Designing Robots for the Battlefield: State of the Art

Bruce A. Swett, Erin N. Hahn, and Ashley J. Llorens

Contents

Introduction	132
AI-Robots Currently Used on the Battlefield	132
Unmanned Aerial Vehicles.....	132
Unmanned Ground, Surface, and Underwater Vehicles.....	132
Integrated Air Defense Systems and Smart Weapons.....	133
Understanding AI and how It Enables Machine Autonomy	134
Elements of the PDATT Framework.....	134
The Risks of AI-Robots on the Battlefield	136
Current Limitations of AI-Robots on the Battlefield and Associated Risks.....	136
The Future of AI-Robots on the Battlefield	137
What AI-Robots May Be Able to Do in the Near Future.....	137
JHU/APL Research Toward AI-Robots for Trusted, Real-World Operations	139
Areas of AI-Robotics Research at JHU/APL.....	139
Conclusions: The Future of AI for Battlefield Robotics	141
Competing Visions of the Future of AI-Robots.....	141
References	143

Abstract

There is currently a global arms race for the development of artificial intelligence (AI) and unmanned robotic systems that are empowered by AI (AI-robots). This paper examines the current use of AI-robots on the battlefield and offers a framework for understanding AI and AI-robots. It examines the limitations and risks of AI-robots on the battlefield and posits the future direction of battlefield AI-robots. It then presents research performed at the Johns Hopkins University Applied Physics Laboratory (JHU/APL) related to the development, testing,

and control of AI-robots, as well as JHU/APL work on human trust of autonomy and developing self-regulating and ethical robotic systems. Finally, it examines multiple possible future paths for the relationship between humans and AI-robots.

Keywords

Artificial intelligence · Robotics · Unmanned vehicles · AI ethics · AI-robots · Battlefield

B. A. Swett (✉) · E. N. Hahn · A. J. Llorens
The Johns Hopkins University Applied Physics Laboratory
(JHU/APL), Laurel, MD, USA
e-mail: bruceswett@verizon.net; erin.hahn@jhuapl.edu;
Ashley.Llorens@jhuapl.edu

Introduction

There is currently an arms race for the development of artificial intelligence (AI) and unmanned robotic platforms enhanced with AI (AI-robots) that are capable of autonomous action (without continuous human control) on the battlefield. The goal of developing AI-empowered battlefield robotic systems is to create an advantage over the adversary in terms of situational awareness, speed of decision-making, survivability of friendly forces, and destruction of adversary forces. Having unmanned robotic platforms bear the brunt of losses during combat—reducing loss of human life—while increasing the operational effectiveness of combat units composed of manned-unmanned teams is an attractive proposition for military leaders.

AI-Robots Currently Used on the Battlefield

Numerous types of fielded robotic systems can perform limited operations on their own, but they are controlled remotely by a human pilot or operator. The key element of these systems is that no human is onboard the platform like there is on a plane flown by a pilot or a tank driven by a commander. The human control component has been removed from the operational platform and relocated away from the battlefield. Real-time communication between the human operator and the unmanned robotic system is required, so that the human can control the platform in a dynamic, complex battlefield. As advanced militaries around the world contest the portions of the electromagnetic spectrum used for the control of unmanned combat systems, the need for increased robot autonomy is heightened. U.S. policy directs that, regardless of the level of autonomy of unmanned systems, target selection will be made by an “authorized human operator” (U.S. Dept. of Defense 2012). The next sections describe the primary categories of deployed robotic systems.

Unmanned Aerial Vehicles

Unmanned aerial vehicles (UAVs) are widely used by the U.S., Chinese, and Russian militaries, as well as by numerous nation states and regional actors (Dillow 2016). The relatively low cost and low technology barrier to entry for short-range UAVs has enabled the rapid proliferation of UAV technologies (Nacouzi et al. 2018) for use individually, in groups, or in large swarms. Long-range UAVs with satellite-based communications are more likely to be used by countries with extensive military capabilities. UAVs were initially used for surveillance to provide an enhanced understanding of the area of conflict, but their role has expanded into target identification and tracking; direct attack; command, control, and

communications; targeting for other weapons systems; and resupply (see Fig. 1). UAVs are divided into Class I (small drones), Class II (tactical), and Class III (strategic) (Brown 2016), or Groups 1–5, depending on size and capabilities (U.S. Army Aviation Center of Excellence 2010).

In actual combat, Russia flew multiple types of UAVs in the conflict with the Ukraine, using UAVs to rapidly target Ukrainian military forces for massed-effect artillery fire (Fox 2017; Freedberg Jr. 2015), and has employed armed UAVs in Syria (Karnozov 2017). The U.S. has an extensive history of using UAVs in combat, with the MQ-9 Reaper targeted strikes in the Middle East being the most notable. Israel has developed the IAI Heron to compete with the Reaper and, along with China, is a major supplier of drone technology (Blum 2018). Armed Chinese CH-4 drones are in operation over the Middle East by countries that are restricted from purchasing U.S. drone technologies (Gambrell and Shih 2018), and China has become a leading developer and distributor of armed UAVs (Wolf 2018). Current UAVs are usually flown by a pilot at a ground station. There are AI algorithms that can set waypoints—a path for the UAV to fly—but the UAV doesn’t make decisions on where to go or to fire weapons. These functions are currently the pilot’s responsibility. There are indications that major militaries, including the U.S. Army, are moving toward taking the human out of the loop by having UAVs perform behaviors autonomously (Scharre 2018; Singer 2001).

Unmanned Ground, Surface, and Underwater Vehicles

The U.S. Army has successfully used small, unarmed, portable unmanned ground vehicles (UGVs) for improvised explosive device (IED) removal and explosive ordnance disposal in Iraq and Afghanistan, and it is purchasing UGVs in large numbers (O’Brien 2018). Recently, the focus has shifted to using both small, armed UGVs and larger, armed robotic tanks. Russia deployed the Uran-9 robotic tank, which is armed with a gun and missiles, in Syria with mixed results, and the U.S. Army is developing an optionally manned replacement for the Bradley fighting vehicle (Roblin 2019). The U.S. Army is also developing robotic tanks and technology kits to convert existing tanks into remotely controlled and/or autonomous UGVs (Osborn 2018). Similar to UAVs, unmanned robotic tanks serving as scouts for manned tanks would reduce the loss of life while increasing the overall firepower that could be brought to bear on the adversary (see Fig. 2, left).

Maritime unmanned platforms have been developed and tested by several countries, including the U.S. and China. China demonstrated control of a swarm of 56 unarmed un-

Fig. 1 *Military uses of UAVs.* UAVs are now a ubiquitous presence on the battlefield and are being used for intelligence, surveillance, and reconnaissance; targeting; communications; and direct attack purposes in militaries around the world (images courtesy of the Marine Corps Warfighting Lab—quadcopter; and U.S. Air Force—Reaper)

Fig. 2 *Military uses of unmanned ground, surface, and underwater vehicles.* UGVs have been used for IED and ordnance removal and are now being developed as mobile weapons systems and semiautonomous

tanks (above, left, from BigStockPhoto, Mikhail Leonov). USV and UUV swarms (above, right, JHU APL) are able to coordinate swarm behavior to inflict significant damage on a larger vessel

manned surface vehicles (USVs) in 2018 (Atherton 2018). The importance of swarm control of USVs or unmanned underwater vehicles (UUVs) lies in the ability to overwhelm the defenses of much larger ships, to inflict disproportionate damage (see Fig. 2 right). The U.S. is developing UUVs that can perform anti-submarine warfare and mine countermeasures autonomously (Keller 2019b; Keller 2019a). As with UAVs and UGVs, USVs and UUVs require communications with human operators or integrated control systems, and that will remain the case until AI autonomy algorithms are sufficiently robust and reliable in performing missions in communications-denied environments.

