NATIONAL SECURITY IMPLICATIONS OF INCREASINGLY AUTONOMOUS TECHNOLOGIES: LEGAL AMBIGUITY, CHALLENGES IN CONTROLLING THIS SPACE, PUBLIC/PRIVATE SECTOR DYNAMICS AND ETHICAL CONCERNS

PART 2

Policy Report
February 2015

Caitríona H. Heinl

RSiS
Nanyang Technological University
S. RAJARATNAM SCHOOL OF INTERNATIONAL STUDIES

Policy Report

NATIONAL SECURITY IMPLICATIONS OF INCREASINGLY AUTONOMOUS TECHNOLOGIES: LEGAL AMBIGUITY, CHALLENGES IN CONTROLLING THIS SPACE, PUBLIC/ PRIVATE SECTOR DYNAMICS AND ETHICAL CONCERNS

Part 2

Caitríona H. Heinl
February 2015

Centre of Excellence for National Security (CENS),
S. Rajaratnam School of International Studies (RSIS),
Nanyang Technological University

Executive Summary

This is the second of a two-part report that highlights the mounting importance for the national security agenda of technologies that are becoming increasingly autonomous, or becoming gradually more independent of human control in other words. At present, it is still relatively unclear how maturing autonomous technologies, including potentially fully autonomous and lethal systems, might impact national security exactly in terms of military and economic implications, or possible misuse by criminals. This two-part report finds that many questions still remain unaddressed and that there are several significant policy gaps that should be further analysed.

While some aspects of this area are still in their infancy, the full report aims to identify the key questions that are beginning to emerge. It also highlights the salient aspects of several discussions that have been recently initiated and will impact national security. Thus far, as a United Nations Institute for Disarmament Research (UNIDIR) report of March 2014 notes, there has been a lack of critical analysis on how the proliferation of increasingly autonomous systems might alter regional security dynamics.[1] China, for instance, recently became the largest buyer of industrial robots, overtaking Japan for the first time with an approximately 60 per cent increase in a one-year period from 2012 to 2013.[2] And while scientists, ethicists, and futurists, amongst others, have hotly debated several gaps marked within the report in the past, wider policy circles are only recently beginning to seriously consider these questions to the same extent. This two-part report argues that these issues now require deeper consideration, and it is an opportune time to shape the strategic debate.

The United Nations Special Rapporteur on extrajudicial, summary or arbitrary executions, Christof Heyns, recently explained that while the technology for drones is already in use and discussions are now being held on their regulation, autonomous robotics presents a unique situation since the technology is not actively used yet.[3] This therefore presents some unique challenges, which are addressed throughout both parts of the report. The first part of the report discusses the nature of maturing autonomous technologies and the significance of potential lethality. It finds that, although there is an increasing military interest in this area, a clear understanding of the nature of these technologies is still lacking in the policy community.

The first part of the report then provides an outline of several broader military implications as well as cyber-related implications that could arise in this area. It is likely that states will pursue technological superiority via increasingly autonomous technologies for both economic and military reasons. Yet, deeper analysis of the long-term implications is needed in terms of possible military advantages and disadvantages that might ensue, including the role of the human vis-à-vis the machine, as well as how military interest in autonomy might evolve globally. Given geopolitical uncertainties in the Asia Pacific region, such developments could also be significant if states seek technological superiority with autonomous technologies.

This second part of the report analyses the challenges of controlling and regulating this space. While various stakeholders have made numerous recommendations, there does not

---

1 UNIDIR Resources, "Framing Discussions on the Weaponization of Increasingly Autonomous Technologies", March 2014, 8.

2 Tanya Powley, "China becomes largest buyer of industrial robots", http://www.ft.com/cms/s/0/a5cca8c0-e70c-11e3-aa93-00144feabdc0.html#axzz35X1ZGoLX, 1 June 2014.

