision-making process.

Alternatively, the autonomous systems farming industry partners with their customer base specifically to obtain better T&E data during the development process. The DOD equivalent to this example would be for a system like the AMAS capability to be operationally tested during development by a deployed military transportation division. Unfortunately, the program stagnated and never progressed past the technology demonstration phase and has since transitioned into the Expedient Leader Follower (ExLF) program (DEFPOST 2018). Real-world developmental testing alongside laboratory test and evaluation is a critical step in advancing the T&E methodology for autonomous programs or record.

V. CONCLUSIONS

A. ADAPTING INDUSTRY'S T&E METHODOLOGY

The purpose of this JAP was to collect information on DOD's current T&E methodology of autonomous systems through the analysis of case studies on four current programs of record (Gray Eagle, SREHD, and AMAS & Project Perdix) and identify corresponding industry T&E trends that could be used to successfully enhance DOD's T&E methodology for the future. Drawing comparisons between the defense acquisition process and private industry development process can be dangerous. The DOD is a non-profit organization that is responsible for ensuring the protection of U.S. citizens and national interests around the world and by proxy, that its military is outfitted with the most effective weaponry.

Private industry on the other hand is more concerned with developing products around commercial interests, maximizing profit and ensuring products meet the minimum safety standard required by government regulation. Pass/fail decisions are less relevant for commercial viability. This JAP acknowledges that not all best practices and commercial trends can be transitioned from industry to DOD's T&E methodology but based off of the analysis of the four systems within the JAP, we can conclude that the current DOD T&E model has significant weaknesses and recommend several industry T&E best practices that could be adopted by the DOD T&E community to advance methodology for military autonomous systems development.

B. ADVANCING DOD'S T&E METHODOLOGY FOR AUTONOMOUS SYSTEMS

As we framed our ideas for the JAP, we started out by seeking answers to two specific questions.

(1) Primary:

- What unique challenges are presented by autonomous systems to the Department of Defense's test and evaluation communities?

Based on the research covered within this JAP, the DOD T&E community face several unique challenges that industry does not have to overcome. The challenges the DOD face in the T&E process for autonomous systems is like the development of other highly complex government acquisition PORs. The critical obstacle is to determine how the T&E community assesses the non-deterministic behavior of autonomous systems. We determined that answer to this question is exacerbated by the four specific challenge areas within the DOD T&E acquisition process for AS.

- DOD’s fragmented T&E process that diminishes chances of fielding autonomous capabilities
- DOD’s lack of test facilities that can accommodate an autonomous capability concurrently alongside the development process
- DOD’s minimal use of M&S during T&E of new systems
- DOD’s inefficient Operational Testing of autonomous systems methodology

This JAP concludes that the hurdles facing the DOD T&E community are due to current infrastructure, equipment and technical personnel that were created around the T&E of man-in-the-loop type systems. The military’s autonomous capabilities of tomorrow require facilities that are capable of testing autonomous subsystems and systems in sync with development, as well as the trained personnel required to evaluate AI and machine learning over the life cycle of the system. The DOD T&E community need access to developmental data earlier in the life cycle of major development programs in order to efficiently evaluate the growth and maturity of autonomous systems. Additionally, DOD’s T&E methodology has not sufficiently embraced M&S from the beginning of the life cycle, prohibiting virtually simulated testing of autonomous functionality early on, that would reduce the amount of time and resources needed to qualify systems. Finally, the DOD T&E methodology for operational testing requires a paradigm shift to be more effective and efficient at evaluating autonomous systems in relevant, realistic environments. Industry already embraces early real-world testing of autonomous systems for the purpose of evaluating the suitability and effectiveness of potential designs. The DOD should initiate methods of parallel verification and validation of operational experimentation, that would enable expeditionary forces to evaluate the effectiveness and suitability of autonomous systems prior to and throughout the EMD phase.

(2) • Secondary:

- Does DOD have the correct resources to conduct tests and evaluations of autonomous systems?