Integrated Air Defense Systems and Smart Weapons

Russia has developed and fielded a highly effective integrated air defense system (IADS) that can engage and destroy

multiple aircraft at range, creating regions where no one can fly (i.e., anti-access/area denial, or A2AD) (see Fig. 3). The Russian IADS combines radar tracking; air defense missiles for short, medium, and long range; electronic warfare (EW) jamming capabilities; guns; surface-to-air missiles; robust communications; and an integration control system (Jarretson 2018). Russia has recently made improvements to the range of the S-500 IADS and has sold the existing S-400 IADS to China for use in the South China Sea (Trevithick 2018). Importantly, because of the speed of incoming aircraft and missiles, these IADS are fully autonomous weapons systems with no identification of friend or foe. In this case, “autonomous” means that the system determines what are targets, which targets to engage, and the timing of engagement and then enacts the targeting decisions. While in this autonomous mode, the autonomous system’s decisions and actions are too fast for human operators to identify and intervene. Similarly, the U.S. Navy has used the MK 15 Phalanx Close-In Weapon System on ship to counter high-

Fig. 3 *Russian IADS.* IADS developed by Russia, which include radars, long- and short-range missiles, EW systems, and machine guns, create areas where no aerial platform can operate (from BigStockPhoto, Komisar)

speed missiles and aircraft without human commands (Etzioni and Etzioni 2017). The argument for this human-on-the-loop approach is that a human operator is not fast enough to respond to the threat in time, so an algorithm has to be used.

Understanding AI and how It Enables Machine Autonomy

In this section we provide a framework for understanding how AI enables robots to act autonomously to pursue human-defined goals. This framework is offered as a basis for understanding the current state of the art in AI as well as the limitations of AI, which are described in the next section. AI, as a field of study, may be thought of as the pursuit to create increasingly intelligent AI that interacts with the world, or AI agents (Russell and Norvig 2003). A straightforward way to understand the term intelligence in this context is as an agent's ability to accomplish complex goals (Tegmark 2018), with the agent's degree of autonomy in pursuing those goals arising from the delegation of authority to act with specific bounds (Defense Science Board 2017). Agents are typically implemented as systems, where the interactions of the system components—sensors, computers, algorithms—give rise to its intelligence.

Whether agents are embodied robotic systems or exist only as programs in cyberspace, they share a common ability to perceive their environment, decide on a course of action that best pursues their goals, and act in some way to carry out the course of action while teaming with other agents. Multi-agent teams may themselves be thought of as intelligent systems, where intelligence arises from the agents' ability to act collectively to accomplish common goals. The appropriate calibration of trust is a key enabler for effective teaming among agents performing complex tasks in challenging environments. Perceive, Decide, Act, Team, Trust (PDATT) is a conceptual framework that allows us to build intuition around the key thrusts within AI research, the current state of the art, and the limits of AI-robots (see Fig. 4).

Elements of the PDATT Framework

Perceive

Perception for an intelligent system involves sensing and understanding relevant information about the state of the operating environment. Salient perception of physical environments for an intelligent system may require sensing of physical phenomena both within and beyond the range of human senses. The perception space for an intelligent system may also lie within a purely informational domain requiring, for example, measurement of activity patterns on a computer

Fig. 4 PDATT AI framework. The PDATT framework outlines the main elements of AI-robots. The light blue boxes illustrate what AI-robots do: they must perceive salient elements of the environment to decide which action to select. An action is then taken, with expected effects in the environment. The AI-robot operates alongside humans in teams, requiring frequent and meaningful communication to ensure trust during mission operations. The red boxes illustrate how AI-robots accomplish the four foundational functions of perceive, decide, act, and team. The green box represents the set of capabilities needed to ensure that the AI-robot is trusted by the human operator in meeting mission objectives (JHU/APL)

network or counting the number of mentions of a particular public figure on Twitter. Recent progress in machine learning has enabled amazing improvements in object recognition and sequence translation, while combining this kind of data-driven pattern recognition with robust human-like reasoning remains a fundamental challenge.

Decide

Decision-making for an intelligent system includes searching, evaluating, and selecting a course of action among a vast space of possible actions toward accomplishing long-term goals. The challenge that long-term decision-making poses to an intelligent system can be illustrated using strategy games such as chess or Go. An algorithm that plays chess must determine the best sequence of moves to accomplish its ultimate goal of winning the game. Although a game like chess is less complex than real-world decision-making, it still presents a difficult decision-making challenge for an agent. In 1950, famous mathematician and engineer Claude Shannon estimated that the game size—the number of different sequences of actions in the game of chess, taking into account all the possible moves by each player—is 10^{43} (Shannon 1950). For the game of Go, the estimated game size is 10^{170} (Tromp 2016). Of course, despite their complexity, these games abstract the dynamics and uncertainty associated with real-world decision-making. While recent progress in developing artificial agents capable of superhuman-level play

in games like these is exciting, significant research will be needed for AI to be capable of long-term decision-making for complex real-world challenges (Markowitz et al. 2018).

Act

Acting for an intelligent system includes the ways in which the system interacts with its environment in accomplishing given goals. Acting may include traversing an environment, manipulating objects, or even turning computers on or off to prevent network intrusion. A system may act in order to improve its perception—for example, moving a sensor to get a better line of sight to an object of interest. An agent's action space is the set of possible actions available to it. In chess, the action space consists of allowable movement of pieces on the board. For a driverless car, the action space consists of control actions such as accelerating and steering. For an agent that controls a centrifuge in a nuclear power plant, the action space may include raising or lowering the speed or temperature of critical system components. For a humanoid robot assistant, the action space may include manipulation of household objects with human-like dexterity. Recent progress in robotics has enabled systems capable of taking complex actions within carefully controlled conditions, such as a factory setting, while enabling robots that act in unstructured real-world setting remains a key challenge.

Team

For an intelligent system, effectively pursuing complex goals in challenging environments almost always requires acting as part of a multi-agent team with both humans and other machines. Teaming typically requires some form of communication, the ability to create sufficient shared situational awareness of the operating environment, and the ability to effectively allocate tasks among agents. Recent demonstrations of multi-agent teaming have involved large choreographed formations, such as the Intel UAV demonstration at the 2019 Super Bowl (Intel 2019). In future applications, we envision teams of agents with heterogeneous capabilities that can seamlessly and flexibly collaborate to accomplish human-specified goals. Flexible multi-agent teaming in unstructured scenarios will require continued research. This level of teaming may require something akin to the development of a machine theory of mind, including the ability to model the goals, beliefs, knowledge, etc. of other agents (Premack and Woodruff 1978).

Trust

Realizing the promise of AI-robots will require appropriate calibration of trust among developers, users, and certifiers of technologies, as well as policy-makers and society at large. Here, trust is defined as "...the intention to accept vulnerability to a trustee based on positive expectations of his or her actions" (Colquitt et al. 2007). Fundamental advance-

ments are needed in trust as it relates to intelligent systems, including the following.