3 Christof Heyns, Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions, "Lethal Autonomous Robotics", United Nations Institute for Disarmament Research (UNIDIR) Conference, http://www.unidir.org/programmes/security-and-society/lethal-autonomous-robotics, 23 May 2013.

seem to be a silver bullet solution at this juncture. Moreover, this part of the report finds that there are several highly significant legal ambiguities, which require clarification.

Furthermore, these technologies often have a dual-use nature – for both military application and civilian purposes, and both the public and private sectors are driving these developments by investing heavily in R&D in pursuit of their own objectives. This part of the report finds that while innovation and economic growth should not be disproportionately stifled, stronger collaboration between the public sector and industry as well as academic research laboratories is advisable to shape policies responsibly and manage unexpected developments that could perhaps be detrimental. Malicious non-state actors also add a further layer of complexity since terrorist groups, organised crime gangs, or proxy actors could possibly obtain or alter commercially available technologies.

Lastly, the final section of this part of the report finds that the ethical implications of these tools require deeper consideration, and public perception of such advanced technologies is another important factor that should be considered.

Policy Uncertainty: Challenges and Opportunities

Legal Ambiguities and Challenges in Regulating this Space

The American people seem to have the image of robots flying around semi-autonomously making their own decisions and conducting kinetic strikes without oversight by responsible human beings…[t]he law of armed conflict, the principles of war, U.S. ethics and legal bases apply no matter what the weapon.[4]

Currently, there is a high level of legal ambiguity in this area. While recommendations have been made for further analysis and examination of several questions that arise, no concrete answers are available yet.

A number of non-governmental organisations (NGOs) are already advocating their positions in this space. For instance, Human Rights Watch (HRW) calls for an international treaty to place a blanket ban on the development, sale, and use of autonomous weapons in its position paper.[5] The Campaign to Stop Killer Robots, a relatively new international campaign of 45 NGOs across 22 countries, demands a pre-emptive ban on the development, production and use of weapons capable of attacking targets without human intervention, in other words fully autonomous "human-out-of-the-loop systems".[6] It recommends that states develop national policies and that negotiations begin on a treaty to ban these weapons. It participated in the May 2014 informal meeting on lethal autonomous weapons, as did the International Committee of the Red Cross, Amnesty International, HRW, the International Committee on Robot Arms Control, and stakeholders from business, academia, and research.

However, the counter-argument is that such a ban could be morally flawed since banning increasing autonomy may prevent the development of tools that cause less harm to civilians than human combatants since it is unlikely they would have desires for revenge.[7] They can also be programmed for cases of doubt so that they only respond if fired upon, and they might be able to better identify targets than humans or respond more rapidly and accurately, thereby causing less collateral damage.[8]

Nevertheless, from a broad legal perspective, it is unclear whether lethal autonomous weapons systems comply with existing international law, in particular the principles of international humanitarian law, the 1949 Geneva Conventions, the Martens Clause and customary law.[9] Analysts argue that these systems might never be able to select and strike targets by analysing a complex situation, identifying human nuances, and using basic instincts of mercy, identification, and morality like humans.[10]

The Chairperson's report for the informal meeting of experts in May 2014 notes that legal reviews were therefore recommended, especially when developing new weapons technologies, and that more discussions would be valuable in the area of the implementation of weapons reviews, including Article 36 of Additional Protocol I to the 1949 Geneva

---

4 U.S. Department of Defense, *Military Uses Remotely Piloted Aircraft Ethically*, Press Release, http://www.defense.gov/news/newsarticle.aspx?id=122308, 22 May 2014.

5 Siboni & Eshpar, "Use of Autonomous Weapons", 80.

6 Stuart Hughes, "Campaigners call for international ban on 'killer robots'", http://www.bbc.co.uk/news/uk-22250664, 23 April 2013.

