Research Paper

M. L. Cummings
International Security Department and US and the Americas Programme
January 2017

Artificial Intelligence and the Future of Warfare

CHATHAM HOUSE
The Royal Institute of International Affairs

Summary

- Both military and commercial robots will in the future incorporate 'artificial intelligence' (AI) that could make them capable of undertaking tasks and missions on their own. In the military context, this gives rise to a debate as to whether such robots should be allowed to execute such missions, especially if there is a possibility that any human life could be at stake.

- To better understand the issues at stake, this paper presents a framework explaining the current state of the art for AI, the strengths and weaknesses of the technology, and what the future likely holds. The framework demonstrates that while computers and AI can be superior to humans in some skill- and rule-based tasks, under situations that require judgment and knowledge, in the presence of significant uncertainty, humans are superior to computers.

- In the complex discussion of if and how the development of autonomous weapons should be controlled, the rapidly expanding commercial market for both air and ground autonomous systems must be given full consideration. Banning an autonomous technology for military use may not be practical given that derivative or superior technologies could well be available in the commercial sector.

- A metaphorical arms race is in progress in the commercial sphere of autonomous systems development, and this shift in R&D effort and expenditure from military to commercial settings is problematic. Military autonomous systems development has been slow and incremental at best, and pales in comparison with the advances made in commercial autonomous systems such as drones, and especially in driverless cars.

- In a hotly competitive market for highly skilled roboticists and related engineers across the sectors most interested in AI, aerospace and defence, where funding is far outmatched by that of the commercial automotive or information and communication sectors, is less appealing to the most able personnel. As a result, the global defence industry is falling behind its commercial counterparts in terms of technology innovation, with the gap only widening as the best and brightest engineers move to the commercial sphere.

- As regards the future of warfare as it is linked to AI, the present large disparity in commercial versus military R&D spending on autonomous systems development could have a cascading effect on the types and quality of autonomy that are eventually incorporated into military systems. One critical issue in this regard is whether defence companies will have the capacity to develop and test safe and controllable autonomous systems, especially those that fire weapons.

- Fielding nascent technologies without comprehensive testing could put both military personnel and civilians at undue risk. However, the rapid development of commercial autonomous systems could normalize the acceptance of autonomous systems for the military and the public, and this could encourage state militaries to fund the development of such systems at a level that better matches investment in manned systems.

Introduction

This is a draft of the author's contribution to a forthcoming Chatham House report on artificial intelligence, to be published in the autumn of 2017.

The rise in the use of unmanned aerial vehicles (UAVs) – commonly known as drones – in both military and commercial settings has been accompanied by a heated debate as to whether there should be an outright ban on what some label 'killer robots' (e.g. Future of Life Institute, 2015; Human Rights Watch, 2013; Human Rights Watch and International Human Rights Clinic, 2016). Such robots, which could be in the air, on the ground, or in and under water, theoretically incorporate 'artificial intelligence' (AI) that would make them capable of executing missions on their own. The debate, which has many dimensions and stakeholders, concerns whether artificially intelligent machines should be allowed to execute such military missions, especially if there is a possibility that any human life could be at stake.

Given the complexity of the matter, a working definition of AI is needed. There is no one commonly agreed definition, even among computer scientists and engineers, but a general definition of AI is the capability of a computer system to perform tasks that normally require human intelligence, such as visual perception, speech recognition and decision-making. This definition is, however, inherently oversimplified, since what constitutes intelligent behaviour is also open to debate. Arguably, by this definition a house thermostat is intelligent because it can perceive and adjust the temperature. This is substantially different from AI whereby a UAV selects and engages targets without meaningful human control, which is the common assumption for autonomous weapons.

Another critical factor to consider in the debate over autonomous weapons is the increasing inability to disambiguate commercial drone autonomy from that of military UAVs. Indeed, with the rapidly expanding commercial market for both air and ground autonomous systems, there is evidence of some shifting in AI expertise from military to commercial enterprises. As a result, banning an autonomous technology for military use may not be practical given that derivative or superior technologies could well be available in the commercial sector. In addition, the asymmetrical development of the commercial autonomous system market would be likely to result in a lack of expertise for governments and militaries, which could lead to compromised and unsafe autonomous systems, both fully and semi-autonomous.

