




**Australian Army  
Research Centre**



**Autonomous Cyber Capabilities  
and International Law on the  
Use of Force, Self-Defence,  
Intervention, and Sovereignty**

Samuli Haataja

**Australian Army Occasional Paper No. 25**



**Australian Army  
Research Centre**

---

# **Autonomous Cyber Capabilities and International Law on the Use of Force, Self-Defence, Intervention, and Sovereignty**

**Samuli Haataja**

**Australian Army Occasional Paper No. 25**

*Serving the Nation*

© Commonwealth of Australia 2024

This publication is copyright. Apart from any fair dealing for the purpose of study, research, criticism or review (as permitted under the Copyright Act 1968), and with standard source credit included, no part may be reproduced by any process without written permission.

The views expressed in this publication are those of the author(s) and do not necessarily reflect the official policy or position of the Australian Army, the Department of Defence or the Australian Government.

ISSN (Print) 2653-0406

ISSN (Digital) 2653-0414

DOI: <https://doi.org/10.61451/267510>

All enquiries regarding this publication should be forwarded to the Director of the Australian Army Research Centre.

To learn about the work of the Australian Army Research Centre visit [researchcentre.army.gov.au](https://researchcentre.army.gov.au)

Cover image: Untitled by The Digital Artist (Source: Pixabay).

---

# Contents

<b>Executive Summary</b>	<b>1</b>
<b>Introduction</b>	<b>3</b>
<b>Autonomy and Cyber Capabilities</b>	<b>5</b>
Background	5
Defining Autonomy	6
Autonomous Cyber Capabilities	9
Concerns and Risks with Autonomous Systems	12
<b>International Law and Cyber Operations</b>	<b>14</b>
Use of Force	15
Self-Defence	15
Intervention	17
Sovereignty	19
Intention and Violations of International Law	21
<b>Autonomous Cyber Capabilities and International Law</b>	<b>24</b>
Offensive Cyber Operations	25
Active Cyber Defence Measures	28
<b>Conclusion</b>	<b>32</b>
<b>About the Author</b>	<b>34</b>
<b>Endnotes</b>	<b>35</b>



---

## Executive Summary

- In line with recommendations in the 2023 Defence Strategic Review, Australia is investing in efforts to strengthen its cyber capabilities to deliver broad and responsive capabilities options that can support Australian Defence Force operations.<sup>1</sup> Australia is also investing in artificial intelligence enabled autonomous cyber capabilities. Given Australia's commitment to upholding international law and a rules-based cyberspace, it must ensure that autonomous cyber capabilities are used in a way consistent with international law.
- Internationally, states have agreed that international law applies to their cyber activities. However, there continues to be debate about *how* the law applies in the cyber context and in relation to the use of autonomous cyber capabilities.
- Cyber capabilities enabled by artificial intelligence, or with autonomous functions, are best considered as tools used by human beings to achieve effects. Where these technologies cause effects in other states then their use may be in violation of international law. Legality will generally be determined based on the nature and extent of the effects of the cyber operation in question.
- In relation to offensive cyber operations involving autonomous cyber capabilities, the expected and foreseeable effects are generally known in advance. Accordingly, these systems can be designed to ensure that their use is lawful. However, in relation to autonomous cyber capabilities used as active defence measures to cause effects in other states (so-called automatic 'hack-backs'), the situation is less straightforward. For their use to be lawful, these systems would need to be capable of

making complex legal and factual assessments. The risk of unintended effects and escalation of conflict exists if these assessments are made incorrectly.

- Australia and other states developing and deploying autonomous cyber capabilities should take measures to mitigate the risk of unintended effects caused by these technologies. Where the application of international law is uncertain, they must be used in a responsible way that limits their proliferation and indiscriminate effects.

---

## Introduction

Australia is making significant investments in cyber capabilities, including the use of autonomous cyber capabilities (ACCs) enabled by artificial intelligence (AI). The 2023 Defence Strategic Review (DSR) recommends enhancing Defence's cyber capabilities in order to deliver broad and responsive capabilities to support Australian Defence Force (ADF) operations.<sup>2</sup> A key component in this effort is the Australian Signals Directorate (ASD) REDSPICE project, which aims to expand Australia's cyber capabilities and capacity by providing 'forward-looking capabilities essential to maintaining Australia's strategic advantage and capability edge over the coming decade and beyond'.<sup>3</sup> The blueprint for this project includes the delivery of 'AI-supported offensive and defensive cyber capabilities'.<sup>4</sup> Australia has also established the Advanced Strategic Capabilities Accelerator 'to deliver advanced technologies needed for Australia's national security'.<sup>5</sup> Further, the Defence Science and Technology Group (DSTG) has recognised the promise of AI for 'autonomous cyber operations' which offer 'the potential for distributed, adaptive defensive measures at machine speed and scale'.<sup>6</sup> At the same time, the 2023–2030 Australian Cyber Security Strategy reaffirms Australia's commitment to upholding international law and a 'rules-based cyberspace'.<sup>7</sup> Given Australia's investment in ACCs, and its commitment to upholding international law, it is important for the Department of Defence (including those developing and deploying these technologies) to understand how they can be used responsibly and lawfully.