7 Siboni & Eshpar, Use of Autonomous Weapons, 82.

8 Ibid

9 Chairperson, *2014 informal Meeting of Experts on LAWS*, 4.

10 Siboni & Eshpar, "Use of Autonomous Weapons", 81.

Conventions.[11] Under this article, there may be an obligation, for those party to this protocol, to examine whether deployment would be prohibited under international law, and if it is not possible to confirm the capacity of lethal autonomous robotics' compliance with international law, then it could be deemed illegal.[12]

In accordance with the rule of distinction, questions include whether a robot can distinguish between a civilian and combatant, and if facing a combatant, what will occur in cases of possible surrender? In particular, what will occur in cases where the enemy is among civilians? However, Heyns observes that humans also make mistakes or act on anger and therefore, we should not necessarily compare these systems to the ideal, but to the human.[13]

In order to meet proportionality requirements during armed conflict, it is still uncertain whether these systems can make value-qualitative judgements, or decisions on the level of collateral damage.[14] Though some commentators suggest that while these systems might not currently comply with international humanitarian law, this may not be the case in the future and it is possible that they might even comply better than humans. For instance, while human intelligence will continue to be needed for the foreseeable future to apply context, judgement, and to account for qualitative insights, advances could allow computers to process huge data sets and portray the data in ways that mimic human intuition and judgement.[15] On the other hand, it is not yet clear whether such level of application may in fact be reached.[16]

Further legal ambiguity occurs under human rights law over the right to life, human dignity, the right to be protected against inhumane treatment, and the right to fair trial. It also seems that under *jus ad bellum*, deeper analyses are required as to whether lethal autonomous weapons systems can change the threshold of use of force.[17] Another significant question that remains unanswered concerns the responsibility and accountability gap associated with the use of these systems. It is not certain whether creating states or commanders of autonomous agents are always responsible if an agent exceeds assigned tasks and makes an unforeseen autonomous decision where creators do not know in advance the precise technique employed or system targeted.[18] Heyns has observed that in these situations, a robot cannot be held responsible, in other words the moral agent is not a primary actor, which begs the question whether it is command or state responsibility.[19] He describes this as an uncharted area that requires examination and the May 2014 meeting equally recommended further analysis on this point.[20]

Under the U.S. DoD Directive, those who authorise the use of, direct the use of, or operate these systems must do so with appropriate

---

11 Chairperson, *2014 informal Meeting of Experts on LAWS*, 4.
Article 36 states that "in the study, development, acquisition, or adoption of a new weapon, means or method of warfare, a High Contracting Party is under an obligation to determine whether its employment would, in some or all circumstances, be prohibited by this Protocol or by any other rule of international law applicable to the High Contracting Party".

12 Peter Herby, "Lethal Autonomous Robotics", United Nations Institute for Disarmament Research (UNIDIR) Conference, http://www.unidir.org/programmes/security-and-society/lethal-autonomous-robotics, 23 May 2013.

13 Heyns, "Lethal Autonomous Robotics", UNIDIR Conference.

14 Ibid

15 Work & Brimley, "War in the Robotic Age", 24.

16 Herby, "Lethal Autonomous Robotics".

17 Chairperson, *2014 informal Meeting of Experts on LAWS*, 5.

18 Guarino, "Autonomous Intelligent Agents". See also: Heinl, "Artificial (Intelligent) Agents".

19 Heyns, "Lethal Autonomous Robotics", UNIDIR Conference.

20 Chairperson, *2014 informal Meeting of Experts on LAWS*, 4.

care and in accordance with the law of war, applicable treaties, weapon system safety rules, and applicable rules of engagement. Any autonomous or semi-autonomous weapons systems intended for use in a manner that falls outside the policies in the Directive, must be approved before formal development and again before fielding. The guidelines for review of such systems include that before a decision to enter formal development is made, a preliminary legal review must be completed. However, with the exception of this legal review requirement, a request may be made for a Deputy Secretary of Defense waiver for these requirements in cases of urgent military operational need.