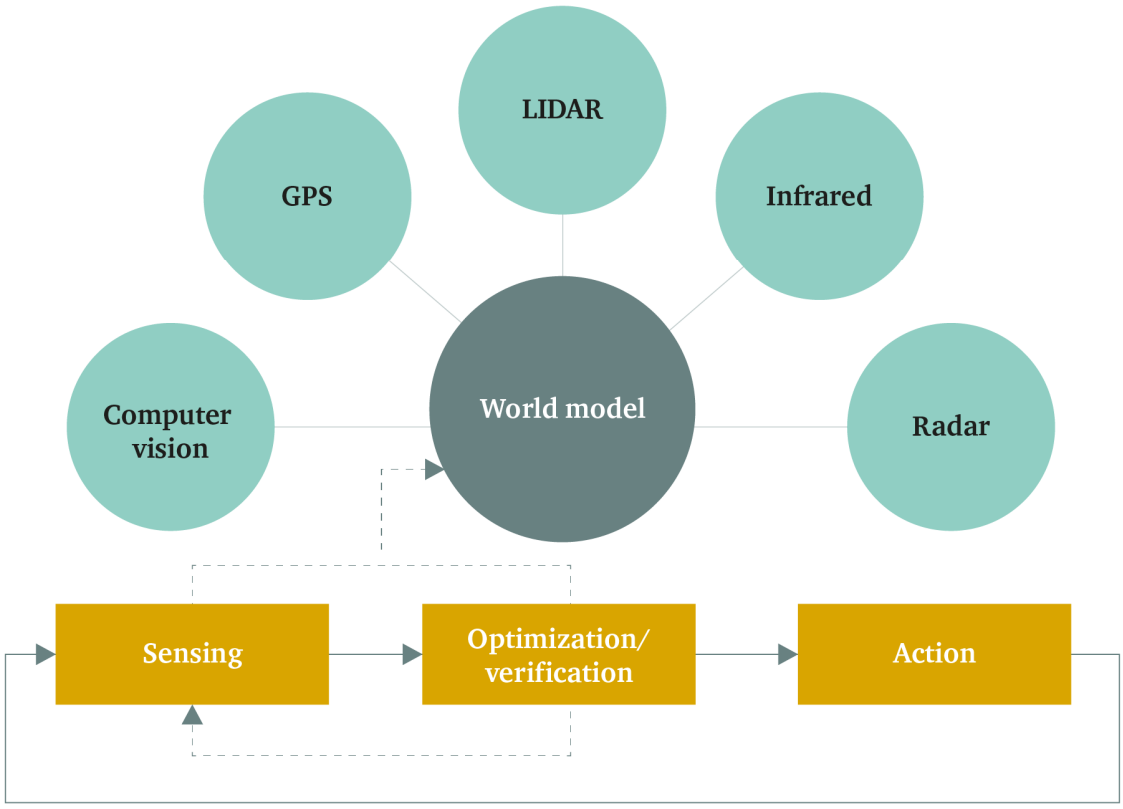
This paper presents first a framework explaining the current state of the art for AI, the strengths and weaknesses of AI, and what the future likely holds. Given that the advancement of AI is inextricably linked to the expertise of the engineers developing these systems, the case is then made that the shift in expertise from the military to the commercial sector will further complicate policy discussions on autonomous weapon, and will make it difficult for governments to deploy and manage these systems.

How robots think

To better understand the nuances of AI, it is important first to understand the difference between an automated and an autonomous system. An automated system is one in which a computer reasons by a clear if–then–else, rule-based structure, and does so deterministically, meaning that for each input the system output will always be the same (except if something fails). An autonomous system is one that reasons probabilistically given a set of inputs, meaning that it makes guesses about best possible courses of action given sensor data input. Unlike automated systems, when given the same input autonomous systems will not necessarily produce the exact same behaviour every time; rather, such systems will produce a range of behaviours.

Human intelligence generally follows a sequence known as the perception–cognition–action information processing loop, in that individuals perceive something in the world around them, think about what to do, and then, once they have weighed up the options, make a decision to act. AI is programmed to do something similar, in that a computer senses the world around it, and then processes the incoming information through optimization and verification algorithms, with a choice of action made in a fashion similar to that of humans. Figure 1 illustrates how an autonomous system embedded with AI 'thinks' and makes decisions in this way.

Figure 1: How AI of an autonomous system works



Source: Adapted from Hutchins, Cummings, Draper and Hughes (2015).

While there are many parallels between human intelligence and AI, there are stark differences too. Every autonomous system that interacts in a dynamic environment must construct a world model and continually update that model (as shown in Figure 1). This means that the world must be perceived (or sensed through cameras, microphones and/or tactile sensors) and then reconstructed in such a way that the computer 'brain' has an effective and updated model of the world it is in before it can make decisions. The fidelity of the world model and the timeliness of its updates are the keys to an effective autonomous system.

Autonomous UAV navigation, for example, is relatively straightforward, since the world model according to which it operates consists simply of maps that indicate preferred routes, height obstacles and no-fly zones. Radars augment this model in real time by indicating which altitudes are clear of obstacles. GPS coordinates convey to the UAV where it needs to go, with the overarching goal of the GPS coordinate plan being not to take the aircraft into a no-fly zone or cause it to collide with an obstacle.

In comparison, navigation for driverless cars is much more difficult. Cars not only need similar mapping abilities, but they must also understand where all nearby vehicles, pedestrians and cyclists are, and where all these are going in the next few seconds. Driverless cars (and some drones) do this through a combination of sensors like LIDAR (Light Detection And Ranging), traditional radars, and stereoscopic computer vision (see Figure 1). Thus the world model of a driverless car is much more advanced than that of a typical UAV, reflecting the complexity of the operating environment. A driverless car computer is required to track all the dynamics of all nearby vehicles and obstacles, constantly compute all possible points of intersection, and then estimate how it thinks traffic is going to behave in order to make a decision to act.

