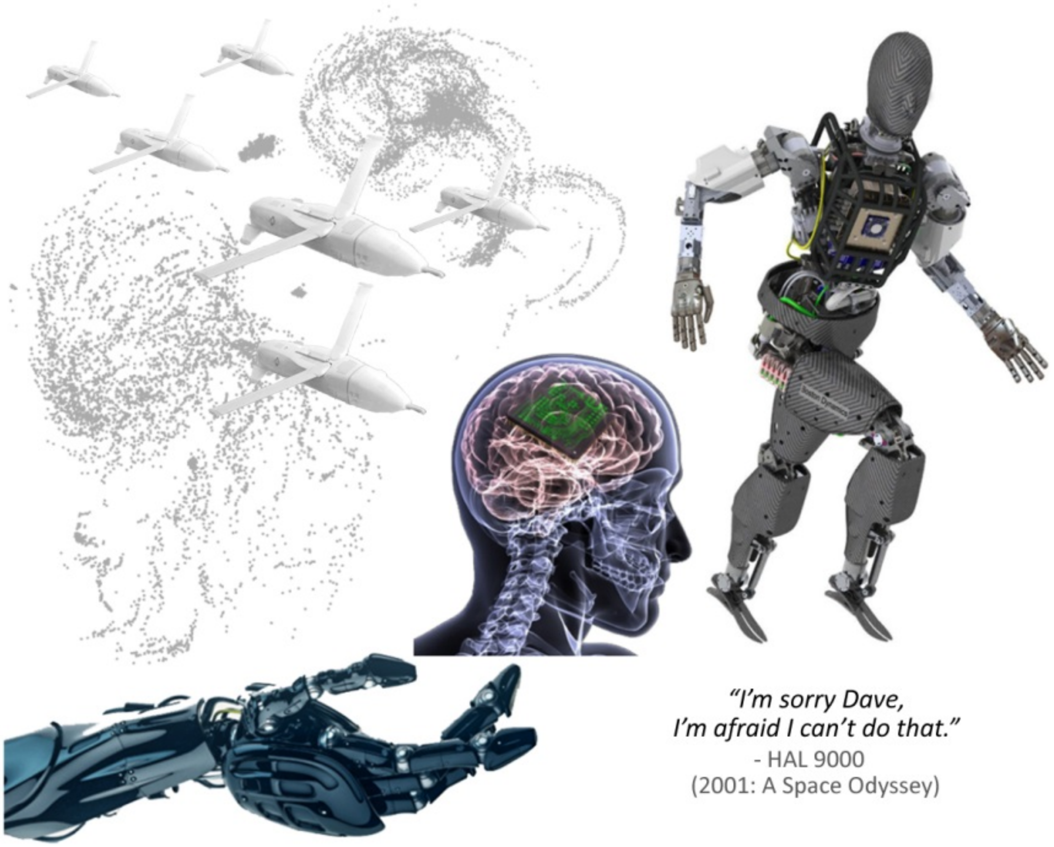
AI, Robots, and Swarms

Issues, Questions, and Recommended Studies

Andrew Ilachinski

January 2017



"I'm sorry Dave,
I'm afraid I can't do that."
- HAL 9000
(2001: A Space Odyssey)

CNA
ANALYSIS & SOLUTIONS

CNA
ANALYSIS & SOLUTIONS

This document contains the best opinion of CNA at the time of issue.
It does not necessarily represent the opinion of the sponsor.

Distribution

Approved for Public Release; Distribution Unlimited. Specific authority: N00014-11-D-0323.

Copies of this document can be obtained through the Defense Technical Information Center at www.dtic.mil or contact CNA Document Control and Distribution Section at 703-824-2123.

Photography Credits: http://www.darpa.mil/DDM_Gallery/Small_Gremlins_Web.jpg; http://4810-presscdn-0-38.pagely.netdna-cdn.com/wp-content/uploads/2015/01/Robotics.jpg; http://i.kinja-img.com/gawker-edia/image/upload/18kxb5jw3e01ujpg.jpg

Approved by: **January 2017**

Dr. David A. Broyles
Special Activities and Innovation
Operations Evaluation Group

Abstract

The military is on the cusp of a major technological revolution, in which warfare is conducted by unmanned and increasingly autonomous weapon systems. However, unlike the last “sea change,” during the Cold War, when advanced technologies were developed primarily by the Department of Defense (DoD), the key technology enablers today are being developed mostly in the commercial world. This study looks at the state-of-the-art of AI, machine-learning, and robot technologies, and their potential future military implications for autonomous (and semi-autonomous) weapon systems. While no one can predict how AI will evolve or predict its impact on the development of military autonomous systems, it is possible to anticipate many of the conceptual, technical, and operational challenges that DoD will face as it increasingly turns to AI-based technologies. This study examines key issues, identifies analysis gaps, and provides a roadmap of opportunities and challenges. It concludes with a list of recommended future studies.

This page intentionally left blank.

Executive Summary / White Paper

A notable number of groundbreaking artificial intelligence (AI)-related technology announcements and/or demonstrations took place in 2016:[1]

1. AI defeated the reigning world champion in the game of Go, a game that is so much more "complex" than chess that, prior to this event, most AI experts believed that *it could not be done for another 15-20 years*.[2]
2. AI learned—*on its own*—where to find the information it needs to accomplish a specific task.[3]
3. AI predicted the immediate future (by generating a short video clip) by *examining a single photograph* (and is also able to predict the future from studying video frames).[4]
4. AI automatically inferred the rules that govern the behavior of individual robots within a robotic swarm *simply by watching*.[5]
5. AI learned to navigate the London Underground *by itself* (by consulting its own acquired memories and experiences, much like a human brain).[6]
6. AI speech recognition reached human parity in conversational speech.[7]

[1] Most of the innovations on this list are described in the *Artificial Intelligence* section of the main narrative of this report (pp. 44-71). A few others also appear in the appendix.

[2] C. Koch, "How the Computer Beat the Go Master," *Scientific American*, 19 March 2016.

[3] K. Narasimhan et al., "Improving Information Extraction by Acquiring External Evidence with Reinforcement Learning," presented at EMNLP 2016, https://arxiv.org/abs/1603.07954.

[4] C. Vondrick, H. Pirsiavash, and A. Torralba, "Generating Videos with Scene Dynamics," presented at the 29th Conference on Neural Information Processing Systems, Barcelona, Spain, 2016: http://web.mit.edu/vondrick/tinyvideo/paper.pdf.

[5] W. Li, M. Gauci, and R. Gross, "Turing learning: a metric-free approach to inferring behavior and its application to swarms," *Swarm Intelligence* 10, no. 3, September 2016: http://link.springer.com/article/10.1007%2Fs11721-016-0126-1.