- *Test and evaluation*: assuring complex intelligent systems meet design goals in the face of significant uncertainty in both the test and operational environments
- *Resilience to adversarial influence*: hardening intelligent systems against subversion through phenomenological, behavioral, and informational attack by an adversary
- *Goal alignment*: ensuring that actions performed by intelligent systems remain aligned with human intent even as those systems are tasked to pursue increasingly high-level goals
- *Policy*: Determining which decisions and actions are appropriate for an intelligent system to perform autonomously in accordance with system capabilities as well as societal values and ethics.

The Risks of AI-Robots on the Battlefield

So, given an understanding of how AI-robots work using PDATT and the current use of AI-robots on the battlefield, what are the risks of accelerated use and enhancement of capabilities of AI-robots on the battlefield? Assuming that artificial general intelligence (AGI)—with human-level perception, understanding of context, and abstract reasoning—won't be available in the next decade, what are the limitations of current AI-robots that induce risk on the battlefield?

Current Limitations of AI-Robots on the Battlefield and Associated Risks

Perception

Computer vision algorithms have made significant gains in the identification of people and objects in video data, with classification rates matching or exceeding human performance in some studies (Dodge and Karam 2017). However, making sense of the world in three dimensions involves more than object detection and classification. Current technical challenges in perception include tracking humans and objects consistently in varied backgrounds and lighting conditions; identifying and tracking individuals in various contexts; identifying behaviors; and predicting intent or outcomes. Actions taken over time also have meaning, as do interactions between people, between people and objects, and between different objects. The human brain has “what” pathways, to link visual identification of a person or object with background information, associations, and prior experiences, also known as “vision-for-perception.” The brain’s “where” pathway not only tracks the location of the person or object relative to

the viewer, but it also encodes “vision-for-action” (Goodale and Milner 2006). Current AI systems lack a sophisticated, multisensory perceptual system that creates rich internal representations of the state of the world, and its changes, in real time. The risks associated with limited perception include the inability to reliably detect friend or foe; relationships between elements in a scene; and the implications of the actions, people, and objects in a scene.

Decision-Making, Reasoning, and Understanding Context

Largely due to the deficits in AI perceptual systems, current AI does not have an ability to understand the context of actions, behaviors, or relationships. This complex web of experience is essential for reasoning and decision-making. The most advanced fielded AI reasoning systems are rules based—that is, they are a set of “if-then” statements that encode human processes and decisions (Kotseruba and Tsotsos 2018). While these rules-based systems have been successful in constrained tasks and limited robotic autonomy, they are statically programmed prior to the AI system operating in the battlefield, so they are necessarily limited and predictable. Current AI algorithms are “narrow,” in that they can solve a specific problem. The eventual goal is to create AGI, which will have human-level reasoning and will be both generative, able to create new behaviors on the fly in response to novel situations, and generalizable, able to apply learning in one area to other operations or missions. The development of AGI is a long-term vision, with estimates of realization between 10 years and never (Robitzski 2018). The risk of AI-robots having insufficient reasoning is compounded by the limitations in perception. Choosing appropriate actions within a specific context and mission requires highly accurate and reliable perception, as well as a wealth of background knowledge of what the context, rules, and perceptions mean for decision-making. None of these highly accurate and reliable perceptual and reasoning capabilities currently exist for fully autonomous AI-robot behavior.

Action Selection, Self-Correction, and Ethical Self-Assessment

If an unmanned platform is to perform independently of a human operator, it will need to encode and function within the legal and ethical constraints of the military operating the system. AI systems have difficulties predicting how action choices will affect the environment, so they are unable to weigh the alternatives from an ethical or legal perspective. A recent project that used AI algorithms to predict the best engineers to hire discriminated against women; another project that used AI algorithms to predict whether criminals would re-offend discriminated against Blacks (Shaw 2019). Humans are able to predict consequences of different pos-

sible actions all the time, weighing which alternative is the most advantageous while adhering to personal chosen norms. Future AI systems for battlefield robots will need to have the ability to predict the impact of actions on the environment, including friendly forces, adversaries, and noncombatants, and then make decisions informed by the Law of Armed Conflict and national policies (Dowdy et al. 2015). If AI-robots do not have continuous, feed-forward prediction capabilities, they will be unable to self-regulate, risking their ability to adhere to mission parameters in the absence of direct human oversight.

Teaming and Trust: Transparency in Human Interactions

A significant concern with today's state-of-the-art AI algorithms is that some neural networks used for battlefield robots are not human readable, or are "black boxes" (Knight 2017). So, while the algorithm that detects the presence of a tank may do so with a certain probability of detection and false-alarm rate, the human operator can't currently query the algorithm's basis of determining the classification beyond a simple confidence factor, as it is encoded in the weights of the neural network. As the role of AI-robots on the battlefield expands, the AI's rationale for selecting a specific course of action, or its determination of an object as a threat, will need to be explainable to the human operator. The AI's course of action must also be coordinated and deconflicted with other blue force actors, to avoid impeding adjacent operations and operational elements. It is also important to remember that military decisions made during war are subject to legal review, so collecting a chain of evidence on what algorithms were used, and their basis of determining actions, is critical. Future AI-powered robots don't need to be perfect, but they do need to perform important functions consistently and understandably, storing reviewable information on perceptions, reasoning, and decisions, so that human soldiers trust them and can determine when their use is acceptable. A lack of transparency could diminish the human operator's trust in the AI-robot, and the operator may not know when to use or not use the AI-robot safely and appropriately.

Trust: Vulnerabilities to Cyber and Adversarial Attacks

In recent studies, neural network-based AI computer vision algorithms were tricked into changing the classification of an object after relatively few pixels in the image to be identified had been changed (Elsayed et al. 2018). The implication for the battlefield is that an adversary could modify the classification outcomes of the AI so that the robot would take unintended actions. Trust in an AI-empowered combat robot requires that the robot perform tasks well and consistently in a manner that is comprehensible to the operator, and that it not be corruptible by the adversary. If an adversary

could use cyberattacks to take over the robot, this would weaken the robot's operational effectiveness by undermining the operator's trust that the robot would complete missions as directed. Recently, Russia claimed that the U.S. took control of 13 UAVs and attempted to use them to attack a Russian military base in Syria (Associated Press 2018). Addressing cyber vulnerabilities will be critical to the successful use of AI-robots on the battlefield, as AI-robots that can be redirected or co-opted from their intended use would pose unacceptable risk to friendly forces and noncombatants.

The Future of AI-Robots on the Battlefield

The U.S., China, and Russia continue to increase defense investments for the application of AI to a broader set of potential missions (Allen 2019; Apps 2019). The U.S. Department of Defense (DoD) released an overarching AI strategy (Moon Cronk 2019), as did China, and Russia is scheduled to do so in mid-2019, demonstrating all three countries' long-term commitment to making AI a bigger and more coordinated part of their defense infrastructures. While the imminent use of human-like robotic killing machines still gets top billing in discussions and articles on AI, this is only one of many potential AI applications on the future battlefield, and it is not necessarily a concern for the near future. Several of the current AI applications in weapons focus on improved human-machine teaming and increased deployment of unmanned vehicles for various missions, mostly nonlethal. If they are capable of lethal force, fielded systems to date have had defensive purposes only. However, this norm may be challenged as future AI applications are likely to cover a wider spectrum, from data sorting for enhanced intelligence analysis to replacement of human decision-making in targeting decisions.