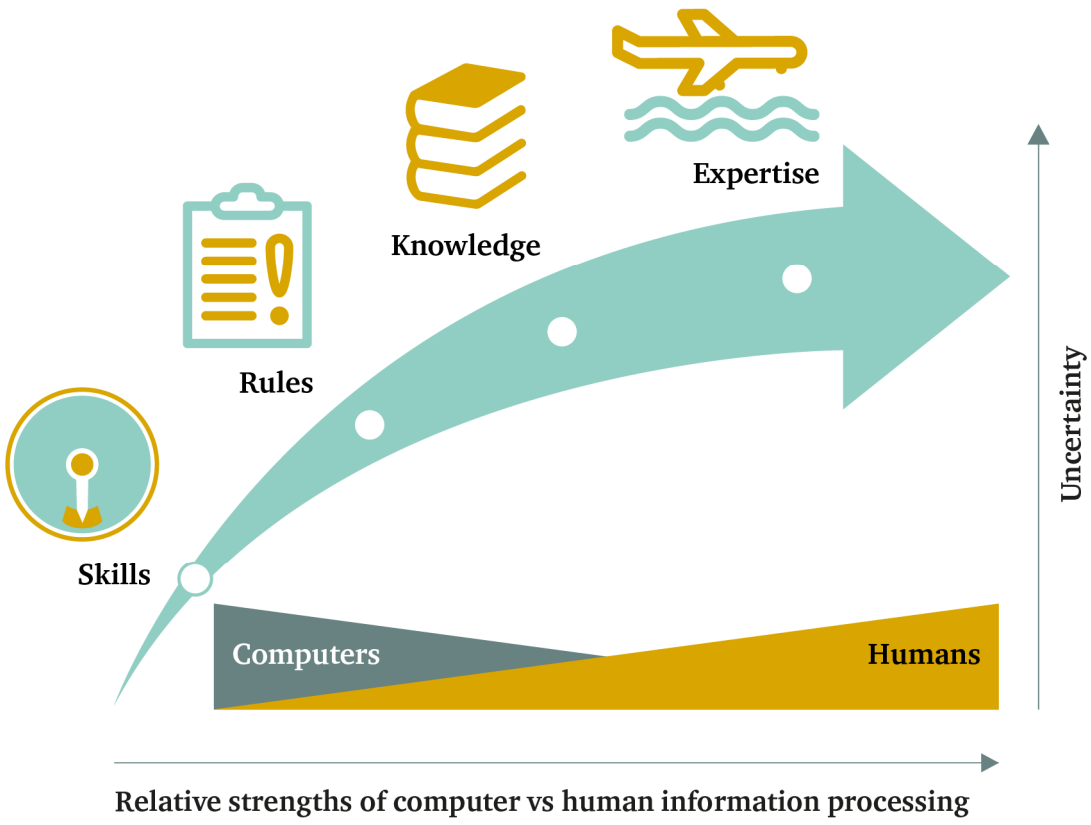
Indeed, this form of estimating or guessing what other drivers will do is a key component of how humans drive, but humans do this with little cognitive effort. It takes a computer significant computation power to keep track of all these variables while also trying to maintain and update its current world model. Given this immense problem of computation, in order to maintain safe execution times for action a driverless car will make best guesses based on probabilistic distributions. In effect, therefore, the car is guessing which path or action is best, given some sort of confidence interval. The best operating conditions for autonomous systems are those that promote a high-fidelity world model with low environment uncertainty, a concept that will be further discussed in the next section.

Balancing tasks between humans and robots

Given this understanding of the basics of autonomous robot reasoning, how should we think about the design of autonomous systems, particularly in terms of the extent to which humans should be involved? It is important first to understand when systems can and should be supervised by humans. This is a decision involving clear technical questions (such as whether a computer vision system can generate an image of sufficient resolution to make an accurate decision) as well as ethical and policy considerations (such as whether a robot should be allowed to take the life of a human being). Detailed understanding of the construction and capabilities of such military AI systems is needed in order to form cogent arguments around this polarizing issue.

Figure 2 depicts the stages of reasoning that any agent must possess in order to deal with increasingly complex decision-making scenarios, including those of autonomous weapons operation. This is an extension of Rasmussen's SRK (skills, rules and knowledge-based behaviours) taxonomy (Rasmussen, 1983), modified to explicitly represent expertise and uncertainty.

Figure 2: The relationship between uncertainty and skill, rule, knowledge and expert reasoning



Source: Adapted from Cummings (2014).

Skills-based behaviours are sensory-motor actions that for humans become highly automatic after some period of training. Such skills rely on a tight coupling of the perception–cognition–action loop, which effectively means that actions must typically come within seconds of a stimulus. An example of a skills-based activity is flying an aircraft. In training, human pilots spend the bulk of their time learning to interpret dials and gauges and thereby adjust the aircraft controls appropriately quickly to make sure the actual state of the aircraft matches the intended state. Automated agents do this as well, through an equations-based feedback loop that relies on the quality of the input data from the sensors that tell the computer where the aircraft is and where it should be.