There are clearly several gaps in terms of legal certainty, and while agreement is unlikely in the short term, it is still worth continuing this dialogue to ensure that this area develops in a responsible way. Several stakeholder recommendations suggested so far include a blanket ban, a moratorium, control mechanisms, international treaty, and technical safeguards. Nevertheless, each approach has inherent challenges. Enforcing a blanket ban could be extremely difficult, if not impossible. Even with a freeze or moratorium, it is unlikely that development of these tools will cease where some steps are taken to delay or prevent development. Although, if the ethical and legal discussions are postponed to when the technology matures, it might be too late then to impose effective restrictions.[21]

Rather than a blanket ban and moratorium or waiting until it is too late, taking regular stock of developments at shorter intervals could possibly assist with this conundrum, given such a rapidly changing strategic and technological environment. For instance, one lesson highlighted in the U.K.'s strategic defence and security review in 2010 was the need for more frequent reviews to reassess a changing strategic environment, so that there could be better focus on decisions that were needed for a shorter period of time such as four years, and to leave those decisions that can be better taken in light of further experience and developments at the end of that period.

Even from a practical perspective, it could be difficult to regulate or control such technological developments. While developing and implementing regulatory and legislative frameworks are important, and ultimately necessary, regulations can often be untimely relative to the speed of technological developments. Moreover, regulatory solutions do not necessarily deter malicious state or non-state actors. Regulating and controlling the development of these new systems might be problematic when, like cyber capabilities, it could be difficult to prevent their development.[22] Limiting development will be difficult, if not impossible, due to countries' investments, the large role that unmanned tools and systems already play in today's battlefield, and their potential in the context of civilian uses such as in science, medicine, services, and industry.[23] Furthermore, these capabilities could affect aspects of our lives in a gradual fashion that might make it difficult to even differentiate between automation and autonomy.[24]

Likewise, agreeing to an international treaty will bring its own innate time difficulties. Several analysts subsequently conclude that the technical matters might be easier to solve than the non-technical.[25] Although enforcing technical safeguards could also be challenging, in light of malicious state and non-state actors, as well as humans' tendency to modify

---

21 Siboni & Eshpar, "Use of Autonomous Weapons", 83.

22 Heinl, "Artificial (Intelligent) Agents".

23 Antebi, "Who Will Stop the Robots", 61.

24 Siboni & Eshpar, "Use of Autonomous Weapons", 83.

25 Antebi, "Who Will Stop the Robots", 70. See also: UNIDIR, "Weaponization of Increasingly Autonomous Technologies", 8. Peter Singer quote from The Robotics Revolution, Brookings Institution, 2012.

technologies to overcome safety controls.[26] Appendix 4 provides a number of proposed technical safeguards for future consideration, in particular for intelligent software that underpins these technologies and autonomous intelligent agents.

There could even be a difficult-to-predict area between the intentions of developers and operators, and their ultimate behaviour in practice.[27] The U.S. DoD Directive addresses some of these concerns by establishing guidelines designed to minimise the probability and consequences of failures in autonomous and semi-autonomous weapons systems that could lead to unintended engagements. This applies to the design, development, acquisition, testing, fielding, and employment of autonomous and semi-autonomous weapon systems (these guidelines are outlined under Appendix 4).

There is still a significant gap between the level of understanding of those working in this field and the policy realm, which becomes especially apparent when seeking appropriate solutions for some of these dilemmas. This gap needs to be addressed and the policy formation process needs to be augmented by a deeper technical understanding of how these technologies actually function. This includes whether systems can transition from one operating state to another, such as from semi-autonomous to fully autonomous. Moreover, these technologies can often have a dual-use nature. In other words, the same systems can be used for both lethal application and civilian purposes, and technical components enabling autonomy can be similar for both military and civil applications.[28] As a result, the interplay between the public and private sector needs careful consideration when analysing the future implications of these tools.

Public-Private Sector Considerations

Defence and security reports assert that the increasingly diverse range of threats is likely to not only include sophisticated military weapons, but also greater innovative application of readily available civil technologies. Operational advantages could then be reduced, when it is easier for adversaries to buy high-technology products on the open market.[29] For now, it seems that both the public and private sectors are driving these technological developments, but it is not certain how this will continue to evolve.