What AI-Robots May Be Able to Do in the Near Future

Countries with sophisticated militaries will continue to invest in AI-enabled weapons. However, these systems are expensive to develop and field, and they are useful for a set of missions but not for all combat-related tasks. Therefore, from a cost-benefit perspective, AI is not likely to fully replace human operators as many assume, but rather to augment them and perhaps take over specific missions or aspects of missions. Refinements in the near future could broaden the number of relevant missions, and better AI-enabled weapons may challenge the current role of the human in the targeting cycle. Below is a discussion of five areas of development that are likely to impact the future battlefield.

Improving What We Already Have and Expanding Missions

Research into and development of systems such as UAVs for use in swarms, as well as UUVs, continues, but the emphasis is on more autonomy (e.g., the ability for drones to accurately and consistently target and engage other drones without human input, the ability of armed UUVs to travel long distances undetected) and greater coordination among multiple platforms (Zhao 2018). However, new payloads represent a growing concern. Russia is thought to be developing nuclear-powered UUVs capable of carrying nuclear payloads (U.S. Dept. of Defense 2018). And the idea of nuclear armed drones is out there but has not yet been realized.

Perhaps the most unsettling future development then is not the AI-enabled platform—although the ability to travel greater distances with stealth is important—but what payload the AI-robot is able to deploy. Debate and calls for international agreements have emerged in response to the idea of applying AI to any part of the nuclear enterprise, out of fear that it will upend nuclear stability entirely (Geist and Lohn 2018). This fear is based on the belief that AI will encourage humans to make catastrophic decisions or that AI will do so on its own. As discussed previously, the current limitations of AI to understand context, reason, and take action would need to be overcome, and even then, the mix of AI with nuclear would benefit from a broader debate among policy-makers and scientists.

Dull, Dirty, and Dangerous Jobs

Numerous “dull, dirty, and dangerous” jobs that soldiers used to do are increasingly being done by AI-empowered machines. Resources are being used to develop better unmanned trucks and logistics vehicles to reduce the number of human operators needed to do these jobs. Currently, unmanned logistics vehicles could enable new missions that would otherwise be precluded because of the inability to safely get materials situated in the desired location. In another use, an AI-enabled unmanned truck could advance ahead of a formation of tactical vehicles if the commander believes the area has buried IEDs. An analog exists for an unmanned lead aircraft advancing first into contested airspace. The application of AI here does two things—it frees up human operators and reduces the operator’s risk.

Augmentation of Human Decision-Making

For both analyzing data and reviewing surveillance footage, people currently pore over data or stare at a screen for hours on end watching video footage. In the case of surveillance footage, even a brief deviation of the eyes from the screen can

result in the analyst missing something critical. Current AI applications have become better at recognizing objects and people, but a machine’s ability to both recognize an object or a human and make an assessment about whether the situation is dangerous, or whether something unexpected is occurring, is still in development. Scientists are using novel methods to train computers to learn how something works by observing it, and in the course of this learning, to detect real from fake behavior (Li et al. 2016). This use of AI has implications for a variety of security applications. For the analyst staring at the screen, it is possible that the AI could make data analysis (such as pattern-of-life detection) much easier by doing some of the work for them, and perhaps doing it better. Such analysis is necessary for the development of strike packages, so the AI would augment the human in developing targeting priorities. While many AI implementations for dull, dirty, and dangerous jobs are instantiated as a physical robotic system, this application is reducing the human burden in the information space.

It is possible that application of AI for information analysis could advance to the point that militaries seek to use it in lieu of a human conducting target identification and engagement. While most Western policy-makers are ok with drones killing drones, when it comes to lethal force, the standing view is that a human must be in the loop. This view is reflected in DoD policy and in its interpretation of international legal obligations (U.S. Dept. of Defense 2012). However, as technology improves and the speed of battle increases, this policy may be challenged, especially if adversaries’ capabilities pose an increasingly greater threat.

Replacing Human Decision-Making

One of the biggest advantages in warfare is speed. AI-robot weapons systems operating at machine speeds would be harder to detect and counter, and AI-empowered information systems allow adversaries to move more quickly and efficiently. This increase in speed means the time for human decision-making is compressed, and in some cases, decisions may need to be made more quickly than humans are able to make them. Advances in the application of AI to weapons may even be used to preempt attacks, stopping them before they occur. In the physical world, this may be achieved through faster defensive systems that can detect and neutralize incoming threats. Examples of these systems already exist (Aegis, Harpy), but new weapons, such as hypersonics, will make traditional detection and interception ineffective. In the information world, AI as applied to cyber weapons is a growing concern. The U.S. DoD’s policy on autonomy in weapons systems specifically excludes cyber weapons because it recognizes that cyber warfare would not permit time for a human decision (U.S. Dept. of Defense 2012). Indeed, countries continue to invest in AI-enabled cyber

weapons that autonomously select and engage targets and are capable of counter-autonomy—the ability of the target to learn from the attack and design a response (Meissner 2016). What is less certain is how cyber warfare will escalate. Will AI systems target infrastructure that civilians rely on, or will an AI-initiated attack cause delayed but lethal downstream effects on civilians?

In addition to speed, a second driver for taking humans out of the loop in AI is related to scale—the volume of information received or actions taken exceeds human capacity to understand and then act. For example, AI-powered chatbots used for social media and information operations can simultaneously analyze and respond to millions of tweets (Carley et al. 2013; Zafarani et al. 2014). This volume of information exceeds a human’s capability to monitor and intervene. So, the increasing scale of data and actions, as well as the compressed time for decision-making, are necessitating autonomy, without the ability for human intervention, during the AI’s action selection and execution.

In the physical domain, countries are developing weapons designed to remove human decision-making. The most common forms are weapons capable of following a target based on a computational determination within the weapon rather than a human input. These are not new, but emerging applications allow weapons to loiter and select targets from a library of options. For some militaries, it is unclear whether the human in the loop in these circumstances simply determines the library of options ahead of time or is able to evaluate the accuracy of the target selected during the operation. As computer vision capabilities improve, it is possible that militaries will seek to remove the human from all aspects of the targeting cycle. In A2AD environments, where there are few if any civilians, this application may be compelling. However, for a more diverse set of missions, it may not be appropriate. This delegation of targeting decisions is at the heart of ongoing international debates focused on the regulation of AI-enabled weapons systems.

Electronic Warfare

Also driven by increasing speed and reduced time for decision-making, the application of AI to EW is an area of increasing importance for several countries. Systems are being developed for use on various platforms that can detect an adversary EW threat while nearly simultaneously characterizing it and devising a countermeasure. Projects have also been underway for applications of AI to space assets, such as DARPA’s Blackjack, to allow autonomous orbital operations for persistent DoD network coverage (Thomas 2019). And, the F-35 is reportedly designed with very sophisticated AI-enabled EW capabilities for self-protection and for countering air defense systems (Lockheed Martin 2019). These are just examples, and this area is likely to grow as research and applications continue to evolve.

JHU/APL Research Toward AI-Robots for Trusted, Real-World Operations

The Johns Hopkins University Applied Physics Laboratory (JHU/APL) has decades of experience developing and testing AI-robots for challenging real-world applications, from defense to health. JHU/APL often serves as a trusted partner to U.S. DoD organizations, helping to facilitate an understanding of capabilities and limitations of state-of-the-art AI as it relates to use in military applications. As AI-robots for air, ground, surface, and underwater use are being developed for use on the battlefield, JHU/APL is committed to ensuring that they are developed, tested, and operated in accordance with U.S. safety, legal, and policy mandates.