As the cognitive continuum increases in complexity (depicted by the arrow in Figure 2), the need for rule-based behaviours arises. Rule-based behaviours are those actions that can be represented by

subroutines, stored rules or procedures. Checklists are common cognitive-aiding tools for humans executing rules. To continue with the example of aviation, pilots rely significantly on procedures to help them manage the complexity of various tasks. For instance, when a fire-light illuminates or another subsystem indicates a problem, pilots are trained to first stabilize the aircraft (a skill) but then turn to the manual to determine the correct procedure (rule following). Such codified procedures are necessary since there are far too many solutions to possible problems to be remembered. Some interpretation of procedures is required in many scenarios, especially as uncertainty and complexity increases, and this is particularly common in cases of multiple and compound problems.

Knowledge-based reasoning occurs when an established set of rules does not necessarily match the current situation, and fast mental simulations could be needed for resolution. Mental models – cognitive representations of the external world – are built over time and assist in the formulation and selection of plans, particularly in the face of uncertainty. The so-called 'miracle on the Hudson', the landing of US Airways Flight 1549 on the Hudson River in 2009, is an example of a knowledge-based behaviour depicted in Figure 2: during an emergency, the flight captain had to decide whether to ditch the aircraft or attempt to land it at a nearby airport. This is the quintessential knowledge-based scenario, reflecting a high degree of uncertainty that required the captain to develop a mental model of the environment and the state of the aircraft. The resultant fast mental simulation meant that he chose the ditching option, with clear success.[1] At the time, no autopilot system had the capability to respond in such a manner (as remains the case, although there is active research in this area).

Expert behaviours sit at the top of the reasoning behaviours, which build on knowledge-based reasoning. Expertise leverages judgment and intuition as well as the quick assessment of a situation, especially in a time-critical environment such as weapons release. Experts typically make difficult decisions in a fast and frugal manner, since comparing all possible plan alternatives is time-intensive – particularly in the face of uncertainty (Gigerenzer, Todd and ABC Research Group, 1999). In humans, the ability to cope with the highest situations of uncertainty is one of the hallmarks of a true expert, but in comparison such behaviour is very difficult for computers to replicate.

As depicted in Figure 2, skill-based tasks are easiest to automate, since they are by definition highly repetitive with inherent feedback loops that can be controlled through mathematical representations. A critical assumption, nonetheless, is that the appropriate sensors are in place. Given the if–then–else structure of rule-based behaviours, these are also potentially good candidates for automation. However, as shown in Figure 2, as uncertainty increases, rule-based reasoning gives way to knowledge-based reasoning, requiring effective management of uncertainty and true expertise.

[1] In 2009 US Airways Flight 1549 suffered a dual engine failure after take-off from New York's LaGuardia airport due to ingestion of birds in both engines. Unable to reach any prepared landing strip for an emergency landing, the pilot, Capt. Chesley Sullenberger, elected to ditch the aircraft into the Hudson River. There were no fatalities, and all 155 passengers and crew were able to evacuate the aircraft; all were subsequently rescued by boat.

Navigation, for example, is very rule-based, in that given a goal state (the destination), the best path can be objectively defined given the rules about traffic flow and knowledge of vehicle dynamics. All the same, uncertainty in such a domain can make the introduction of autonomy difficult. As already noted, navigation for drones is relatively straightforward, since mature sensor capabilities and a low-density obstacle environment mean that uncertainty is low. In the car-driving scenario, by contrast, uncertainty is much higher because the sensors are not as reliable as they are in aircraft, and the potential obstacle field is significantly more dense.

It is at the rule-based level of reasoning where the shift between needing automated versus autonomous behaviours starts to become evident. Some higher-level reasoning begins here, but the uncertainty also starts to grow as well – especially in the presence of an incomplete rule set. For instance, the Global Hawk military UAV works at a rule-based level whereby it is able to land itself if it loses communication, but it has not yet been demonstrated that such an aircraft can reason under all the situations it might encounter. The latter would require a higher level of reasoning.

Knowledge-based behaviours and resulting expertise represent the most advanced forms of cognitive reasoning that typically occur in domains where uncertainty is the highest, as shown in Figure 2. Rule-based reasoning may assist decision-makers (human and/or computer) in determining possible courses of action. Where there is high uncertainty, however, it is often difficult to understand which set of rules applies. It is in these uncertain scenarios – which are by definition vague and ambiguous – that algorithms may not be able to understand the solution space, much less achieve a feasible solution.

The key question for any autonomous system engaged in a safety-critical task (such as weapons release) is whether that system can resolve ambiguity in order to achieve acceptable outcomes. It is conceivable that an autonomous drone could be given a mission to hit a static target on a military installation with a high probability of success. Indeed, many countries have missiles that can do just that. However, could an autonomous drone targeting an individual understand from its real-time imagery that a specific person has been found and that releasing a weapon will kill only that person and no innocent bystanders? Currently, the answer to this question is no.

The power of human induction – i.e. the ability to form general rules from specific pieces of information – is critical in a situation that requires both visual and moral judgment and reasoning. For humans, induction that drives such judgments is necessary to combat uncertainty. Computer algorithms – especially those that are data-driven like typical algorithms that fall under the category of AI – are inherently brittle, which means that such algorithms cannot generalize and can only consider the quantifiable variables identified early on in the design stages when the algorithms are originally coded (Smith, McCoy and Layton, 1997). Replicating the intangible concept of intuition, knowledge-based reasoning and true expertise is, for now, beyond the realm of computers. There is significant research currently under way to change this, particularly in the machine learning/AI community, but progress is slow.