The DOD T&E community face unique challenges based on regulatory and statutory guidelines in the Federal Acquisition Regulation process, which prohibit adaptive and reflexive strategies required to keep pace with the constantly changing technologies of tomorrow. The DOD is also constrained by not having specialized personnel and processes in place that maximize the unique capabilities of autonomy. Examples like Mcity in Michigan developing specific test facilities and technologies to evaluate vehicles with autonomous capabilities or agencies like the FAA and USDOT developing policy's for the advancement of T&E research are evidence that the T&E community supporting the autonomous systems industry has rapidly developed new and better technologies. Even though most POR are developed in conjunction with defense contractors, most of the verification and validation of requirements still take place at DOD run facilities, limiting access to the latest innovative T&E methods and techniques. Some of these issues can be addressed, through the use of the adaptive acquisition framework, providing PMO's the authority to determine and tailor the most efficient method of acquiring, maturing and fielding a capability, delivering a better solution faster to the warfighter (USDAS 2020). Through the use of MTA, the PMO has the ability to leverage new technologies, contractor test data and government experimentation data for the purpose of maturing and fielding a capability that would otherwise take several years based on the typical major capability acquisition pathway.

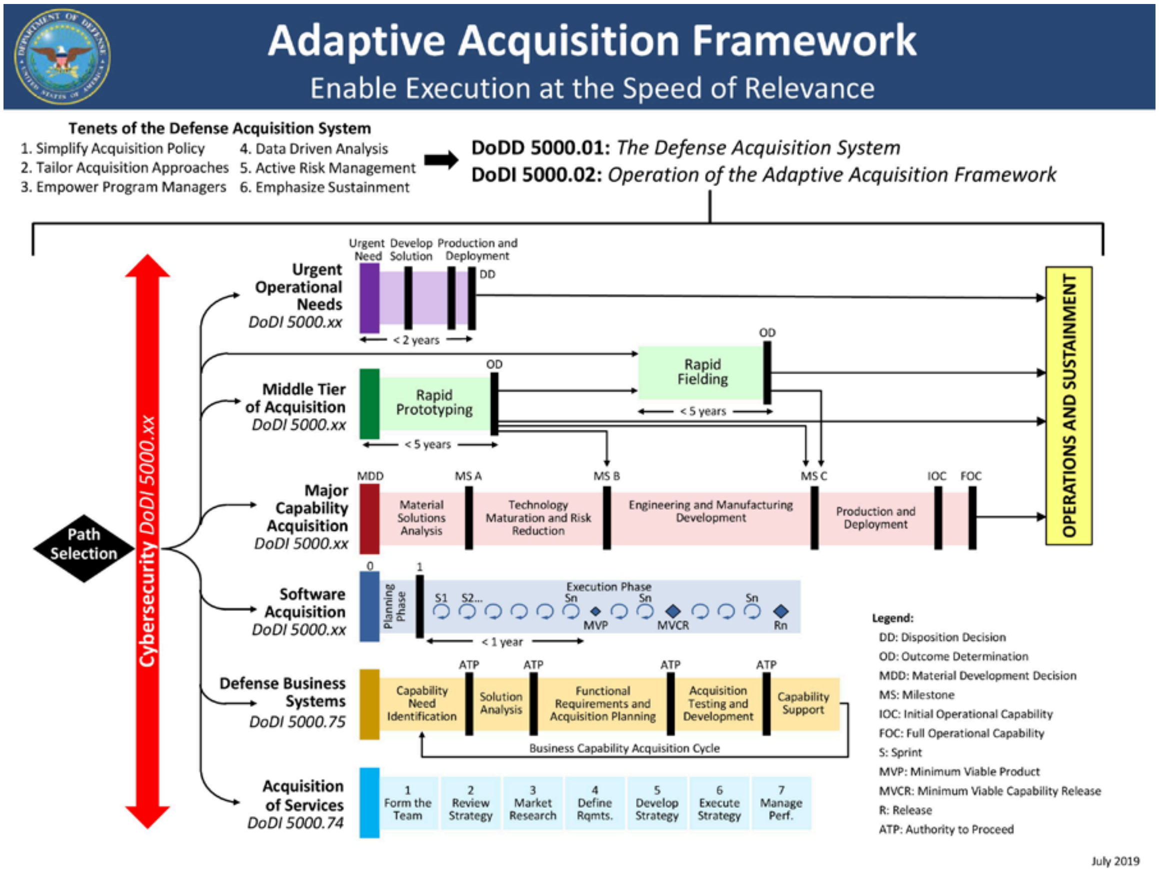

Figure 12. DoDI 5000.02 Adaptive Acquisition Framework. Source: Office Under Secretary of Defense for Acquisition and Sustainment [USDAD] AAF (2020).