As a 2012 U.K. Ministry of Defence (MoD) report explains, advanced technology development for defence and security that was "once the realm of government research organisations is now carried out almost exclusively in the civil and commercial sectors".[30] Governments may not even be able to sustain deep expertise in all areas of science and technology, and the rapid pace of innovation means that new technology can often appear faster than it may be integrated.[31] Likewise, the public sector may not always match the speed of innovation in the private sector. The U.K. MoD has also reiterated the point that civil applications, rather than defence and security applications, drive innovation in many fields.[32]

Likewise, UNIDIR's report finds that the civilian technology sector is under-appreciated so far in the context of these discussions, and that this sector is in fact developing a more far

---

26 Heinl, "Artificial (Intelligent) Agents". See also: UNIDIR, "Weaponization of Increasingly Autonomous Technologies", 8.

27 Siboni & Eshpar, "Use of Autonomous Weapons", 81.

28 Heyns, "Lethal Autonomous Robotics", UNIDIR Conference. See also: Chairperson, *2014 informal Meeting of Experts on LAWS*, 4.

29 UK MoD, *National Security Through Technology*, Executive Summary.

30 Ibid, 38.

31 Ibid, 34 & 36.

32 Ibid, 20.

ranging set of autonomous applications than those considered by the military.[33] The CNAS report similarly argues that the Robotic Age is not being led by the American military-industrial complex, but by companies producing goods and services that are driving key enabling technologies such as advanced computing, big data, autonomy, AI, miniaturisation, additive manufacturing, and small but high density power systems, and that these technologies could be exploited to build increasingly sophisticated and capable unmanned and autonomous military systems.[34]

For example, Google acquired eight companies in late 2013 that included Boston Dynamics, a robotics company often known for its military robots, and DeepMind, a British AI start-up that specialises in an advanced form of machine learning called reinforcement learning to solve high-dimensional decision-making problems and improve mechanisms for knowledge representation, search and human-level reasoning.[35] This tool is deeply rooted in behavioural psychology and neuroscience to improve predicted modelling, reduce the amount of human intervention, and enhance decision-making.[36] These eight companies have created some of the best-engineered arms, hands, motion systems and vision processors in the robotics industry, while Boston Dynamics is regarded as one of the most accomplished robotics companies globally.[37] According to some analysts, while it was initially suggested that Google aimed to further automate factories, which are highly controlled environments suitable for a fleet of semi-independent robots, it is now clear that the vision is for "truly dexterous, autonomous robots".[38] Consequently, stronger collaboration between the public sector and industry, as well as academic research laboratories, is now advisable to not only shape policies responsibly, but to also prevent strategic surprises in the near future.

Financial constraints and decreasing defence budgets might also restrain some governments in this space, particularly if cuts are made at the expense of national security or Science, Technology, Engineering and Maths (STEM) investments in R&D are reduced. In the U.K., while it has the fourth largest defence budget, cuts were significant in defence and security science and technology over the last 15 years. Thus, in 2012, it identified a need to invest strongly where these cuts were made, as well as recognising the strategic importance of STEM subjects.[39] While in the U.S., given reduced defence resources, there could be a tendency to give preference to capabilities that are perceived as more affordable and good enough, rather than investing in R&D or pursuing more expensive, advanced systems that focus on potential future high-tech warfare.[40] That said, the budget for cyber capabilities in the U.S. remained relatively unscathed as compared to cuts in other areas. In comparison, other countries like China and Russia are investing heavily in advanced technologies such as cyber warfare tools, stealth and counter-stealth, and in capabilities designed specifically to exploit perceived vulnerabilities in U.S. made systems.[41]

Another important factor that should be considered includes the probability of a heightened need for and competition surrounding

---

33 UNIDIR, "Weaponization of Increasingly Autonomous Technologies", 6.

34 Work & Brimley, "War in the Robotic Age", 6.

35 Colin Lewis, "Google's DeepMind acquisition in reinforcement learning", http://robotenomics.com/2014/01/27/googles-deepmind-acquisition-in-reinforcement-learning/, 27 January 2014.