Areas of AI-Robotics Research at JHU/APL

In addition to developing simulated and real-world semiautonomous and autonomous AI-robots, JHU/APL performs research on the evaluation, testing, and security of these systems. This research includes developing external control systems for AI-robots to ensure safety, AI systems that are self-correcting, explainable and transparent, and AI with ethical reasoning. The following sections describe each of these areas of research.

Safe Testing of Autonomy in Complex Interactive Environments

JHU/APL is the lead developer of the Safe Testing of Autonomy in Complex Interactive Environments (TACE) program for the Test Resource Management Center and other DoD sponsors. TACE provides an onboard “watchdog” program that monitors the behaviors of an AI-robot and takes control of the platform if the AI produces behaviors that are out of bounds relative to the safety, policy, and legal mandates for the mission (see Fig. 5). TACE operates in live, virtual, and constructive environments, which combine simulation and reality to allow AI algorithms to learn appropriate behavior before operating in the real world. TACE is currently used during the development and testing of autonomous systems to maintain safety and control of AI-robot systems. TACE is independent from the robot’s AI autonomy system, has separate access to sensor and control interfaces, and is designed to deploy with the AI-robot to ensure compliance with range safety constraints (Lutz 2018).

Self-Regulating AI

JHU/APL has invested independent research and development funding to examine novel methods of AI self-regulation. Self-regulating AI is a hybrid model that uses both rules-based encoding of boundaries and limitations with neural network-based AI that produces adaptive behaviors in

Fig. 5 TACE. TACE is an architecture for autonomy testing that serves as an onboard watchdog over AI-robots. TACE monitors all perceptions and behaviors of the AI-robot, takes over when the AI attempts to

perform a behavior that is outside of range safety constraints, and records all outcomes for post-mission evaluation (Lutz 2018)

complex environments. While TACE is an external process that ensures that AI behaviors meet mission requirements, and takes over if they don't, self-regulation incorporated the rules into the AI's decision-making process. A key element to this research is the AI's ability to predict environmental effects of sets of possible actions. Developed in simulation, the AI is penalized for decision choices that result in environmental effects that violate mission rules and parameters. In this manner, the AI learns to make decisions that don't violate the mission requirements, which include the Rules of Engagement, the Law of Armed Conflict (Dowdy et al. 2015), and U.S. policies. Self-regulating AI can be used in conjunction with an external watchdog governor system to provide multiple layers of protection against aberrant AI behaviors.

Explainable AI and Human-Machine Interactions

Even if AI-robots perform mission-critical tasks well and reliably, how humans and AI-robots work together in manned-unmanned teams is an open area of research. Simple AI

systems that perform a function repeatedly without making decisions are merely tools; the issues in human-machine interactions arise when the AI-robot acts as an agent, assessing the environment, analyzing its mission goals and constraints, and then *deciding to do something*. How predictable will the AI-robot be to human teammates? How transparent will the robot's perception, decision-making, and action selection be—especially in time-critical situations? JHU/APL is researching how humans develop trust in other humans, and how that process is altered when humans work with AI-robots. This includes having the AI learn when it needs to provide information to its human teammate, when to alert the human to a change in plans, and, importantly, when to not make a decision and ask the human for help. Figure 6 presents an integration of research on the elements of human trust in autonomy, including human, autonomous entity, team, and task factors. Developing an AI system that understands what it knows and doesn't know, or what it can and cannot do, is especially challenging. It requires that the AI have the capability to self-assess and provide a level of confidence

Fig. 6 *The theoretical underpinnings of trust.* Important elements in the development of human trust in autonomy are presented—including human, autonomous entity, team, and task factors—that, taken together, present a method for evaluating trust and the breakdown in trust (Marble and Greenberg 2018)

for every perception, decision, and action. The ability to know when an AI-robot can and can't be used safely and appropriately in the battlefield is a critical step in developing meaningful interactions and trust between humans and AI-robots.

Ethical AI

A crucial step in designing autonomous AI-robots is to develop methods for the AI to encode our values, morals, and ethics. JHU/APL teamed up with the Johns Hopkins Berman Institute of Bioethics to study how to enable an AI to perform moral reasoning. Starting from the first of Isaac Asimov's laws of robotics—"A robot may not injure a human being or, through inaction, allow a human being to come to harm" (Asimov 1950)—this research explores the philosophical, governance, robotics, AI, cognitive science, and human-machine interaction aspects of the development of an AI with moral reasoning. For an AI to perform moral reasoning, it would necessarily need to encode what a human is, what injuries to humans are, and the ways by which humans are injured. Moral and ethical decision-making to prevent an AI from harming a human starts with perception. Even basic capabilities—such as differentiating living people from objects—are challenges for current AI algorithms. The JHU/APL research formulates a method of moral-scene assessment for intelligent systems, to give AI systems a "proto-conscience" that allows them to identify in a scene the elements that have moral salience (see Fig. 7). This moral perception reifies potential insults and injuries to persons for the AI to reason over and provides a basis for ethical

behavioral choices in interacting with humans (Greenberg 2018).

Conclusions: The Future of AI for Battlefield Robotics

Rapid advancements in the fields of AI-robots have led to an international arms race for the future of the battlefield. AI systems will become more capable and more lethal, and wars eventually will be fought at machine speed—faster than humans can currently process information and make decisions. As AI reasoning systems mature, more complex behaviors will be delegated to AI-robot systems, with human control on the loop or out of the loop entirely. Proliferation will result in large numbers—hundreds of thousands to millions—of AI-robots on the battlefield. JHU/APL is working to develop multiple means by which AI-robots on the battlefield can be used within the legal, ethical, and policy mandates of the U.S. Government.

Competing Visions of the Future of AI-Robots

Beyond the near-term development of new AI and AI-robot capabilities is the question of the long-term relationship between humans and AI-enhanced robots. As of now, AI-empowered robots are extensions of human activities. As AI systems evolve to operate autonomously, self-assess and self-regulate, and work effectively alongside humans to accomplish tasks, what will the roles be for the humans and

Fig. 7 *Moral-scene assessment.* Starting from Asimov's Three Laws of Robotics (Asimov 1950), this decomposition (above, left) extracts the perceptual and reasoning capabilities needed for an AI-robot to engage

in moral consideration. Ethical decision-making to prevent an AI from harming a human starts with perception (above, right), providing a basis for ethical behavior choices (Greenberg 2018)

the robots? The following sections review two projections of the future relationship between humans and robots and then present an alternative possibility.

Human Subservience to AI-Robots

The most popular envisioned future relationship between humans and AI-robots is exemplified by the movie "The Terminator," in which an AI defense system (Skynet) becomes self-aware and decides to enslave and destroy the human race (Cameron 1984). The central premise is that humans build a synthetic reasoning system that is able to process faster and control more sensors and weapons than humans can. In the movie, humans become obsolete. Elon Musk recently stated that autonomous AI-powered weapons could trigger the next world war, and that AI could become "an immortal dictator" (Holley 2018). A number of experts have called for a ban on "killer robots," to ensure that humans are not made to be subservient to, or destroyed by, robots (Gibbs 2017). The concerns about the continuing advancement of AI technologies has led the Institute of Electrical and Electronics Engineers (IEEE) to develop a global initiative on AI development, including a primer on ethically aligned design, to ensure that human well-being is prioritized in

the development and deployment of AI and AI-empowered autonomous systems (IEEE 2019).