IBM's Watson, composed of 90 servers, each with a 3.5 GHz core processor (Deedrick, 2011), is widely referenced as a computer that has the capacity to outmatch human reasoning abilities. As such, it is often popularly stated to be an 'intelligent' computer. However, people confuse the ability of a computer – which has been tuned by humans for a highly specific task of searching vast databases and generating formulaic responses – with an entity that exhibits knowledge. Watson

leverages AI through natural language processing and pattern matching, but it is operating in environments in which uncertainty is low.

Some claim that machine learning and deep learning approximate human intelligence, but at present these tools basically detect patterns that are significantly tuned by humans and must be interpreted by humans to be useful. As a result, the advances that they represent are evolutionary and not revolutionary. One critical limitation of machine learning is that, as a data-driven approach, it fundamentally relies on the quality of the underlying data and thus can be very brittle. There are no guarantees that a computer leveraging machine learning algorithms can detect a pattern or event never previously encountered, or even scenarios that are only slightly different. As uncertainty grows, therefore, these tools become less useful.

For example, a highly publicized machine learning algorithm has been able to identify objects (e.g. dog, cat, truck) in images with 15.8 per cent accuracy in a world containing some 22,000 object categories (Le et al., 2012). When that world is collapsed into 1,000 categories of objects, other algorithms can achieve up to 60–70 per cent accuracy (Sermanet et al., 2014). In both of these studies, 10 million labelled images were required to 'train' the algorithms. In comparison, humans need far fewer examples to 'learn' from, and have the ability to distinguish and accurately name far more than 22,000 individual objects.

So while computer vision algorithms may be able to distinguish a car from a bicycle at close range for the purposes of driverless car navigation, such algorithms are far from perfect. Moreover, it is multiple orders of magnitude more difficult for a computer to identify a specific truck carrying specific people with enough certainty to launch a weapon. So while UAVs are excellent platforms for obtaining an image, letting the UAV or any autonomous system decide whether a specific target meets the criteria for weapons release is still several years away.

The inability of computers to identify specific targets with a high degree of certainty highlights the probabilistic nature of knowledge-based reasoning. Whether humans or computers do it, both are guessing with incomplete information based on prior probabilities about an outcome. While the consequences of an autonomous central heating thermostat wrongly guessing a homeowner's arrival time are relatively trivial, the same cannot be said of an autonomous weapon making an inaccurate guess in its particular operating environment.

The big picture

The future of AI in military systems is directly tied to the ability of engineers to design autonomous systems that demonstrate independent capacity for knowledge- and expert-based reasoning as illustrated in Figure 2. There are no such autonomous systems currently in operation. Most ground robots are teleoperated, essentially meaning that a human is still directly controlling a robot from some distance away as though via a virtual extension cord. Most military UAVs are only slightly more sophisticated: they have some low-level autonomy that allows them to navigate, and in some cases land, without human intervention, but almost all require significant human intervention to execute their missions. Even those that take off, fly over a target to capture images, and then return home still operate at an automated and not autonomous level, and do not reason on the fly as true autonomous systems would.

While current operational systems are more automatic than autonomous, there are significant global efforts in the research and development (R&D) of autonomous systems. Incremental progress in such military system development is occurring in many countries in air, ground, on-water and underwater vehicles with varying degrees of success. Several types of autonomous helicopter that can be directed with a smartphone by a soldier in the field are in development in the US, in Europe and in China. Autonomous ground vehicles such as tanks and transport vehicles are in development worldwide, as are autonomous underwater vehicles. In almost all cases, however, the agencies developing these technologies are struggling to make the leap from development to operational implementation.

There are many reasons for the lack of success in bringing these technologies to maturity, including cost and unforeseen technical issues, but equally problematic are organizational and cultural barriers. The US has, for instance, struggled to bring autonomous UAVs to operational status, primarily as a result of organizational in-fighting and prioritization in favour of manned aircraft (Spinetta and Cummings, 2012). For example, despite the fact that the F-22 aircraft has experienced significant technical problems and has flown little in combat, the US Air Force is considering restarting the F-22 production line – in itself an extremely costly option – as opposed to investing in more drone acquisitions. Beyond the production line, moreover, the hourly operational cost of the F-22 is $68,362, as compared with the Predator's $3,679 (Thompson, 2013); the latter can perform most of the same core functions of an F-22 save for air-to-air combat missions, which the F-22 itself could not previously perform due to technical problems.