[6] E. Gibney, "Google's AI reasons its way around the London Underground," *Nature*, Oct 2016.

[7] X. Xiong et al., "Achieving Human Parity in Conversational Speech Recognition," *arXiv*, 2016: https://arxiv.org/abs/1610.05256.

7. An AI communication system *invented its own encryption scheme*, without being taught specific cryptographic algorithms (and without revealing to researchers how its method works).[8]
8. An AI translation algorithm invented its own "interlingua" language to more effectively translate between any two languages (*without being taught to do so by humans*).[9]
9. An AI system *interacted with its environment* (via virtual actuators) to learn and solve problems in the same way that a human child does.[10]
10. An AI-based medical diagnosis system at the Houston Methodist Research Institute in Texas achieved 99% accuracy in reviewing millions of mammograms (at a rate 30× faster than humans).[11]

These and other recent similar breakthroughs (e.g., IBM's *Watson's* defeat of the two highest ranked *Jeopardy!* players of all time in 2011),[12] are notable for several reasons. First, they collectively provide evidence that we, as a species, have already crossed over into an era in which seeing AI outperform humans—at least for specific tasks—is *almost* routine (perhaps in the same way that landing on the moon was "almost" routine after the first few Apollo missions).[13] Second, they offer a glimpse of how *different* AI is from human intelligence, and how inaccessible its "thinking" is to outside probes. And third, they demonstrate the power of AI to *surprise* us (including AI system developers, who nowadays are closer in spirit to "data collectors" and "trainers" than to traditional programmers)—i.e., AI, at its core, is fundamentally *unpredictable*. In the second game of the Go match between the AI that defeated Lee SeDol (an 18-time world champion in Go), the AI made a move so surprising that

[8] M. Abadi and D. Andersen, "Learning to Protect Communications with Adversarial Neural Cryptography," arXiv:1610.06918v1: https://arxiv.org/abs/1610.06918.

[9] Q. Le and M. Schuster, "A Neural Network for Machine Translation, at Production Scale," Google Research Blog, 27 Sep 2016: https://research.googleblog.com/2016/09/a-neural-network-for-machine.html.

[10] M. Denil, P. Agrawal, T. Kulkarni, et al., "Learning to perform physics experiments via deep reinforcement learning," under review as a conference paper to ICLR 2017: https://arxiv.org/pdf/1611.01843v1.pdf.

[11] T. Patel et al., "Correlating mammographic and pathologic findings in clinical decision support using NLP and data mining methods," *Cancer* 123, 1 Jan 2017.

[12] S. Baker, *Final Jeopardy: Man vs. Machine and the Quest to Know Everything*, Houghton Mifflin Harcourt, 2011.

[13] Unlike the Apollo program, however, AI is here to stay: *Artificial Intelligence and Life in 2030: One Hundred Year Study on Artificial Intelligence*, Report of the 2015 Study Panel, Stanford University, Sep 2016.

SeDol had to leave the room for 15 minutes to recover his composure: "It's not a human move. I've never seen a human play this move. So beautiful."[14]

The breakthroughs listed above are also notable for a fourth reason—a more subtle one, but the one that directly inspired this study. Namely, they portend a set of deep conceptual and technical challenges that the Department of Defense (DoD) must face, now and in the foreseeable future, as it embraces *AI*–, *robot*–, and *swarm*–related technologies to enhance (and weaponize) its fleet of unmanned systems with higher levels of autonomy. The subtlety lies in unraveling the true meaning of the deceptively "obvious" word, *autonomy;* indeed, as of this writing, there is no universally accepted definition.

Autonomous weapons—colloquially speaking—have been used since World War II (e.g., the German *Wren* torpedo's passive acoustic homing seeker effectively made it the world's first autonomously guided munition).[15] Human-supervised automated defensive systems have existed for decades, and aerial drones were first used more than 20 years ago (i.e., the RQ-1 Predator was used as an intelligence, surveillance, and reconnaissance platform in former Yugoslavia).[16] But it was only after the September 11, 2001, terrorist attacks that the military's burgeoning interest in, and increasing reliance on, unmanned vehicles started in earnest. In just 10 years, DoD's inventory of unmanned aircraft grew from 163, in 2003, to close to 11,000, in 2013 (and, in 2013, accounted for 40% of *all* aircraft).[17] And the United States is far from being alone in its interest in drones: by one recent tally, at least 30 countries have large military drones, and the *weaponized* drone club has recently grown to 11 nations, including the United States.[18]

DoD procured most of its medium-sized and larger unmanned aerial vehicles (UAVs), the MQ-1/8/9s and RQ-4/11s, for the counterinsurgency campaigns in Iraq and Afghanistan, where the airspace was largely uncontested. Now the United States is withdrawing from those campaigns and the military is shifting its strategic focus to less permissive operating environments (i.e., the Asia-Pacific region) and to adversaries with modern air defense systems. Thus, there is a growing emphasis on developing new, more *autonomous*, systems that are better equipped to survive in more contested airspaces.

[14] C. Metz, "The Sadness and Beauty of Watching Google's AI play Go," *Wired*, 11 March, 2016.

[15] J. Campbell, *Naval Weapons of World War Two*, Naval Institute Press, 2002.

[16] P. Springer, *Military Robots and Drones: A Reference Handbook*, ABC-CLIO, 2013.

[17] *Unmanned Systems Integrated Roadmap: FY2013-2038*, U.S. Department of Defense, 2013.

[18] *World of Drones: Military*, International Security Data Site, New America Foundation: http://securitydata.newamerica.net/world-drones.html.

Fundamentally, an autonomous system is a system that can independently compose and select among alternative courses of action to accomplish goals based on its knowledge and understanding of the world, itself, and the local, dynamic context. Unlike automated systems, autonomous systems must be able to respond to situations that are not pre-programmed or anticipated prior to their deployment. In short, autonomous systems are inherently, and irreducibly, *artificially intelligent robots*. In the remaining pages of this summary, we explicate the analytical implications of this assertion (leaving details and supporting evidence to the main narrative).

To start, if and when autonomous systems, in the sense just described, finally arrive, they will offer a variety of obvious advantages to the warfighter. For example, they will eliminate the risk of injury and/or death to the human operator; offer freedom from human limits on workload, fatigue, and stress; and be able to assimilate high-volume data and make "decisions" based on time scales that far exceed human ability. If robotic swarms are added into the mix, entirely new mission spaces potentially open up as well—e.g., wide-area, long-persistence, surveillance; networked, adaptive electronic jamming; and coordinated attack. There are also numerous advantages to using swarms rather than individual robots, including: *efficiency* (if tasks can be decomposed and performed in parallel), *distributed action* (multiple simultaneous cooperative actions can be performed in different places at the same time), and *fault tolerance* (the failure of a single robot within a group does not necessarily imply that a given task cannot be accomplished).