Human Dominance over AI-Robots

An alternative view of the future relationship between humans and robots that is equally concerning is one in which AI-empowered robots are developed and evolved as slaves to humans. While this is less of a concern in the near term when AI-robots are primarily tools, as we move toward AGI, we move into a space where AI-robots act as independent agents and become indistinguishable from humans. DeepMind has developed an AI (AlphaGo) that beat the best humans in chess, shogi, and Go (Silver et al. 2018) and an AI system (AlphaStar) that can beat the best human StarCraft II video game players (AlphaStar Team, DeepMind 2019). It is not far-fetched to imagine the continued development of super-human AI capabilities in the next 20–50 years. The question arises: if an AI approaches AGI—and is not distinguishable from human rationality using human perception—do humans still have the right to "turn it off," or does a self-aware AI have the rights of personhood (van Hooijdonk 2018; Sheliazhenko 2017)?

Co-Evolution with AI-Robots

In contrast with the two previous envisioned futures, where humans are either subservient to or destroyed by robots or vice versa, another possibility exists. It is important to remember that machine learning and AI algorithms were developed by studying mammalian (including human) neural computations in the brain, and that future developments in AI will occur through an increased understanding of the brain (Hassabis et al. 2017). In this respect, developing AI is a lot like looking in a mirror. We are currently exploring a very minute part of what a human brain can do, through experimental and computational research, and then creating mathematical models to try to replicate that neural subsystem. It is vital to understand that this process also helps us learn about how we, as humans, work—what our computational strengths and weaknesses are—and how human diseases and limitations could be overcome.

An example of this is the study of memory—both human memory storage and retrieval—and AI uses of memory. Human memory is quite limited, with only a few independent, short-term memory slots (Baddeley and Hitch 1974) that are quite volatile and long-term memory that is often inaccurate, error prone, and retroactively alterable (Johnson 2010). Persons with Alzheimer’s disease have lost the ability to locate individual memories, as the index to memory location has been degraded by the disease (RIKEN 2016). Recent research used a mathematical model to encode episodic memories (i.e., remembering where you put your car keys) in the hippocampus, a function that is degraded in people with Alzheimer’s disease. These enhanced neural signals were then sent back into the patient’s brain, resulting in an improvement in short-term memory of 35–37% (Wake Forest Baptist Medical Center 2018). Future integration of AI with brain-computer interfaces could result in improved memory for both Alzheimer’s patients and healthy humans. Similar improvements to human perception, reasoning, decision-making, and learning are possible.

So, an alternative vision of the future of humans and robots is one where humans and AI co-evolve, each learning from the other and improving each other. If the development of AGI is the ultimate human mirror showing us our potential for self-destruction and destruction of others, it is also a mirror that shows us the potential for humans to grow, change, and evolve in positive ways. As in healthcare research, where the more we understand human physiology and pathologies, the more creative solutions we can develop, research toward AGI offers us the opportunity to understand ourselves and to improve the human condition—not as enforced by external organizations or movements, but as a matter of individual choice. In this way, co-evolution between humans and AI becomes like any other method of human self-improvement—education, healthcare, use of technology—in

that we can choose to what level we participate, weighing the pros and cons of that decision.

This concept of co-evolution between humans and AI-robots also addresses the fears of the first two envisioned futures. A central concern of the “Terminator” future is that humans will be “left behind” by AI systems, leading to human destruction or subordination (Caughill 2017). By co-evolving, humans will be able to understand and keep up with emerging AI systems. If humans are able to process information at machine speeds, and at massive scale, we will be able to intervene in decisions made by AI-robots before actions are taken. In this manner, human operators will be able to ensure that AI systems are operating legally and ethically, within the human-defined mission parameters. Also, by understanding our human limbic systems and emotions better—including our fight or flight responses and fear of the “other”—we have the potential to co-evolve with AI to improve our self-control beyond the current human tendencies to dominate and subordinate, as described in the enslavement future projection. Co-evolution between humans and AI-robots could lead to the voluntary improvement of both individual humans and human society.

In conclusion, AI and AI-robots are an increasing presence on the battlefield, and are being developed with greater speed, scale of operation, autonomy, and lethality. Understanding how AI and AI-robots operate, the limitations of current AI-robots, and the likely future direction of these systems can inform policy-makers in their appropriate uses and potential for misuse. JHU/APL has leveraged its expertise in the development of AI and AI-robot systems to perform research in how these systems can be developed responsibly and safely, and used within U.S. safety, legal, and policy mandates. While there are highly divergent opinions on the relationship of humans to future AI systems—including fears of humans being dominated by AI, or vice versa—there is an opportunity for humans to learn and advance from interactions with AI. In a future of co-evolution, humans would be able to ensure that AI reflects human ethical and legal values because humans would be able to continue to understand and interact with AI systems.

References

- Allen, G. C. (2019). *Understanding China’s AI strategy*. Washington, DC: CNAS. Available via CNAS. Retrieved March 20, 2019, from <https://www.cnas.org/publications/reports/understanding-chinas-ai-strategy>.
- AlphaStar Team, DeepMind. (2019). *AlphaStar: Mastering the real-time strategy game StarCraft II*. Available via DeepMind. Retrieved March 6, 2019, from <https://deepmind.com/blog/alphastar-mastering-real-time-strategy-game-starcraft-ii/>
- Apps, P. (2019). *Commentary: Are China, Russia winning the AI arms race?* Available via Reuters. Retrieved March 20, 2019, from