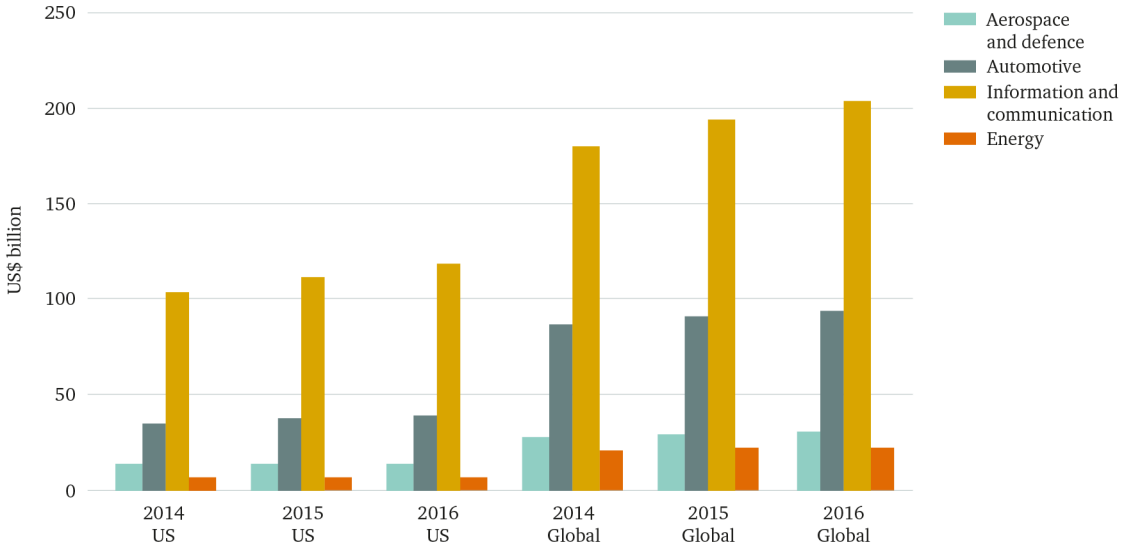
To give another example, the US Navy's X-47 was intended to be developed as an autonomous fighter/bomber aircraft, but despite many successful sea trials it is now slated to operate as a tanker for aerial refuelling – a far cry from its original (achievable) purpose. Both the US Air Force and Navy have chosen to use the vast majority of aircraft acquisition funds for the manned F-35 Joint Strike Fighter, even though the programme has been beset with management and engineering problems, and whose relevance is contested particularly in light of advancing autonomous technologies (Hendrix, 2015). For many in the military, UAVs are acceptable only in a support role, but they threaten the status quo if allowed to take the most prestigious, 'tip-of-the-spear' jobs.

There are, however, other organizational issues limiting the operational implementation of autonomous systems, and one that is increasingly problematic is the shift in advanced development from military to commercial settings. A metaphorical arms race is in progress in the commercial sphere of autonomous systems development. Military autonomous systems development has been slow and incremental at best, and pales in comparison with the advances made in commercial autonomous systems such as drones, and especially in driverless cars.

Driverless car development originated with a Defense Advanced Research Projects Agency (DARPA) programme in 2004. When the programme ended in 2007, driverless cars could move only slowly through closed courses, and not without accidents. A decade later, industry is on the verge of commercializing driverless cars around the world. This rapid progress is a result of the significant industry-sponsored R&D investment, as well as competition for the multi-billion-dollar automotive consumer market. Meanwhile – and paradoxically, given the origins of the technology – there has been very little progress in military autonomous vehicle development.

The inability of the military to advance its autonomy programmes, not only on the ground but also in the air and in other domains, is evidently linked to the growth in autonomous systems in the commercial market, particularly driverless cars. Figure 3 shows R&D spending in the three-year period 2014–16 in three key sectors: aerospace and defence; automotive; and information and communication. These sectors are core to the development of autonomous systems, and so tracking spending there gives insight into the speed and extent of innovation.

Figure 3: R&D spending by sector, 2014–16 (US$ billion)



Source: Industrial Research Institute (2016).

The aerospace and defence sector is responsible for the bulk of the development of military autonomous systems. However, as shown in Figure 3, R&D spending is far below that of the other two sectors (only around 15 per cent of the global information and communication sector, for example). In the US, moreover, spending has actually been declining. Autonomous systems development is not a priority for the US defence industry, and competes for investment with development of traditional platforms like narrowly capable manned fighter aircraft (e.g. the F-35) and extremely costly laser weapons. The outcome is that only a very small proportion of defence R&D money is invested in autonomous military systems.

In contrast, not only is R&D spending in the automotive sector three times that of the aerospace and defence industry, but it is growing in both the US and the global markets. A key factor in spending in this sector is the development of driverless car technology (Industrial Research Institute, 2016) – as reflected in the intensely competitive market for start-ups such as Cruise Automation, bought by General Motors for more than $1 billion in 2016.

The information and communication sector is another critical stakeholder in the development of autonomous systems. Industries within the sector specialize in the development of software, including machine learning and AI that are core capabilities for these advanced systems. To give some examples, X (formerly known as Google X) has both drone and driverless car research

programmes; Facebook and Amazon have drone development programmes; and Apple is thought to have an autonomous car development project. Given that the information technology companies are spending far more on R&D than does the aerospace and defence sector (Figure 3), it is clear why there has been far greater progress in commercial autonomous systems development.