C. FINAL THOUGHTS

Through the exploration of the DOD T&E methodology for autonomous systems, this JAP was able to identify several keyways in which DOD can advance its T&E process to be more in line with industry trends.

- DOD autonomous systems PORs should consider the addition of clauses in contracts that require contractors to create, maintain and utilize virtual models for the specific purpose of efficiently simulating test of autonomous functionality. These clauses can be used to increase the level of confidence of virtual models that can expedite development test and evaluation.

- DOD needs to increase funding and education in support of advancing the autonomous test capabilities at APG and YPG and test facilities. The education of the T&E community and restoration of DOD test facilities are critical to providing the warfighter the cutting edge in capability. To include MRTFBs retaining the capability to replicate multiple terrain types, environments, and threats at a single test facility versus having to fund and transport equipment, test hardware, and people to multiple test sites.

- DOD should consider incentiving PORs to develop T&E strategies that include parallel verification and validation of operational effectiveness and suitability during the EMD phase. Continuous operational evaluation of autonomous subsystems would provide contractors valuable data on the effects of machine learning outside of the laboratory environment, provide decision makers valuable information on how better to manage funding needed for corrective action strategies, and give users information on how to tailor requirements to only what is critical for operational effectiveness. With the use of expeditionary forces providing globally deployed soldiers' innovative solutions to address urgent needs operational, T&E development process should be included to provide as much information as soon as possible on operational effectiveness and suitability.

- DOD must establish a more consistent process for experimentation, T&E and updating already fielded autonomous capabilities. To note, defense programs like Gray Eagle, have a process in place to incrementally add enhancements but the ability to improve upon autonomy post initial fielding is only reserved for the larger MDAP initiatives. Advances in technology have enabled the tech industry to sell products in the marketplace with a limited feature set initially and continue to increase

> that products' capabilities over time. If DOD were to adopt a similar methodology of fielding autonomous systems with limited capability and over time increasing functionality to bring enhanced autonomy retroactively, then DOD autonomous systems would require less test and evaluation, at a reduced cost and less schedule delays.

Ironically, it is now industry leading the way in the technology fields of artificial intelligence, machine learning, and virtual simulated testing that originally were government defense research initiatives. Fortunately, during the time it took to evolve our JAP, the DOD has initiated policies and postures that are intended to totally change the dynamic of how the military brings advancements and capability to the field. Army Future Command is the Army's primary command stood up in 2019, for the purpose of soliciting industry's best and brightest to rapidly evolve technology and demonstrate capability quickly for the warfighter. The Army has recognized that only through embracing rapid innovation of industry will we be able to keep ahead of our near-pear threats. It is our hope that the DOD will keep pushing ahead and use the information contained within this JAP to continue creating new policies and regulations that will facilitate the T&E of autonomous systems within DOD systems.

LIST OF REFERENCES

Ahner, D.K, & Parson, C.R. (2016). Workshop report: Test and evaluation of autonomous systems.: STAT T&E Center of Excellence.

Brabbs, J. et al. (2019, August 13–15). “M&S as the key Enabler for Autonomy Development, Acquisition and Testing,” In Proceedings of the Ground Vehicle Systems Engineering and Technology Symposium (GVSETS), NDIA, Novi, MI

Business Insider Intelligence. (2016, April 24). The Drone Report: Market forecasts, regulatory barriers, top vendors, and leading commercial applications. https://www.businessinsider.com/the-drones-report-market-forecasts-regulatory-barriers-top-vendors-and-leading-commercial-applications-2015-8

Defense Acquisition Management Information Retrieval (DAMIR) (2010). Selected Acquisition Report (RCS: DD-A&T(Q&A)823-420). MQ-1C Gray Eagle.