36 Lewis, "Google's DeepMind acquisition".

37 Illah Nourbakhsh, "Google's Robot Army", http://www.newyorker.com/tech/elements/googles-robot-army, 16 December 2013.

38 Nourbakhsh, "Google's Robot Army".

39 UK MoD, *National Security Through Technology*, 12.

40 Work & Brimley, "War in the Robotic Age", 20.

41 Work & Brimley, "War in the Robotic Age", 20.

technical talent, especially if there may not be enough STEM graduates to innovate or secure ICT systems in the near future.[42] A recent report by Microsoft concludes that by 2025, emerging economies will produce nearly 16 million graduates in STEM fields annually, which will be almost five times greater than the 3.3 million per year in developed countries.[43] There is an imbalance in STEM graduate rates and emerging economies like China, India, and Brazil show a different pattern to the U.S., where only four per cent of undergraduate degrees are in engineering compared to 31 per cent in China.[44] This is significant because not only might there be competition for talent between countries or regions, but also very possibly between the public and private sectors. Unless arrangements are made for enhanced collaboration between government and industry, governments could have major difficulties in sourcing the right skills sets.

When developing or implementing policies for this area, a balance must also be struck so as to ensure that innovation or economic growth opportunities will not be stifled. A number of delegations at the 2014 informal meeting on lethal autonomous weapons systems similarly recognised that the peaceful uses of autonomous technologies in the civilian field must be acknowledged and current technological development efforts should not be undermined.[45]

Nevertheless, if for instance, Google and Amazon plans succeed, this could usher in a new era of human-robot interaction where the public regularly faces robots in both public and private spaces.[46] Public perception and acceptance of some of these advanced technologies is therefore also very important, and it is likely to play a significant role in the shaping of future policies.

Ethical Concerns and Importance of the Public's Response

Heyns asks whether we want to enter a world where a robot can kill humans and there is no humanity or compassion.[47] Yet ethicists argue that, especially in the domain of technology ethics, answers to pressing questions like these are often unclear, law is often undefined, applications of new technologies are unclear, and social and political values conflict.[48]

After acquiring DeepMind, Google established an internal ethics committee. Such an ethics board might suggest that research takes a direction that avoids problems or that open discussions are held to defuse issues before public backlash occurs.[49] However, the danger is that such committees could apparently focus solely on consumer risk, improving public safety, and minimising corporate liability, which ethicists argue is an infamously grey moral area for emerging technologies because they are usually unanticipated and therefore unaddressed by regulations.[50] Nevertheless, such internal ethics boards could have real value if they deal with more than risk avoidance like the medical industry in the U.S. where they comprise lawyers, doctors, bioethicists, theologians and philosophers.[51] While such boards have been less common in other industries, it seems that not only are external

---

42 Microsoft, "Cyberspace 2025", 12.

43 Ibid, 4.

44 Ibid, 11.

45 Chairperson, *2014 informal Meeting of Experts on LAWS*, 3.

46 Nourbakhsh, "Google's Robot Army".

47 Heyns, "Lethal Autonomous Robotics", UNIDIR Conference.

48 Evan Selinger, "Inside Google's Mysterious Ethics Board", http://www.forbes.com/sites/privacynotice/2014/02/03/inside-googles-mysterious-ethics-board/, 02 March 2014.

49 Ibid

50 Ibid

51 Ibid

technology ethicists being called upon more frequently, several companies like BMW are also establishing internal ethics teams in recent times to guide the development of advanced technologies.[52] In recognising the significance of public response, the U.S. DoD Directive also notes that provision is made for the coordination and approval of guidance on public affairs matters concerning these systems and their use.