As regards the future of warfare as it is linked to AI, the large disparity in commercial versus military autonomous R&D spending could have a cascading effect on the types and quality of autonomy that are eventually incorporated into military systems. One critical issue in this regard is whether defence companies will have the capacity to develop and test safe and controllable autonomous systems, especially those that fire weapons.

Engineers who design such systems must have strong hardware as well as software expertise. However, there are a limited number of universities graduating students in robotics, controls, mechatronics and related fields with the technical prowess for such jobs, and so there is fierce competition for highly qualified personnel. This demand is what drives $1 billion start-up acquisitions, as well as buy-outs of university research groups such as that by Uber, in 2015, of 40 highly skilled roboticists from the National Robotics Engineering Center at Carnegie Mellon University. Boston Dynamics, at one time the leading US military R&D robotics company, was bought by Google in 2013 for domestic robot development, and was reported by a number of sources in 2016 to be a target for acquisition by Toyota for its driverless car development programme (e.g. Brian, M., 2016).

With such a hotly competitive market for roboticists and related engineers across these industries, the aerospace and defence sector, where funding lags behind, is less appealing to the most able personnel. As a result, the global defence industry is falling behind its commercial counterparts in terms of technology innovation, with the gap only widening as the best and brightest engineers move to the commercial sphere. This comparative lack of expertise means that military autonomous systems eventually fielded could be deficient, or lacking in appropriate safeguards and testing. So while the debate over whether autonomous weapons should be banned is clearly an important one, a more immediate issue is the ability of defence industries to field safe semi-autonomous systems, much less fully autonomous ones.

While some may say that the current distribution of both R&D and expertise is the inevitable outcome of competition and a free market, this does not fully acknowledge the reality of a fundamental shift in technology prowess whereby militaries will start to significantly lag in autonomous system capabilities as compared with commercial systems. The average American is more likely to have a driverless vehicle before soldiers on the battlefield do, and terrorists will potentially be able to buy drones on the internet with as much or greater capability than those available to the military.

This imbalance in technology access will no doubt introduce new unforeseen and disruptive dynamics for military operations. For example, if defence companies and governments continue down a path of relative AI illiteracy, could this enable a potential power shift such that critical AI services will be leased via Google, Amazon or Facebook? Google has long distanced itself from military contracts, while also acquiring highly advanced robotics companies and letting these companies' pre-existing military contracts expire. If militaries are relegated to buying robots and AI services such as image analysis from commercial off-the-shelf suppliers, this would undoubtedly

affect military readiness in both the short and the long term.

The gap between historical military superiority in UAV development and the present capabilities of the commercial sector is closing, as evidenced by the increasing number of military-grade drones offered for sale via the internet. Footage showing weaponization of drones, once thought to be the exclusive domain of established militaries, can now be regularly be found on YouTube. It is unlikely that small entrepreneurial companies will produce drones that could compete with highly advanced UAVs such as the US Air Force's Predator, but there will certainly be competition from such companies in terms of developments in surveillance and communication. And while the US military's advanced research arm, DARPA, has had difficulty in fielding a drone that can transport troops, China's EHang has purportedly built a commercial drone that will transport passengers, and several commercial companies are developing versions of autonomous flying cars for general public use (e.g. Forum, 2016).

Given the current extent of commercial development of drones and other robotic systems, there are other important considerations such as the possible latent consequences of companies and countries that rush AI technologies to market – as against nation states that tend to take more conservative approaches. Fielding nascent technologies without comprehensive testing could put both military personnel and civilians at undue risk. However, the rapid development of commercial autonomous systems could normalize the acceptance of autonomous systems for the military and the public, and this could encourage state militaries to fund the development of such systems at a level that better matches investment in manned systems. Meanwhile, it remains unclear how the rise of autonomous drones for civilian use could influence popular attitudes and perceptions concerning autonomous military platforms, including weapons.

The thorny path ahead

Although it is not in doubt that AI is going to be part of the future of militaries around the world, the landscape is changing quickly and in potentially disruptive ways. AI is advancing, but given the current struggle to imbue computers with true knowledge and expert-based behaviours, as well as limitations in perception sensors, it will be many years before AI will be able to approximate human intelligence in high-uncertainty settings – as epitomized by the fog of war.

Given the present inability of AI to reason in such high-stakes settings, it is understandable that many people want to ban autonomous weapons, but the complexity of the field means that prohibition must be carefully scoped. Fundamentally, for instance, does the term autonomous weapon describe the actual weapon – i.e. a missile on a drone – or the drone itself? Autonomous guidance systems for missiles on drones will likely be strikingly similar to those that deliver packages, so banning one could affect the other. And how will technologies be treated that emerge from the growing commercial market, which is expected to leapfrog some aspects of military capability and possibly change public perception?

The impact of the rapid expansion of the commercial market on autonomous systems development cannot be overstated, and an even bigger problem in the short term is how to fully understand the global implications of the discernible shift in the power base of AI expertise from the military to commercial enterprises. Machines, computers and robots are getting 'smarter' primarily because

roboticists and related engineers are getting smarter, so this relatively small group of expert humans is becoming a critical commodity. Universities have been slow to respond to this demand, and governments and industry have also lagged behind in providing scholarship mechanisms to incentivize students in the field of AI.