However, the design and development of autonomous systems also entails significant conceptual and technical challenges, including:

- *"Devil is in the details" research hurdles:* Developers of autonomous systems must confront many of the same fundamental problems that the academic and commercial AI and robotic research communities have struggled for decades to "solve." To survive and successfully perform missions, autonomous systems must be able to sense, perceive, detect, identify, classify, plan for, decide on, and respond to a diverse set of threats in complex and uncertain environments. While aspects of all these "problems" have been solved to varying degrees, there is, as yet, no system that fully encompasses all of these features.

- *Complex and uncertain environments:* Autonomous systems must be able to operate in complex—possibly, a priori unknown—environments that possess a large number of potential states that cannot all be pre-specified or be exhaustively examined or tested. Systems must be able to assimilate, respond to, and adapt to dynamic conditions that were not considered during their design. This "scaling" problem—i.e., being able to design systems that are developed and tested in static and structured environments, and then have them perform as required in dynamic and unstructured environments—is highly nontrivial.

- *Emergent behavior:* For an autonomous system to be able to adapt to changing environmental conditions, it must have a built-in capacity to learn, and to do so without human supervision. It may be difficult to predict, and be able to account for *a priori* unanticipated, emergent behavior (a virtual certainty in sufficiently "complex" systems-of-systems dynamical systems).

- *Human-machine interactions/I:* The operational effectiveness of autonomous systems will depend on the dynamic interplay between the human operator and the machine(s) in a given environment, and on how the system responds, in real time, to changing operational objectives, in concert with the human's own adaptation to dynamic contexts. The innate unpredictability of the human component in human-machine collaborative performance only exacerbates the other challenges identified on this list.

- *Human-machine interactions/II:* The interface between human operators and autonomous systems will likely include a diverse space of tools that include visual, aural, and tactile components. In all cases, there is the challenge of translating human goals into computer instructions (e.g., "solving" a long-standing "AI problem" of natural language processing), as well as that of depicting the machine's "decision space" in a form that is understandable by the human operator (e.g., allowing the operator to answer the question, "Why did the system choose to take action X?").

- *Control:* As autonomous systems increase in complexity, we can expect a commensurate decrease in our ability to both predict and control such systems—i.e., the "spectre of complacency in complexity." As evidenced by the general nature of recent AI breakthroughs, there is a fundamental tradeoff: either the AI can achieve a given performance level (e.g., it can play the game Go as well as, or better than, a human), or humans can be able to understand how its performance is being achieved).

Apart from these innately technical challenges to developing autonomous systems, there are a set of concomitant acquisition challenges, the origin of which is a recent shift in DoD's innovation-related procurement practices. While the U.S. government has always played an important role in fostering AI research (e.g., ARPA, DARPA, NSF, ONR), most key innovations in AI, robotics, and autonomy are now being driven by the *commercial sector*,[19] and at a pace that DoD's relatively plodding stove-piped

[19] The development of most of the UAVs used in Iraq and Afghanistan was driven not by DoD requirements, but rather by commercial research and development. Ref: "Microsoft, Google,

acquisition process is ill equipped to accommodate: it takes 91 months (7.6 years), on average, from the start of an analysis of alternatives (AoA) study to initial operational capability (IOC).[20] Even information technology programs—under whose rubric most AI-derived acquisitions naturally fall—have averaged 81 months. By way of comparison, note that within roughly this same interval of time, the commercial AI research community has gone from just *experimenting* with (prototypes of dedicated hardware-assisted) deep learning techniques,[21] to beating the world champion in Go (along with achieving many other major breakthroughs).

Of course, DoD acquisition challenges, particularly for weapons systems that include a heavy coupling between hardware and software, have been known for decades.[22] However, despite numerous attempts by various stakeholders to address these challenges, the generic acquisition process (at least on the traditional institutional level) remains effectively unchanged. Whatever progress has been made in recent years derives more from *workarounds* instituted by DoD to facilitate "rapid acquisition" of systems,[23] than from wholesale changes applied to stove-piped processes of the acquisition process itself. Some recent progress has been made—e.g., the 2009/2011 National Defense Authorization Acts (NDAA/Sec 804), mandated a new IT acquisition process, which, in turn led to multiple Defense Science Board (DSB) Task Force (TF) studies of the acquisition process. Yet, a notable absence in any of these DSB/TF studies is any explicit mention of autonomy.

Complicating the issue still further is a basic dichotomy between DoD's existing directive on autonomy (DoD Directive 3000.09, issued Nov 2012) and current Test and Evaluation (T&E) and Verification and Validation (V&V) practices. Specifically,

Facebook and more are investing in artificial intelligence: What is their plan and who are the other key players?" *TechWorld*, September 29, 2016.

[20] *Policies and Procedures for the Acquisition of Information Technology*, Department of Defense, Defense Science Board, Task Force Report, Office of the Under Secretary of Defense for Acquisition, Technology and Logistics, March 2009.

[21] The first graphics-processor-based unsupervised deep-learning techniques were introduced in 2009: R. Raina, A. Madhavan, and A. Ng, "Large-scale deep unsupervised learning using graphics processors,“ *Proceedings of the 26th Annual International Conference on Machine Learning*, ACM, 2009.

[22] J. Merritt and P. Sprey, "Negative marginal returns in weapons acquisition," *in American Defense Policy*, Third Edition, edited by R. Head and E. Roppe, John Hopkins Univ. Press, 1973.

[23] Examples include: the U.S. Air Force Rapid Capabilities Office, the U.S. Army's Asymmetric Warfare Group and Rapid Capabilities Office, DoD's Strategic Capabilities Office, and, most recently, SecDef Ashton Carter's Defense Innovation Unit Experimental (DIUx). Ref: B. Fitzgerald, A. Sander, J. Parziale, *Future Foundry: A New Strategic Approach to Military-Technical Advantage*, Center for a New American Security, 2016.