- <https://www.reuters.com/article/us-apps-ai-commentary/commentary-are-china-russia-winning-the-ai-arms-race-idUSKCN1P91NM>
- Asimov, I. (1950). *The Isaac Asimov collection*. New York: Doubleday.
- Associated Press. (2018). *Russia claims US aircraft took control of drones in attempted attack on its Syrian base*. Available via Military Times. Retrieved March 6, 2019, from <https://www.militarytimes.com/flashpoints/2018/10/25/russia-claims-us-aircraft-took-control-of-drones-in-attempted-attack-on-syrian-base/>.
- Attherton, K. (2018). *See China's massive robot boat swarm in action*. Available via C4ISRNET. Retrieved March 5, 2019, from <https://www.c4isrnet.com/unmanned/2018/06/01/see-chinas-massive-robot-boat-swarm-in-action/>.
- Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *Psychology of learning and motivation*, vol. 8 (pp. 47–89). Cambridge, MA: Academic Press. [https://doi.org/10.1016/S0079-7421\(08\)60452-1](https://doi.org/10.1016/S0079-7421(08)60452-1).
- Blum, B. (2018). *9 Israeli drone startups that are soaring to success*. Available via ISRAEL21c. Retrieved March 5, 2019, from <http://www.israel21c.org/9-israeli-drone-startups-that-are-soaring-to-success/>.
- Brown, J. (2016). *Types of military drones: The best technology available today*. Available via MyDroneLab. Retrieved March 5, 2019, from <http://mydronelab.com/blog/types-of-military-drones.html>.
- Cameron, J. (1984). *The Terminator* [Motion picture]. https://www.imdb.com/title/tt0088247/?ref_=nv_sr_4
- Carley, K. M., Pfeffer, J., Liu, H., Morstatter, F., & Goolsby, R. (2013). Near real time assessment of social media using geo-temporal network analytics. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (pp. 517–524). New York: ACM. doi: <https://doi.org/10.1145/2492517.2492561>.
- Caughill, P. (2017). *Artificial intelligence is our future. But will it save or destroy humanity?* Available via Futurism.com. Retrieved March 7, 2019, from <https://futurism.com/artificial-intelligence-is-our-future-but-will-it-save-or-destroy-humanity>.
- Colquitt, J. A., Scott, B. A., & LePine, J. A. (2007). Trust, trustworthiness, and trust propensity: A meta-analytic test of their unique relationships with risk taking and job performance. *The Journal of Applied Psychology*, 92(4), 909–927. <https://doi.org/10.1037/0021-9010.92.4.909>.
- Defense Science Board. (2017). *Summer study on autonomy*. Washington, DC: Defense Science Board. Available via Homeland Security Digital Library. Retrieved February 21, 2020, from <https://www.hsdl.org/?abstract&did=794641>.
- Dillow, C. (2016). *All of these countries now have armed drones*. Available via Fortune. Retrieved March 4, 2019, from <http://fortune.com/2016/02/12/these-countries-have-armed-drones/>.
- Dodge, S., & Karam, L. (2017). A study and comparison of human and deep learning recognition performance under visual distortions. In: *2017 26th International Conference on Computer Communication and Networks (ICCCN)* (pp. 1–7). doi: <https://doi.org/10.1109/ICCCN.2017.8038465>.
- Dowdy, L. R., Cherry, L. J., Berry, L. J., Bogar, L. T., Clark, L. B., Bogar, L. T., Blank, M. J., MJ, D. S., Festa, M. M., Lund, M. M., Radio, M. K., & Thorne, T. (2015). *Law of armed conflict deskbook* (5th ed., p. 253). Charlottesville, VA: International and Operational Law Department, U.S. Army Judge Advocate General's Legal Center and School.
- Elsayed, G., Shankar, S., Cheung, B., Papernot, N., Kurakin, A., Goodfellow, I., & Sohl-Dickstein, J. (2018). Adversarial examples that fool both computer vision and time-limited humans. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, & R. Garnett (Eds.), *Advances in neural information processing systems* (Vol. 31, pp. 3910–3920). Red Hook, NY: Curran Associates. Retrieved from <http://papers.nips.cc/paper/7647-adversarial-examples-that-fool-both-computer-vision-and-time-limited-humans.pdf>.
- Etzioni, A., & Etzioni, O. (2017). Pros and cons of autonomous weapons systems. *Military Review*, 2017(3), 72–80.
- Fox, M. A. (2017). Understanding modern Russian war: Ubiquitous rocket, artillery to enable battlefield swarming, siege warfare. *Fires*, 2017(5), 20–25. Retrieved from <https://sill-www.army.mil/firesbulletin/archives/2017/sep-oct/additional-articles/Understanding%20modern%20Russian%20war.pdf>.
- Freedberg, S. J. Jr. (2015). *Russian drone threat: Army seeks Ukraine lessons*. Available via Breaking Defense. Retrieved March 4, 2019, from <https://breakingdefense.com/2015/10/russian-drone-threat-army-seeks-ukraine-lessons/>.
- Gambrell, J., & Shih, G. (2018). *Chinese armed drones now flying over Mideast battlefields. Here's why they're gaining on US drones*. Available via Military Times. Retrieved March 4, 2019, from <https://www.militarytimes.com/news/your-military/2018/10/03/chinese-armed-drones-now-flying-over-mideast-battlefields-heres-why-theyre-gaining-on-us-drones/>.
- Geist, E., & Lohn, A. (2018). *How might artificial intelligence affect the risk of nuclear war?* Santa Monica, CA: RAND Corporation. <https://doi.org/10.7249/PE296>.
- Gibbs, S. (2017). *Elon Musk leads 116 experts calling for outright ban of killer robots*. Available via The Guardian. Retrieved February 21, 2020, from <https://www.theguardian.com/technology/2017/aug/20/elon-musk-killer-robots-experts-outright-ban-lethal-autonomous-weapons-war>.
- Goodale, M., & Milner, D. (2006). *One brain—two visual systems*. Available via The Psychologist. Retrieved March 5, 2019, from <https://thepsychologist.bps.org.uk/volume-19/edition-11/one-brain-two-visual-systems>.
- Greenberg, A. (2018). *Moral-scene assessment for intelligent systems. In Presented at the 2018 IEEE Resilience Week*, Denver, CO, 2018.
- Hassabis, D., Kumaran, D., Summerfield, C., & Botvinick, M. (2017). Neuroscience-inspired artificial intelligence. *Neuron*, 95(2), 245–258. <https://doi.org/10.1016/j.neuron.2017.06.011>.
- Holley, P. (2018). *Elon Musk's nightmarish warning: AI could become 'an immortal dictator from which we would never escape'*. Available via Washington Post. Retrieved March 6, 2019, from <https://www.washingtonpost.com/news/innovations/wp/2018/04/06/elon-musks-nightmarish-warning-ai-could-become-an-immortal-dictator-from-which-we-would-never-escape/>.
- IEEE. (2019). *IEEE-Standards Association*. The IEEE global initiative on ethics of autonomous and intelligent systems. Available via IEEE. Retrieved March 6, 2019, from <https://standards.ieee.org/industry-connections/ec/autonomous-systems.html>.
- Intel. (2019). *Intel drone light show and Intel True View technology enhance Pepsi Super Bowl LIII Halftime Show and Super Bowl LIII viewing experience*. Available via Intel Newsroom. Retrieved March 20, 2019, from <https://newsroom.intel.com/news-releases/intel-drones-true-view-super-bowl-liiii/>.
- Jarretson, P. (2018). *How does Russia's integrated air defense network work?* Available via Quora. Retrieved March 5, 2019, from <https://www.quora.com/How-does-Russias-integrated-air-defense-network-work>.
- Johnson, J. (2010). Chapter 7: Our attention is limited, our memory is imperfect. In J. Johnson (Ed.), *Designing with the mind in mind* (pp. 121–130). Waltham, MA: Elsevier.
- Karnozov, V. (2017). *Extent of Russian UAV use over Syria revealed*. Available via AIN Online. Retrieved March 5, 2019, from <https://www.ainonline.com/aviation-news/defense/2017-09-21/extent-russian-uav-use-over-syria-revealed>.