Ultimately, the growth in the commercial information technology and automotive sectors, in terms of both attracting top talent and expanding autonomous systems capabilities in everyday commercial products, could be a double-edged sword that will undoubtedly affect militaries around the world in as yet imagined ways.

References

Brian, M. (2016), 'Toyota is the top bidder for robotics pioneer Boston Dynamics', Engadget, 1 June 2016, https://www.engadget.com/2016/06/01/toyota-alphabet-boston-dynamics.

Cummings, M. L. (2014), 'Man vs. Machine or Man + Machine?' *IEEE Intelligent Systems*, 29(5), pp. 62–69.

Deedrick, T. (2011), 'It's Technical, Dear Watson', *IBM Systems Magazine*, http://www.ibmsystemsmag.com/ibmi/trends/whatsnew/It%E2%80%99s-Technical,-Dear-Watson.

Forum (2016), 'Future of urban mobility: My kind of flyover', Airbus Forum, http://www.airbusgroup.com/int/en/news-media/corporate-magazine/Forum-88/My-Kind-Of-Flyover.html.

Future of Life Institute (2015), 'Autonomous Weapons: An Open Letter from AI & Robotics Researchers', http://futureoflife.org/open-letter-autonomous-weapons.

Gigerenzer, G., Todd, P. M. and ABC Research Group (1999), *Simple Heuristics That Make Us Smart*, Oxford and New York: Oxford University Press.

Hendrix, J. (2015), *Retreat from Range: The Rise and Fall of Carrier Aviation*, Washington, DC: Center for a New American Security.

Human Rights Watch (2013), 'Arms: New Campaign to Stop Killer Robots', https://www.hrw.org/news/2013/04/23/arms-new-campaign-stop-killer-robots.

Human Rights Watch and International Human Rights Clinic (2016), 'Killer Robots and the Concept of Meaningful Human Control: Memorandum to Convention on Conventional Weapons (CCW) Delegates', https://www.hrw.org/news/2016/04/11/killer-robots-and-concept-meaningful-human-control.

Hutchins, A. R., Cummings, M. L., Draper, M. and Hughes, T. (2015), 'Representing Autonomous Systems' Self-Confidence through Competency Boundaries', paper presented at the 59th Annual Meeting of the Human Factors and Ergonomics Society, Los Angeles, CA, 26–30 October 2015.

Industrial Research Institute (2016), '2016 Global R&D Funding Forecast', supplement to *R&D Magazine*, Winter 2016.

Le, Q. V., Ranzato, M. A., Monga, R., Devin, M., Chen, K., Corrado, G. S., Dean, J. and Ng, A. Y. (2012), 'Building High-level Features Using Large Scale Unsupervised Learning', paper presented at the 29th International Conference on Machine Learning, Edinburgh, 26 June–1 July 2012.

Rasmussen, J. (1983), 'Skills, rules, and knowledge: Signals, signs, and symbols, and other distinctions in human performance models', *IEEE Transactions on Systems, Man, and Cybernetics,* 13(3), pp. 257–266.

Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R. and LeCun, Y. (2014), 'OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks', paper presented at the International Conference on Learning Representations, Banff, AB, 14–16 April 2016.

Smith, P. J., McCoy, C. E. and Layton, C. (1997), 'Brittleness in the design of cooperative problem-solving systems: The effects on user performance', *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, 27(3), pp. 360–371.

Spinetta, L. and Cummings, M. L. (2012), 'Unloved Aerial Vehicles: Gutting its UAV plan, the Air Force sets a course for irrelevance, *Armed Forces Journal*, November 2012, pp. 8–12.

Thompson, M. (2013), Costly Flight Hours, *Time*, 2 April 2013.

About the author

Mary 'Missy' L. Cummings is the director of the Humans and Autonomy Laboratory at Duke University. She received her bachelor's degree in mathematics from the US Naval Academy in 1988, her master's in space systems engineering from the Naval Postgraduate School in 1994, and her PhD in systems engineering from the University of Virginia in 2004. A naval officer and military pilot from 1988–99, she was one of the US Navy's first female fighter pilots. She is currently a professor in the Duke University Department of Mechanical Engineering and Materials Science, the Duke Institute of Brain Sciences, and the Duke Electrical and Computer Engineering Department. She is also a member of the US Department of Transportation's advisory committee on autonomous transportation, and co-chair for the World Economic Forum's Global Future Council on Artificial Intelligence and Robotics. Her research interests include human supervisory control, human–unmanned vehicle interaction, human–autonomous systems collaboration, human–robot interaction, human systems engineering, and the ethical and social impact of technology.

Acknowledgments

The author would like to thank the many sponsors of her research that helped to inform this work, including the National Science Foundation, the Office of Naval Research, NASA, and the US Army and Air Force.

Independent thinking since 1920

ISBN 978 1 78413 198 2

This publication is printed on recycled paper.

The Royal Institute of International Affairs
Chatham House
10 St James's Square, London SW1Y 4LE
T +44 (0)20 7957 5700 F +44 (0)20 7957 5710
contact@chathamhouse.org www.chathamhouse.org
Charity Registration Number: 208223