Directive 3000.09 requires that weapons systems (italics added by author of this report):[24]

- Go through rigorous hardware and software T&E/V&V, "including analysis of *unanticipated emergent behavior* resulting from the effects of complex operational environments on autonomous or semiautonomous systems."
- "Function as anticipated in realistic operational environments against *adaptive adversaries.*"
- "Are sufficiently robust to minimize failures that could lead to *unintended engagements.*"

Directive 3000.09 further requires that T&E/V&V must "assess system performance, capability, reliability, effectiveness, and suitability under realistic conditions, including possible adversary actions, consistent with the *potential consequences of an unintended engagement or loss of control of the system.*"

Yet, existing T&E/V&V practices do not make accommodations for any of the italicized parts of these quoted requirements. Among the many reasons why autonomous systems are particularly difficult to test and validate are: (1) *complexity of the state-space* (it is impossible to conduct an exhaustive search of the vast space of possible system "states" for autonomous systems); (2) *complexity of the physical environment* (the behavior of an autonomous system cannot be specified—much less tested and certified—in situ, but must be tested in concert with interaction with a dynamic environment, rendering the space of system inputs/outputs and environmental variables combinatorically intractable); (3) *unpredictability* (to the extent that autonomous systems are inherently complex adaptive systems, novel or unexpected behavior can be expected to arise naturally and unpredictably in certain dynamic situations; existing T&E/V&V practices do not have the requisite fidelity to deal with emergent behavior); and (4) *human operator trust in the machine* (existing T&E/VV&A practice is limited to testing systems in closed, scripted environments, since "trust" is not an innate trait of a system).

Trust also entails grappling with the issue of *experience* and/or *learning:* to be more effective, autonomous systems may be endowed with the ability to accrue information and learn from experience. But such a capability cannot be certified monolithically, during one "check the box" period of time. Rather, it requires periodic retesting and recertification, the periodicity of which is necessarily a function of the system's history and mission experience. Existing T&E/V&V practices are wholly inadequate to address these issues.

[24] Enclosures 2 and 3 of DoD Directive 3000.09 (*Autonomy in Weapon Systems*, Nov 2012) address T&E and V&V issues, and generally review guidelines, respectively.

Defining autonomy

"Autonomy" applies to a vastly greater range of processes than those that pertain to unmanned vehicles—as physical entities—alone, including the myriad factors needed to describe human-machine interactions. It represents a range of *context-dependent capabilities* that may appear at different scales, and in varying degrees of sophistication, that collectively enable the coupled human-machine system to perform specific tasks. Autonomy—by itself—does not reductively "fix" any existing problems; rather, it redefines, extends, and potentially opens up entirely new mission spaces. And its value can only be assessed in the context of specific mission requirements, the operating environment, and its coupling with human operators.

A major impediment to the development of autonomous weapon systems is the current lack of a common language by which AI, robot, and other technology experts, systems developers, and program managers can communicate (in a manner consistent with autonomy's multi-dimensional, context-dependent nature). There is not an even a single definition of the *word* "autonomy," much less a universally agreed upon taxonomy that might be used as basis for forming a common language. Some taxonomies emphasize the details related to a system's output functions (i.e., to its decision capability), while others focus on making detailed distinctions between input functions, such as how a system acquires information and how it formulates options. And, while sliding scales have been used to delineate between levels of "human control" that a given system might require (e.g., the "autonomy" of a system may be ranked from, say, 0, meaning that it is under complete control, to 10, meaning it is fully autonomous, albeit, typically, without the term "fully" being well defined), the practical utility of these kinds of taxonomies is limited because they ignore critically important contextual factors. For this reason, a recent U.S. Defense Science Board report recommended doing away with defining levels of autonomy altogether and replacing such taxonomies with a comprehensive conceptual framework. However, to date, despite a handful of ongoing attempts, no useable framework yet exists.

Ethical concerns

The emerging use of autonomous weapons—and the spectre (if not yet the reality) of *lethal* autonomous weapon systems (LAWS), that can select and engage targets on their own[25]—raises a host of ethical and moral questions. For example, "Will soldiers

[25] Although there are a number of weapon systems in use today that depend on varying degrees of human supervision, there are none that are fully autonomous (with the only possible exception being the Israel Defense Forces *Harpy*, a "fire-and-forget" loitering munition

be willing to go to battle alongside robots?" "Will robots be able to distinguish between military and civilian targets, and be able to use force proportionately?" "Will an AI be able to recognize enemy signs of surrender?" "Who will be responsible for an unjustified robotic kill?" and "How does one codify an innately subjective body of ethical standards and practices?"

Such questions have led to several international movements against "killer robots."[26] For example, in July 2015, over 1,000 robotics and artificial intelligence researchers signed an open letter calling for a ban on offensive autonomous weapons (with 20K+ signatories as of Dec 2016).[27] And, at the most recent United Nations Convention on Conventional Weapons, the 123 participating nations voted to convene a group of government experts to meet (during two sessions) in 2017 to formally address the LAWS issue, which could potentially lead to an international ban.[28]

While the outcome of these upcoming meetings is uncertain, it is clear is that the political, cultural, and basic human-rights dimensions of this issue are only beginning to be explored. An analysis of the *operational* impact that any limitations on (or an outright ban of) the use of offensive autonomous weapons may entail for U.S. military forces obviously deserves attention.

Transitioning to new autonomy-enabled mission areas

Figure ES-1 illustrates, schematically, the key steps involved in extending the existing unmanned systems mission space (e.g., reconnaissance, route clearance, and search and rescue) to one that more fully embraces all that autonomy potentially offers (e.g., self-organized, and self-healing, adaptive swarms). Leaving aside details of the pipeline to the main text, the key (mutually entwined) steps include, starting from bottom of the figure and working our way to the top:

- *Step 1:* Conducting basic AI research across multiple domains (the green-to-red overlay emphasizing that research in different AI areas—e.g., deep learning,

designed to detect, attack and destroy radars). Autonomy policy for U.S. weapon systems is spelled out in DoD Directive 3000.09, which expressly prohibits use of lethal *fully* autonomous weapons, which it defines as weapon systems that, once activated, may select and engage targets without further intervention by a human. Ref: DoD Directive 3000.09, "Autonomy in Weapon Systems," Nov 2012: http://www.dtic.mil/whs/directives/corres/pdf/ 300009p.pdf.

[26] M. Wareham and S. Goose, "The Growing International Movement Against Killer Robots," *Harvard International Review*, 5 Jan 2017.

[27] http://futureoflife.org/open-letter-autonomous-weapons/.

[28] Final Document of the Fifth Review Conference, CCW, Dec 2016: http://www.reaching critical will.org/disarmament-fora/ccw/2016/revcon.

image recognition, and robotic swarms—necessarily proceeds at different rates and exists, at any one time, at different levels of maturation).