However, responses to these technologies might vary across countries or regions, and a number of possible scenarios arise. For instance, it is likely that as increasingly autonomous technologies become more deeply prevalent in daily life, the public might gradually accept developments and grow increasingly dependent on them. On one hand, consumer demand could drive development. Perceptions and trust in machine decision-making might in fact improve and several reports expect that industry, the scientific community and consumers will drive expectations and investment in further advances.[53] Autonomous self-driving cars are already expected to be rolled out, for example, to see how the public interact with them in the U.S., U.K., Singapore, and China.

If such advanced technologies are found to be more effective and a more affordable way of achieving national security objectives, this might even alleviate taxpayer concerns and justify their development and use. Conversely, if they are not found to be more cost-efficient and effective, large government expenditures could end up restricted in several countries. Likewise, public concern over casualties could make such systems more attractive if the public perceives that lives are saved through their use. Although, increasing accuracy might in fact mean that systems are more lethal.

It is likely that negative public response could sometimes end up restricting the public sector from using these technologies in certain countries. For example, the public's understanding of AI and autonomous systems might be overly influenced by doomsday scenarios, science fiction and popular culture, which possibly over-exaggerate robots' decision-making abilities.[54] Or according to one forecast, the public could tire of "robot smog", whereby autonomous robots "displace our sense of control precisely because they are out of our control but occupy the physical world and demand our attention".[55] Equally, solutions will be needed for issues such as a lack of information parity, where a machine knows everything about a human because it is connected to large cloud databases, and where it also acts as highly distributed sensors feeding information back to either corporate or government databases such as tracking where a person looks with computer vision, discerning emotions through facial analysis, and reading body language through gesture recognition.[56] There is little doubt that the impact of such technologies on civil liberties and fundamental human rights will also need extensive examination. Ultimately though, where the public sector might be sometimes restrained, it is likely that such technologies will still enter the commercial sector if they are not already developed by it. And it is equally likely that some states and malicious non-state actors will not be deterred by ethical concerns or negative public responses.

In conclusion, while there is still some uncertainty as to how maturing autonomous technologies, including potentially fully autonomous and lethal systems, will develop and impact national security, it is clear that several major policy questions are already evident. Key questions are identified throughout both parts of this

---

52 Ibid

53 UNIDIR, "Weaponization of Increasingly Autonomous Technologies", 7.

54 Heinl, "Artificial (Intelligent) Agents". See also, Heyns, "Lethal Autonomous Robotics", UNIDIR Conference.

55 Nourbakhsh, "Google's Robot Army".

56 Nourbakhsh, "Google's Robot Army".

report, which should be considered given the increasing interest in these technologies from both military circles and industry.

The first part of the report finds that a clearer understanding of the nature of these technologies would assist this debate especially since it is likely that states will pursue technological superiority via increasingly autonomous technologies for both economic and military reasons. Deeper analysis is required on the possible military advantages and disadvantages that might ensue, including the role of the human vis-à-vis the machine.

This second part of the report finds that currently, there are also major challenges in controlling and regulating this space as well as highly significant legal ambiguities and ethical question marks. It argues that the relationship between the public sector and industry should also be better managed to ensure that while innovation and economic growth are not restrained, this area will be developed responsibly. Lastly, while policy guarantees that the operation of systems will always be under human control, it does not seem certain from a technical standpoint that the human might always be in a position to control such systems.

Appendix 4

Possible Technical Safeguards:[57]

- Mandatory signatures or watermarks for purposes of identification;
- Guarantee of appropriate control under any circumstance;
- Setting strict constraints on their behaviour;
- Careful testing, although thorough verification of their safety and possible behaviours is apparently difficult;
- Restricting the environment as much as possible by only permitting an agent to operate on known platforms.
- Examine the extent to which an agent could communicate with its base, and whether communication should be one-way (intelligence gathering from the agent for instance) or two-way in that the command and control structure could issue instructions like target selection or self-destruct commands.
- Examine more carefully the possible cooperative behaviour of agents, in other words what is described as the "multi-agent" threat.
- Build safeguards such as backdoors and forced destruction into agents or self-destruction if loss of contact occurs.