- Keller, J. (2019a). *Future unmanned underwater vehicles with machine autonomy for deep-sea missions are focus of DARPA angler program*. Available via Military and Aerospace Electronics. Retrieved March 5, 2019, from <https://www.militaryaerospace.com/articles/2019/01/unmanned-deep-sea-machine-autonomy.html>.
- Keller, J. (2019b). *Navy eyes sonar, communications, and power upgrades for black pearl unmanned underwater vehicles (UUVs)*. Available via Military and Aerospace Electronics. Retrieved March 5, 2019, from <https://www.militaryaerospace.com/articles/2019/01/unmanned-underwater-vehicles-uuv-sonar-power.html>.
- Knight, W. (2017). *The dark secret at the heart of AI*. Available via MIT Technology Review. Retrieved March 5, 2019, from <https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/>.
- Kotseruba, I., & Tsotsos, J. K. (2018). 40 years of cognitive architectures: Core cognitive abilities and practical applications. *Artificial Intelligence Review*, 53, 17. <https://doi.org/10.1007/s10462-018-9646-y>.
- Li, W., Gauci, M., & Groß, R. (2016). Turing learning: A metric-free approach to inferring behavior and its application to swarms. *Swarm Intelligence*, 10(3), 211–243. <https://doi.org/10.1007/s11721-016-0126-1>.
- Lockheed Martin. (2019). *Unprecedented battlefield access*. Available via Lockheed Martin. Retrieved March 18, 2019, from <https://www.f35.com/about/capabilities/electronicwarfare>.
- Lutz, R. (2018). Taming the Terminator. *Presented at the Inter-service/Industry Training, Simulation, and Education Conference (IITSEC)*, Orlando, FL, November 26–30, 2018.
- Marble, J. L., & Greenberg, A. (2018). ESCAPE with PARTNER: Experimental suite for cooperatively-achieved puzzle exercises with platform assessing risk and trust in non-economic relationships. In *Proceedings of the 2018 IEEE Resilience Week: Avoiding Skynet*. Denver, CO.
- Markowitz, J., Gardner, R. W., & Llorens, A. J. (2018). On the complexity of reconnaissance blind chess. *ArXiv:1811.03119 [Cs]*. <http://arxiv.org/abs/1811.03119>
- Meissner, C. (2016). *The most militarily decisive use of autonomy you won't see*. Available via Defense One. Retrieved March 23, 2019, from <https://www.defenseone.com/ideas/2016/11/most-militarily-decisive-use-autonomy-you-wont-see-cyberspace-ops/132964/>.
- Moon Cronk, T. (2019). *DOD unveils its artificial intelligence strategy*. Available via U.S. Dept. of Defense. Retrieved March 18, 2019, from <https://dod.defense.gov/News/Article/Article/1755942/dod-unveils-its-artificial-intelligence-strategy/>.
- Nacouzi, G., Williams, J. D., Dolan, B., Stickells, A., Luckey, D., Ludwig, C., Xu, J., Shokh, Y., Gerstein, D. M., & Decker, M. (2018). *Assessment of the proliferation of certain remotely piloted aircraft systems: Response to section 1276 of the National Defense Authorization act for fiscal year 2017*. Santa Monica, CA: RAND Corporation. <https://doi.org/10.7249/RR2369>.
- O'Brien, M. (2018). *Army looks for a few good robots, sparks industry battle*. Available via Army Times. Retrieved March 5, 2019, from <https://www.armytimes.com/news/your-army/2018/12/30/army-looks-for-a-few-good-robots-sparks-industry-battle/>.
- Osborn, K. (2018). *The Army is building robot attack tanks [blog post]*. Available via National Interest. Retrieved March 5, 2019, from <https://nationalinterest.org/blog/buzz/army-building-robot-attack-tanks-26541>.
- Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *The Behavioral and Brain Sciences*, 1(4), 515–526. <https://doi.org/10.1017/S0140525X00076512>.
- RIKEN. (2016). *Flipping a light switch recovers memories lost to Alzheimer's disease mice*. Available via Science Daily. Retrieved March 7, 2019, from <https://www.sciencedaily.com/releases/2016/03/160316140419.htm>.
- Robitzski, D. (2018). *When will we have artificial intelligence as smart as a human? Here's what experts think*. Available via Futurism. Retrieved March 5, 2019, from <https://futurism.com/human-level-artificial-intelligence-agi>.
- Roblin, S. (2019). *Russia's Uran-9 robot tank went to war in Syria (it didn't go very well)*. Available via National Interest. Retrieved March 5, 2019, from <https://nationalinterest.org/blog/buzz/russias-uran-9-robot-tank-went-war-syria-it-didnt-go-very-well-40677>.
- Russell, S., & Norvig, P. (2003). *Artificial intelligence: A modern approach*. Upper Saddle River, NJ: Prentice Hall. Retrieved from <http://aima.cs.berkeley.edu/>.
- Scharre, P. (2018). *Why drones are still the future of war*. Available via CNAS. Retrieved March 22, 2019, from <https://www.cnas.org/publications/commentary/why-drones-are-still-the-future-of-war>.
- Shannon, C. E. (1950). XXII. Programming a computer for playing chess. *Philosophical Magazine*, 41(314), 256–275. <https://doi.org/10.1080/14786445008521796>.
- Shaw, J. (2019). *Artificial intelligence and ethics*. Available via Harvard Magazine. Retrieved March 5, 2019, from <https://harvardmagazine.com/2019/01/artificial-intelligence-limitations>.
- Sheliazhenko, Y. (2017). Artificial personal autonomy and concept of robot rights. *European Journal of Law and Political Sciences*, 2017(1), 17–21. <https://doi.org/10.20534/EJLPS-17-1-17-21>.
- Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T., Lillicrap, T., Simonyan, K., & Hassabis, D. (2018). A general reinforcement learning algorithm that masters chess, shogi, and go through self-play. *Science*, 362(6419), 1140–1144. <https://doi.org/10.1126/science.aar6404>.
- Singer, P. W. (2001). *In the loop? Armed robots and the future of war*. Available via Brookings. Retrieved March 5, 2019, from <https://www.brookings.edu/articles/in-the-loop-armed-robots-and-the-future-of-war/>.
- Tegmark, M. (2018). *Life 3.0: Being human in the age of artificial intelligence*. New York: Knopf. Retrieved from <https://www.penguinrandomhouse.com/books/530584/life-30-by-max-tegmark/9781101970317/>.
- Thomas, P. (2019). *Blackjack*. Available via DARPA. Retrieved March 18, 2019, from <https://www.darpa.mil/program/blackjack>.
- Trevithick, J. (2018). *Russia's s-500 air defense system reportedly hits target nearly 300 miles away*. Available via The Drive. Retrieved March 5, 2019, from <http://www.thedrive.com/the-war-zone/21080/russias-s-500-air-defense-system-reportedly-hits-target-nearly-300-miles-away>.
- Tromp, J. (2016). Number of legal 12xn go positions. Available via John Tromp. Retrieved from <https://tromp.github.io/go/L12.html>
- U.S. Army Aviation Center of Excellence. (2010). *"Eyes of the Army": U.S. Army roadmap for unmanned systems, 2010–2035*. Available via Homeland Security Digital Library. Retrieved from <https://www.hsdl.org/?abstract&did=705357>
- U.S. Dept. of Defense. (2012). *Directive 3000.09: Autonomy in weapons systems*. Available via Executive Services Directorate. Retrieved March 4, 2019, from <https://www.esd.whs.mil/portals/54/documents/dd/issuances/dodd/300009p.pdf>.
- U.S. Dept. of Defense. (2018). *Nuclear posture review Feb 2018*. Available via U.S. Dept. of Defense. Retrieved March 20, 2019, from <https://www.defense.gov/News/Special-Reports/NPR-2018>.
- van Hooijdonk, R. (2018). *If AI becomes self-aware, will we need to give it the rights we normally reserve for humans?*. Available via Richard van Hooijdonk. Retrieved March 6, 2019, from <https://www.richardvanhooijdonk.com/en/blog/ai-becomes-self-aware-will-need-give-rights-normally-reserve-humans/>.
- Wake Forest Baptist Medical Center. (2018). *Prosthetic memory system successful in humans*. Available via Science Daily. Retrieved

- March 7, 2019, from <https://www.sciencedaily.com/releases/2018/03/180327194350.htm>.
- Wolf, H. (2018). *3 Reasons why China is the global drones leader*. Available via World Economic Forum. Retrieved March 5, 2019, from <https://www.weforum.org/agenda/2018/09/china-drones-technology-leader/>.
- Zafarani, R., Ali Abbasi, M., & Liu, H. (2014). *Social media mining*. New York: Cambridge University Press. Retrieved from <http://dmml.asu.edu/smm/>.
- Zhao, T. (2018). *Tides of change: China's nuclear ballistic missile submarines and strategic stability*. Available via the Carnegie Endowment for International Peace website. Retrieved March 20, 2019, from https://carnegieendowment.org/files/Zhao_SSBN_final.pdf.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.