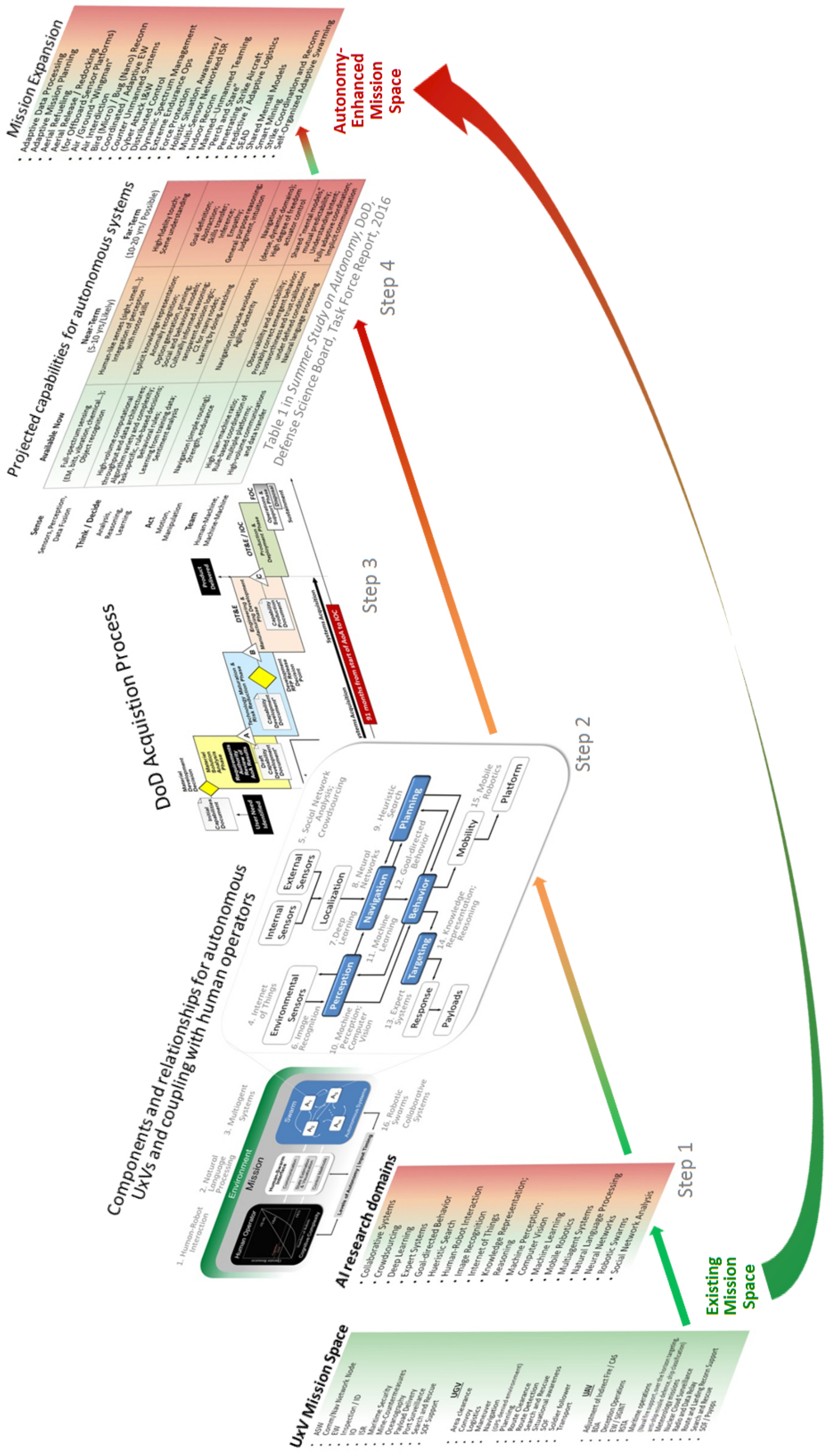
- *Step 2:* Understanding how individual AI research domains feed into the myriad components that make up autonomous systems, including their coupling with human operators (which further involves the understanding of how human-machine collaborative systems function in specific mission environments).

- *Step 3:* Moving design, development, testing, and accreditation through the DoD acquisition process (and accommodating autonomy's unique set of technical challenges while doing so).

- *Step 4:* Interpreting and projecting the requisite levels of maturity of system capabilities that autonomous systems must possess for specific missions. The autonomous systems that are shown in figure ES-1 are characterized as functions of four broad categories of AI (i.e., *sensing*, *thinking*, *acting*, and *teaming*). Their projected capabilities are indicated as follows: shades of green indicate capabilities that are available now; shades of orange denote near-term capabilities; and increasingly darker shades of red indicate the far-term regime. This table is taken from the DoD's Defense Science Board's most recent study on autonomy,[29] but is intended mostly as a notional place-holder for the kinds of conceptual, technical, and analytical considerations that must be taken into account as the raw capabilities of the autonomous systems that come out of the acquisition process are transformed into new and operationally meaningful missions and missions areas.

[29] Table 1 in *Summer Study on Autonomy*, Department of Defense, Defense Science Board, Task Force Report, Office of the Under Secretary of Defense for Acquisition, Technology and Logistics, June 2016: https://www.hsdl.org/?view&did=79464.