This paper examines the application of public international law to the use by states of ACCs. The focus is on the laws regulating the use of force, self-defence, intervention, and sovereignty. The paper demonstrates that the core legal issues around the use of autonomous capabilities in the cyber context are not novel compared to those applying to cyber operations more generally. Nevertheless, the capacity for autonomy creates added legal complexity, and the use of autonomous capabilities increases the risk that international law will be violated. This is particularly the case with AI-enabled ACCs, which increases the likelihood of unpredictable or unintended effects.

This paper comprises five sections. Following the introduction in section one, section two provides background to international debates about



autonomy and international law concerning military weapons systems and debates about the extent to which international law applies to state activities in the cyber context. This section includes a definitional survey of 'autonomy' as a concept, some examples of current and prospective uses of autonomous cyber technologies, and observations around the key concerns and risks associated with the development and use of these technologies. Section three considers how international law on the use of force, self-defence, intervention and sovereignty is generally considered to apply to state activities in cyberspace. It also examines the relevance of a state's intention in relation to violations of international law. Section four analyses international law relating to the use of ACCs by states in cyber operations, with a focus on both offensive cyber operations and active cyber defence measures. This section demonstrates that, while a state can develop and lawfully use ACCs where the effects are limited to their own territory or jurisdiction, where effects are caused in other states the use of these capabilities can constitute an unlawful use of force (unless in self-defence), a violation of the non-intervention principle, and/or a violation of sovereignty. However, residual legal ambiguity makes uncertain the threshold at which these violations are likely to occur.

The paper concludes that, while the use of ACCs in cyber operations does not raise novel legal issues distinct from those raised by cyber operations generally, the use of ACCs as active cyber defence measures is particularly complicated. This is because of the need for these systems to be capable of making complex assessments that must take into account a range of legal and non-legal factors based on both technical and non-technical information. States developing and using these technologies should therefore ensure proper safeguards are in place to limit the risk of unintended effects, the unnecessary proliferation of ACCs, and indiscriminate harmful effects.

---

# Autonomy and Cyber Capabilities

## Background

In recent years, there has been much discussion about the international legal issues surrounding the use of autonomous weapons systems (AWSs). Following the publication of military policy documents in 2011 by both the UK and the US that referred to the prospect of such systems,<sup>8</sup> prominent human rights organisations called for a pre-emptive ban on lethal AWSs.<sup>9</sup> The supporting premise was that these weapons systems lack the requisite degree of human judgment and human qualities needed to comply with key principles of international humanitarian law.<sup>10</sup> Since 2014, these and other concerns have been regularly discussed under the auspices of the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects (CCW).<sup>11</sup> In this context, parties to the CCW, as well as other international stakeholders have considered various legal, ethical and technical issues surrounding AWSs.<sup>12</sup> A formal group of governmental experts (the 'CCW GGE') has met annually since 2017, and in this context states have agreed that international humanitarian law applies to the potential development and use of AWSs.<sup>13</sup>

In addition to the ongoing discussions about AWSs, there have been separate debates about the applicability of international law to state activities in cyberspace. Since the late 1990s, states have met within the United Nations General Assembly's First Committee on Disarmament and International Security to discuss 'Developments in the field of information and telecommunications in the context of international security'.<sup>14</sup> In 2004, a UN GGE was established to examine existing and potential threats from cyberspace.<sup>15</sup> As part of this process, in 2013, 2015 and 2021 states agreed to the general application of international law to their activities in cyberspace.<sup>16</sup> Parallel to the UN GGE meetings, a UN open-ended working group (OEWG) was also established, which convened in 2019 to develop rules of state behaviour in cyberspace.<sup>17</sup> Here too, states agreed on the general application of international law in the cyber context.<sup>18</sup>