Defense Modeling & Simulation Coordination Office. (2019, December 13). Home page. https://www.msco.mil/

Defense Science Board. (2012). Task force report: The role of autonomy in DOD Systems. https://fas.org/irp/agency/dod/dsb/autonomy.pdf

Department of the Army. (2018). Test and evaluation policy. (AR 73-1). Headquarters Department of Army.

Director Operational Test and Evaluation (December, 2010). FY2010 Annual Report.https://www.dote.osd.mil/Publications/Annual-Reports/2010-Annual-Report/

Edan, Y., Han, S., & Kondo, N. (2009). Spring Handbook of Automation: Automation of Agriculture.

Federal Aviation Administration. (2013). FAA selects six sites for unmanned aircraft research. https://www.faa.gov/news/updates/?newsId=75399

General Atomics. (2012), Gray Eagle UAS Completes 20,000 Automated Takeoffs and Landings. http://www.ga.com/gray-eagle-completes-20000-automated-takeoffs-and-landings

Hansman, J. (2011). Project Perdix, Beaver Works Lincoln Laboratory School of Engineering. https://beaverworks.ll.mit.edu/CMS/bw/projectperdixcapston

Helle, P., & Schamai, W., Strobel., C. (2016). Testing of Autonomous Systems-Challenges and Current State-of-the-Art. www.incose.org>docs>default-source>enchantment.

James, L. Maj, & Chan, K., (6 March 2017). SREHD Shreds the handheld competition. Army ALT Magazine, Science and Technology.

Lockheed Martin Corporation (2018). Autonomous Mobility Applique System (AMAS): https://www.lockheedmartin.com/en-us/products/autonomous-mobility-applique-system-amas.html.

Moore, N (2017). New way to test self-driving cars could cut 99.9 percent of validation costs. https://news.umich.edu/new-way-to-test-self-driving-cars-could-cut-99-9-percent-of-validation-costs/

Nielson, M (2016, August 30) Autonomous Solutions, Inc. and CNH Industrial unveil concept autonomous tractor. https://www.asirobots.com/autonomous-solutions-inc-and-cnh-industrial-unveil-concept-autonomous-tractor/

Office of the Under Secretary of Defense for Acquisition and Sustainment. (2020 Jan 23). Operation of the Adaptive Acquisition Framework. Department of Defense. https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodi/500002p.pdf?ver=2020-01-23-144114-093

Office of the Under Secretary of Defense for Acquisition and Sustainment. (2019 Dec 30). Operation of the Middle Tier of Acquisition (MTA).Department of Defense. https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodi/500080p.PDF?ver=2019-12-30-095246-043

Office of the Under Secretary of Defense for Acquisition, Technology and Logistics Washington, D.C. (2012). Defense Science Board. The Role of Autonomy in DOD System.

Pecheur, C (2000, August). Verification and Validation of Autonomy Software at NASA. National Aeronautics and Space Administration. https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20010000882.pdf

Peng, H., & Zhao, D. (15 July, 2017), From the lab to the Street: Solving the Challenge of Accelerating Automated Vehicle Testing. https://www.semanticscholar.org/paper/From-the-Lab-to-the-Street%3A-Solving-the-Challenge-Zhao-Peng/855c84035458abcac6ab510b187bddb3df86d793

Pirozzo, D., Hecker, J.P., & Dickinson., & A, Schulteis,, & T., Ratowski, J., & Theisen, B,. (2019, August 15). Integration of the Autonomous Mobility Appliqué System into the Robotic Technology Kernel, In Proceedings of the Ground Vehicle Systems Engineering and Technology Symposium (GVSETS).

Scharre, P., (2018). Army of None: Autonomous Weapons and the Future of War. W. W. Norton & Company.

University of Michigan (20 August, 2019). New Software Allows Users to Control MCity Test Facility. https://mcity.umich.edu/new-software-allows-users-to-control-mcity-test-facility/.

Zhao, D., & Peng, H. (2017). From the Lab to the Street: Solving the Challenge of Accelerating Automated Vehicle Testing. https://arxiv.org/abs/1707.04792

Zilberstein, S. (2015). Building Strong Semi-Autonomous Systems. [Unpublished manuscript].

THIS PAGE INTENTIONALLY LEFT BLANK

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California