Figure ES-1. Key steps in transitioning to new autonomy-enabled mission areas


UxV Mission Space
ASW
Comm/Nav Network Node
EW
Inspection / ID
IO
ISR
Maritime Security
Mine-Countermeasures
Oceanography
Payload Delivery
Port Surveillance
Search and Rescue
SOF Support
UGV
Area clearance
Convoy
Logistics
Maneuver
Navigation
(GPS denied environment)
Planning
Route Clearance
Route Detection
Search and Rescue
Situational awareness
SOF
Soldier follower
Transport
UAV
Adjustment of Indirect Fire / CAS
BDA
Deception Operations
EW / SIGINT
RSTA
Maritime operations
(Naval fire support, over the horizon targeting, anti-ship missile defence, ship classification)
Meteorology missions
Nuclear Cloud Surveillance
Radio and Data Relay
Route and Landing Reconn Support
Search and Rescue
SOF / Psyops
Existing Mission Space
Step 1
AI research domains
Collaborative Systems
Crowdsourcing
Deep Learning
Expert Systems
Goal-directed Behavior
Hueristic Search
Human-Robot Interaction
Image Recognition
Internet of Things
Knowledge Representation; Reasoning
Machine Perception; Computer Vision
Machine Learning
Mobile Robotics
Multiagent Systems
Natural Language Processing
Neural Networks
Robotic Swarms
Social Network Analysis
Components and relationships for autonomous UxVs and coupling with human operators
1. Human-Robot Interaction
2. Natural Language Processing
3. Multiagent Systems
Human Operator
Environment
Mission
Swarm
Autonomous Systems
Levels of Autonomy / Input Timing
16. Robotic Swarms Collaborative Systems
4. Internet of Things
Environmental Sensors
6. Image Recognition
10. Machine Perception; Computer Vision
Perception
Internal Sensors
External Sensors
5. Social Network Analysis; Crowdsourcing
Localization
7. Deep Learning
8. Neural Networks
Navigation
11. Machine Learning
9. Heuristic Search
Planning
12. Goal-directed Behavior
Behavior
14. Knowledge Representation; Reasoning
Targeting
13. Expert Systems
Response
Payloads
Mobility
15. Mobile Robotics
Platform
Step 2
DoD Acquistion Process
User Need Identified
Initial Capabilities Document
Material Development Decision
Material Solution Analysis Phase
Requirements Authority Review of AoA Results
Draft Capability Development Document
A
Technology Maturation & Risk Reduction Phase
Capability Development Document
Development RFP Release Decision Point
B
Engineering & Manufacturing Development Phase
Capability Production Document
DT&E
C
Production & Deployment Phase
OT&E / IOC
Product Delivered
FOC
Operations & Support Phase
Disposal
Sustainment
Systems Acquisition
91 months from start of AoA to IOC
Step 3
Projected capabilities for autonomous systems
Table 1 in Summer Study on Autonomy, DoD, Defense Science Board, Task Force Report, 2016
Step 4
Mission Expansion
Adaptive Data Processing
Adaptive Mission Planning
Aerial Refueling
Aerial Release / Redocking (for Offboard Sensor Platforms)
Air /Ground "Wingman"
Air Interdiction
Bird (Micro) / Bug (Nano) Reconn
Coordinated / Adaptive EW
Counter Unmanned Systems
Cyber Attack I&W
Distributed Control
Dynamic Spectrum Management
Extreme Endurance Ops
Force Protection
Holistic Situation Awareness / Multi-Sensor Networked ISR
Indoor Reconn
Manned–Unmanned Teaming
"Perch and Stare"
Penetrating Strike Aircraft
Predictive / Adaptive Logistics
SEAD
Shared Mental Models
Smart Mining
Strike Coordination and Reconn
Self-Organized Adaptive Swarming
Autonomy-Enhanced Mission Space

	Available Now	Near-Term (5-10 yrs/Likely)	Far-Term (10-20 yrs/ Possible)
Sense Sensors, Perception, Data Fusion	Full-spectrum sensing (EM, bits, vibration, chemical...); Object recognition	Human-like senses (sight, smell...); Integration of perception with motor skills	High-fidelity touch; Scene understanding

Think / Decide
Analysis, Reasoning, Learning
	High-volume computational throughput and data architectures;
	Algorithm variety and complexity;
	Task-specific, rule-based decisions;
	Behavioral rules;
	Learning from training data;
	Sentiment analysis
		Explicit knowledge representation;
		Anomaly recognition;
		Option generation, pruning;
		Social and behavioral models;
		Culturally informed reasoning;
		ransparent decision logic;
		C2 for many nodes;
		Learning by doing, watching
			Goal definition;
			Abstraction;
			Skills transfer;
			Inference;
			Empathy;
			General purpose reasoning;
			Judgment, intuition

Act Motion, Manipulation	Navigation (simple routing); Strength, endurance	Navigation (obstacle avoidance); Agility, dexterity	Navigation (dense, dynamic domains); High degree of freedom actuator control

Team
Human-Machine, Machine-Machine
	High man-machine ratio;
	Rule-based coordination of multiple platforms;
	High-volume communications and data transfer
		Observability and directability;
		Provably correct emergent behavior;
		Trustworthiness and trust calibration under defined conditions;
		Natural language processing
			Shared "mental models";
			mutual predictability;
			Understanding intent;
			Fully adaptive coordination;
			Implicit communication

Gestalt of main findings

The military is on the cusp of a major technological revolution as it enters the *Robotic Age,*[30] in which warfare is conducted by unmanned and increasingly autonomous weapon systems, operating across all domains (air, sea, undersea, land, space, and cyber), and across the full spectrum of military operations. The question is not *whether* the future of warfare will be filled with autonomous, AI-driven robots, but *when* and in what *form*. However, unlike the last "sea change" during the Cold War (i.e., the so-called "2nd Offset"),[31] when advanced technologies such as precision-strike weapons, stealth aircraft, smart weapons and sensors, and GPS were developed primarily by DoD-sponsored research and development programs, a successful transition into the Robotic Age (spurred on by DoD's recent "Third Offset Strategy" innovation initiative)[32] will depend critically on how well DoD is able to embrace technologies and innovations that are now being developed mostly in the commercial world. And, while the human warfighter is not going away anytime soon, if ever (even as the depth and breadth of autonomy steadily expand), human operators will not suddenly lose control of existing unmanned systems. A telltale sign that DoD has made a "no looking back" cross-over into the Robotic Age will be when human operators can no longer fully understand, or *predict*, how autonomous systems behave—i.e., when, for the first time, a human operator is as stunned by some weapon system's action as 18-time world Go champion Lee SeDol was by a single move of the AI that defeated him.

In preparation for DoD's cross-over into the Robotic Age, whenever it arrives, this study has identified four key technical gaps in developing AI-based autonomous systems, wherein opportunities for future analytical studies naturally arise (see figure ES-2).

These gaps are:

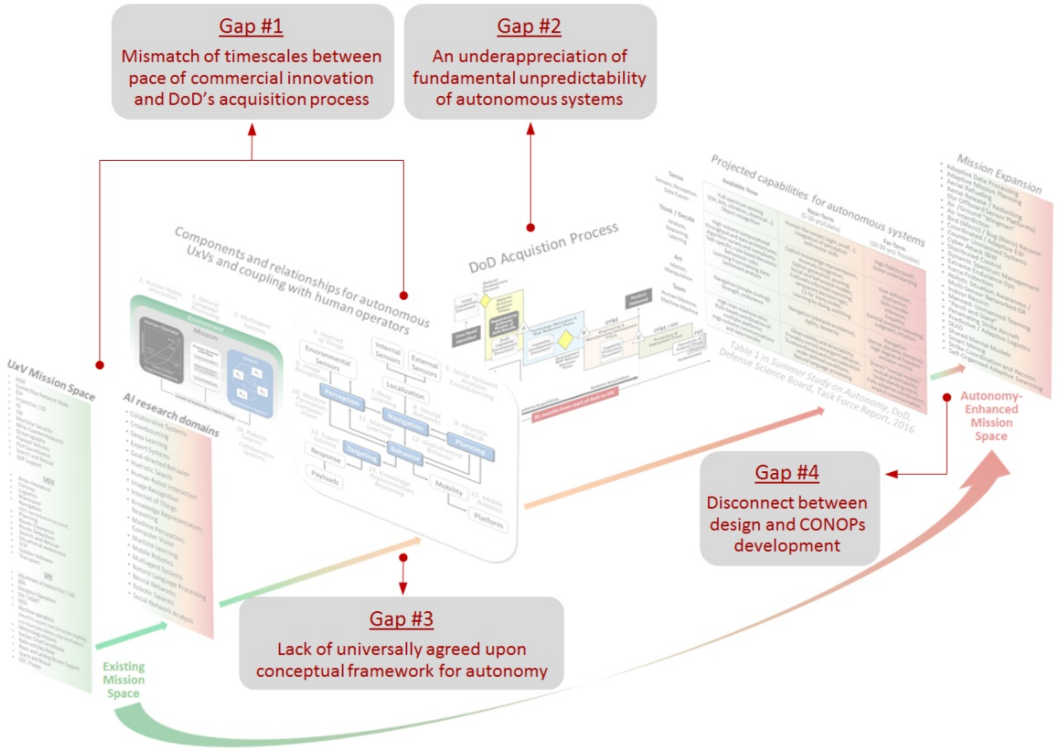
- *Gap 1*: A fundamental mismatch—even *dissonance*—between the accelerating pace (and manner of development and evolution) of technology innovation in commercial and academic research communities, and the timescales and assumptions underlying DoD's existing acquisition process.