Guidelines under the United States Directive:

- These systems shall be designed to allow commanders and operators to exercise appropriate levels of human judgement over the use of force;
- Systems will undergo thorough rigorous hardware and software verification, including analysis of unanticipated emergent behaviour resulting from the effects of complex operational environments on autonomous or semi-autonomous systems;
- Measures will ensure that systems function as anticipated, that they complete engagements in a timeframe consistent with commander and operator intentions, and if unable to do so that they terminate engagements or seek additional human operator input before continuing the engagement;
- They are sufficiently robust to minimise failures that could lead to unintended engagements or to loss of control of the system to unauthorised parties.
- For the potential consequences of an unintended engagement or loss of control of the system to unauthorised parties, hardware and software will be designed with appropriate safeties, anti-tamper mechanisms, and information assurance, and human machine interfaces and controls.
- In order for operators to make informed and appropriate decisions in engaging targets, the interface between people and machines for these systems will be readily understandable to operators, provide traceable feedback on system status, and provide clear procedures for operators to activate and deactivate system functions.

---

[57] Tyugu, "Command and Control of Cyber Weapons". See also: Heinl, "Artificial (Intelligent) & Guarino, "Autonomous Intelligent Agents".

- Any autonomous or semi-autonomous weapons systems intended to be used in a manner that falls outside the policies in the above table, must be approved before formal development and again before fielding. The guidelines for review of such systems include that before a decision to enter formal development is made that: 1) the system design incorporates the necessary capabilities to allow commanders and operators to exercise appropriate levels of human judgement in the use of force; 2) the system is designed to complete engagements in a timeframe consistent with commander and operator intentions, and if unable to do so, to terminate engagements or seek additional human operator input before continuing the engagement; 3) the system design including safeties, anti-tamper mechanisms, and information assurance addresses and minimizes the probability or consequences of failure that could lead to unintended engagements or to loss of control of the system; 4) plans are in place for to establish reliability, effectiveness, and suitability under realistic conditions, including possible adversary actions.

About the Author

Caitríona H. Heinl is a Research Fellow responsible for research on cybersecurity matters under the Homeland Defence Programme at the Centre of Excellence for National Security (CENS) within the S. Rajaratnam School of International Studies (RSIS). CENS is a research unit which works closely with the National Security Coordination Secretariat (NSCS) within the Prime Minister's Office, Singapore.

About the Centre of Excellence for National Security

The **Centre of Excellence for National Security (CENS)** is a research unit of the S. Rajaratnam School of International Studies (RSIS) at the Nanyang Technological University, Singapore.

Established on 1 April 2006, CENS raison d'être is to raise the intellectual capital invested in strategising national security. To do so, CENS is devoted to rigorous policy-relevant analysis across a range of national security issues.

CENS is multinational in composition, comprising both Singaporeans and foreign analysts who are specialists in various aspects of national and homeland security affairs. Besides fulltime analysts, CENS further boosts its research capacity and keeps abreast of cutting edge global trends in national security research by maintaining and encouraging a steady stream of Visiting Fellows.

About the S. Rajaratnam School of International Studies

The **S. Rajaratnam School of International Studies (RSIS)** is a professional graduate school of international affairs at the Nanyang Technological University, Singapore. RSIS' mission is to develop a community of scholars and policy analysts at the forefront of security studies and international affairs. Its core functions are research, graduate education and networking. It produces cutting-edge research on Asia Pacific Security, Multilateralism and Regionalism, Conflict Studies, Non-Traditional Security, International Political Economy, and Country and Region Studies. RSIS' activities are aimed at assisting policymakers to develop comprehensive approaches to strategic thinking on issues related to security and stability in the Asia Pacific.

For more information about RSIS, please visit www.rsis.edu.sg.

RSiS
Nanyang Technological University
S. RAJARATNAM SCHOOL OF INTERNATIONAL STUDIES

Nanyang Technological University
Block S4, Level B4, 50 Nanyang Avenue, Singapore 639798
Tel: +65 6790 6982 | Fax: +65 6794 0617 | www.rsis.edu.sg