[30] Robert O. Work and Shawn Brimley, *20YY: Preparing for War in the Robotic Age*, Center for a New American Security, Jan 2014.

[31] J. McGrath, "Twenty-First Century Information Warfare and the Third Offset Strategy," *Joint Forces Quarterly*, National Defense University, Issue 82, 3rd Quarter 2016.

[32] C. Hagel, Transcript of Keynote speech delivered at Reagan National Defense Forum Keynote, Ronald Reagan Presidential Library, Simi Valley, CA, Nov. 15, 2014.

Figure ES-2. Key *gaps* in transitioning to new autonomy-enabled mission areas


Gap #1
Mismatch of timescales between pace of commercial innovation and DoD's acquisition process
Gap #2
An underappreciation of fundamental unpredictability of autonomous systems
Components and relationships for autonomous UxVs and coupling with human operators
DoD Acquistion Process
Projected capabilities for autonomous systems
Mission Expansion
UxV Mission Space
AI research domains
Table 1 in Summer Study on Autonomy, DoD Defense Science Board, Task Force Report, 2016
Autonomy-Enhanced Mission Space
Gap #4
Disconnect between design and CONOPs development
Gap #3
Lack of universally agreed upon conceptual framework for autonomy
Existing Mission Space

- *Gap 2*: An underappreciation of the unpredictable nature of autonomous systems, particularly when operating in dynamic environment, and in concert with other autonomous systems. Existing T&E/V&V practices accommodate neither the basic properties of autonomous systems, as expected by AI and indicated by decades of deep fundamental research into the behavior of complex adaptive systems, nor the requirements they must meet, as weapon systems (as spelled out by DoD Directive 3000.09).

- *Gap 3*: A lack of a universally agreed upon conceptual framework for autonomy that can be used both to anchor theoretical discussions and to serve as a frame-of-reference for understanding how theory, design, implementation, testing, and operations are all interrelated. A similar deficiency exists for understanding the role that trust plays in shaping a human operator's interaction with an autonomous system. The Defense Science Board's most recent study on autonomy[33] warns that "inappropriate calibration" of trust during "design, development, or operations will lead to misapplication" of autonomous systems, but offers only a tepid definition of trust, and little guidance on how to apply it.

- *Gap 4*: DoD's current acquisition process does not allow for a timely introduction of "mission-ready" AI/autonomy, and there is a general disconnect between system design and the development of concepts of operations (CONOPS). Unmanned systems are typically integrated into operations from a *manned*-centric CONOPS point of view, which is unnecessarily self-limiting by implicitly respecting human performance constraints.

Recommended studies

While not even AI experts can predict how AI will evolve in even the near–term future (much less project its possible course over 10 or more years,[34] or predict AI's impact on the development of military autonomous systems), it is still possible to anticipate many of the key conceptual, technical, and operational challenges that DoD will face in the coming years as it increasingly turns to and more deeply embraces AI-based technologies, and fully enters the "Robotic Age." From an operational analysis standpoint, these challenges can also be used to help shape future studies:

Recommendation 1: ***Help establish dialog between commercial research and development and DoD.***

Institutions specializing in operational analysis are well suited to act as "go betweens" linking the academic and commercial research communities with military culture / operational needs. Assuming that Secretary of Defense

[33] *Summer Study on Autonomy*, Department of Defense, Defense Science Board, Task Force Report, Office of the Under Secretary of Defense for Acquisition, Technology and Logistics, June 2016: https://www.hsdl.org/?view&did=79464.

[34] S. Armstrong, K. Sotala, and S. hÉigeartaigh, "The errors, insights and lessons of famous AI predictions – and what they mean for the future," *Journal of Experimental & Theoretical Artificial Intelligence* 26, no. 3, 2014; D. Fagella, "Artificial Intelligence Risk – What Researchers Think is Worth Worrying About," *Tech Emergence,* 20 March 2016: http://techemergence.com/artificial-intelligence-risk/. For the most recent survey of expert opinion see: V. Muller and N. Bostrom, "Future Progress in Artificial Intelligence: A Survey of Expert Opinion," in *Fundamental Issues of Artificial Intelligence*, edited by V. Muller, Springer-Verlag, 2016.

Ashton Carter's Defense Innovation Unit-Experimental (DIUx) program survives into the next administration,[35] operationally informed and technically knowledgeable analysts can help stakeholders better "understand" each other. Cross-fertilization with the Naval Postgraduate School (NPS) may also pay dividends.[36]

Recommendation 2: ***Develop an operationally meaningful conceptual framework for autonomy.***

For example, build on lessons learned from the National Institute of Standards and Technology's (NIST's) stalled evolution of its ALFUS (Autonomy Levels for Unmanned Systems) framework, and develop the skeleton of an idea proposed by DoD's Defense Science Board's 2012 report on autonomy.[37]

Recommendation 3: ***Develop measures of effectiveness (MOEs) and measures of performance (MoP) for autonomous systems.***

Develop a methodology by which the effectiveness of autonomous systems can be measured at all levels (e.g., developers, program managers, decision-makers, and warfighters) and across all required functions, missions, and tasks (e.g., coordination, mission tasking, training, survivability, situation awareness, and workload).

Recommendation 4: ***Use nontraditional modeling and simulation (M&S) techniques to help mitigate AI/autonomy-related dimensions of uncertainty.***

As DoD moves into the Robotic Age, M&S is moving away from "simulations as distillations" of real systems (for which M&S has traditionally been used to develop models in order to gain insights into the *real* system), to "simulation-based rules and algorithms as descriptions" of real (i.e., engineered)

[35] DIUx has been established to help facilitate the discovery and development of capabilities and technologies outside DoD's normal acquisition pipeline. Ref: https://www.diux.mil/.

[36] For example: NPS's Consortium for Robotics and Unmanned Systems Education and Research (CRUSER: https://my.nps.edu/web/cruser), and *Autonomous Systems Track* (http://my.nps.edu/web/ast).

[37] *The Role of Autonomy in DoD Systems*, DoD Defense Science Board, Task Force Report, Office of the Under Secretary of Defense for Acquisition, Technology and Logistics, July 2012.

robots and behaviors. It is here, at the cusp between exploring behaviors and prescribing rules that generate them (e.g., engineering *desired* swarm behaviors), that M&S can help mitigate some of the challenges and uncertainties of developing autonomous systems and robotic swarms. For example, while "swarm engineering" methods exist to facilitate the unique design requirements of robotic swarms, no general method exists that maps individual rules to (desired) group behavior.[38]

Multi-agent based modeling techniques[39] are particularly well suited for developing these rules, and, more generally, for studying the kinds of self-organized emergent behaviors expected to arise in coupled autonomous systems (e.g., "How sensitive is an autonomous system's behavior to changes in its physical environment?", "What new command and control architectures will be needed for robotic swarms?", and "How will the control and behavior of a swarm scale with its size and mission complexity?").

Recommendation 5: ***Apply wargaming techniques to help develop new CONOPS.***

Wargaming can be used to help identify and develop new CONOPS, apply lessons-learned from the experience of using deployed systems, explore options to counter uses of autonomy by potential adversaries, and assist in training (e.g., by exploring trust issues in human-machine collaboration). Wargames can also stimulate and nurture a more unified approach to understanding autonomous system performance and behavior, provided that they are conducted with the support and participation from across all military services and domains.

[38] I. Navarro and F. Matia, "An Introduction to Swarm Robotics," *International Scholarly Research Notes*, Vol. 2013, 2013: https://www.hindawi.com/ journals/isrn/2013/608164/.

[39] A. Ilachinski, *Artificial War: Multiagent-Based Simulation of Combat*, World Scientific, 2004. See also: A. Ilachinski, "Modelling insurgent and terrorist networks as self-organized complex adaptive systems," *International Journal of Parallel, Emergent and Distributed Systems* 27, 2012; A. Ilachinski, *AOEWSim: An Agent Based Model for Simulation Interactions Between Off-Board EW Systems and Anti-Ship Missiles*, CNA, DWP-2013-U-004757, 2013; A. Ilachinski and M. Shepko, *FAC/FIAC Simulation (FFSim): User's Guide*, CNA, Annotated Briefing, 2015.

Recommendation 6: ***Develop new T&E/V&V standards and practices appropriate for the unique challenges of accrediting autonomous systems.***

For example, help ameliorate basic gaps in testing in terms of accommodating complexity, uncertainty, and subjective decision environments, by appealing to and exploiting lessons learned from the development and accreditation practices established by the complex system theory and multiagent-based modeling research communities.

Recommendation 7: ***Explore basic human-machine collaboration and interaction issues.***

As autonomy increases, human operators will be concerned less with the manual control of a vehicle, and more with controlling swarms and directing the overall mission: "What are the operator's informational needs (and workload limitations) for controlling multiple autonomous vehicles?" "How do humans keep pace with an accelerating pace of autonomy-driven operations?" "What kinds of command-and-control relationships are best for human-machine collaboration?" "How are human and autonomous-system decision-making practices optimally integrated?" and "What data practices are key to developing shared situation awareness?"

Recommendation 8: ***Explore the challenges of force-integration of increasingly autonomous systems.***

Essentially all force-integration issues are, as yet, undetermined. They must consider not just "low hanging fruit" extensions of existing CONOPS, in which the human component is simply replaced with unmanned systems and "operational value" of human performance is scaled to accommodate "better" performance (e.g., endurance, survivability), but brainstorm heretofore nonexistent tactics, operations, and missions that fully embrace existing and anticipated future autonomous capabilities. What is the tradeoff between large numbers of simple, low-cost (i.e., "disposable") vehicles and small numbers of complex (multi-functional) ones?

The operationalization of robotic swarms, in particular, represents a heretofore largely untapped dimension of the mission space, and will require the development of new CONOPS. The swarm may be used as a radically new form of precision coordinated "en masse" guided munition; as a self-healing area surveillance network (which includes collecting and assimilating data on an adversary's Internet-of-Things (IoT);[40] or as an adaptive distributed electronic jammer.

Recommendation 9: ***Explore the cyber implications of autonomous systems.***

Explore what new features increased AI-driven autonomy brings to the general risk assessment of increasingly autonomous unmanned systems. On one hand, autonomy may potentially reduce a force's overall vulnerability to jamming or cyber hacking. For example, communications loss over a jammed data link may be compensated for by the ability of autonomous vehicles to continue performing their mission). On the other hand, autonomy itself may also be *more*, not less, vulnerable to a cyber intrusion. For example, an adversary may gain "control," or otherwise deliberately "perturb" the behavior of an autonomous system; it may also be more difficult to detect embedded malware. In the latter context, consider some future variants of incidents such as the Iranian capture of an RQ-170 *Sentinel* in 2011,[41] and the "keylogging" virus that infected the UAV-control-computers at the Creech Air Force Base in Nevada.[42]

Recommendation 10: ***Explore operational implications of ethical concerns over the use of lethal autonomous weapon.***

Analyze issues of accountability, legality, and liability in arguments put forth by various "Ban LAWS" movements. Examine the possible constraints on missions (along with other associated impediments to the design and development of autonomous systems) that may result from an international ban (or set of limits) imposed on the development or deployment of LAWS, such as might come out of the United-Nations-sponsored government experts' negotiations scheduled to take place sometime in 2017.

[40] G. Seffers, "Defense Department Awakens to Internet of Things," *Signal*, 1 Jan 2015: http://www.afcea.org/content/?q=defense-department-awakens-internet-things.

[41] The Iranian government announced that the RQ-170 was captured by its cyber warfare unit: "Iran shows film of captured US drone," BBC News, 8 Dec 2011: http://www.bbc.com/news/world-middle-east-16098562.

[42] N. Shachtman, "Exclusive: Computer virus hits U.S. drone fleet," *Wired*, 7 Oct 2011.