There have evidently been extensive discussions at the international level about the legal implications of autonomy in relation to weapons systems, and discussions about how international law applies to state conduct in the cyber context. However, these deliberations have largely occurred in silos, focused on a single technology (either AWS or cyber technologies), despite the relevance of autonomy in both of these contexts.<sup>19</sup> The two areas are also conceived differently in government policy. For example, the US policy concerning AWSs expressly provides that it '[d]oes not apply to autonomous or semi-autonomous cyberspace systems for cyberspace operations'<sup>20</sup>—in other words, the policy addresses AWSs but not autonomy in the context of cyber capabilities.<sup>21</sup> States do not generally have specific public policies in relation to ACCs. For example, it is only recently that the US National Security Commission on Artificial Intelligence recommended that the US create a policy specifically addressing the use of AI in cyber operations.<sup>22</sup> Similarly, while questions about autonomy have been central to the CCW GGE discussions, they have been largely ignored in the UN GGE and OEWG discussions on cyber technologies.<sup>23</sup>

The limited consideration of ACCs and international law is also evident in academic literature. It is certainly possible to find scholarship on 'autonomous cyber weapons'.<sup>24</sup> However, the volume of such literature is low compared to AWS-focused research. Indeed, it is evident that the literature on the legal issues surrounding ACCs has only begun to develop,<sup>25</sup> with a paucity of literature dealing with the application of international law.<sup>26</sup> The *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations* (Tallinn Manual) only deals indirectly with the topic.<sup>27</sup> It provides an influential but non-binding account by international law experts about how international law applies in the cyber context.<sup>28</sup> While autonomy is implicit in some of its analysis of the law, the Tallinn Manual does not explicitly focus on the legal issues surrounding ACCs.<sup>29</sup>

## Defining Autonomy

In relation to military weapons systems, the notion of autonomy is itself a central point of contention in ongoing discussions at the CCW GGE on AWSs.<sup>30</sup> Part of the problem arises from the multiple ways in which the term is defined across different disciplines. Originally from the Greek words *autos* (self) and *nomos* (law), autonomy generally refers to self-governance or self-regulation. This meaning is reflected in social sciences, where the concept

is closely connected to independence. For example, in political science an institution (such as a nation state) is considered to be autonomous when it can regulate its own affairs.<sup>31</sup> In psychology, human beings or society are considered autonomous when they are in a state of independence and self-determination.<sup>32</sup> Then again, in philosophy autonomy refers to the capacity for self-government. Specifically, an agent is considered autonomous if its actions are ‘truly its own’ in contrast to circumstances where the agent’s will is under the control of another.<sup>33</sup> In more technical disciplines, such as AI and robotics, the term autonomy has been used in a ‘loose and undisciplined way’.<sup>34</sup> While different definitions are favoured by different professions, none provide a definitive approach.

In AI, autonomy is often associated with intelligence. This is evident in Peter Norvig’s and Stuart Russell’s writings. They maintain that an agent lacks autonomy if it ‘relies on the prior knowledge of its designer rather than its own precepts’.<sup>35</sup> For these authors, autonomy involves an agent’s ability to learn from its environment so that it ‘can become effectively *independent* of its prior knowledge’.<sup>36</sup> Using this approach, an agent must have the ability to learn (using machine learning, for example) to be considered autonomous. Others, however, argue that in the field of robotics:

‘autonomous’ carries [at a minimum] *some* of its philosophical meaning in the sense that an autonomous agent should be able to make informed decisions (based on its knowledge, rules and sensory input) and act accordingly.<sup>37</sup>

By contrast, others hold the view that the ways the term is used in social science and in technical disciplines are quite distinct. Specifically, in philosophy it is about the ‘freedom to choose goals’ for oneself, whereas in robotics autonomy refers to the ‘capacity for independent (unsupervised) action’.<sup>38</sup>

Some favour a broad technical definition of the term ‘autonomy’ that is not conflated with social science based definitions involving independence. For example, Patrick Lin, George Bekey and Keith Abney define autonomy in relation to robots as:

the capacity to operate in the real-world environment without any form of external control, once the machine is activated and at least in some areas of operation, for extended periods of time.<sup>39</sup>

Similarly, Stan Franklin and Art Graesser define an autonomous agent as:

a system situated within and a part of an environment that senses that environment and acts on it, over time, in pursuit of its own agenda and so as to effect what it senses in the future.<sup>40</sup>

Likewise, Tim McFarland defines autonomy as:

the ability of a system to behave in a desired manner or achieve the goals previously imparted to it by its operator, without needing to receive the necessary instructions from outside itself on an ongoing basis.<sup>41</sup>

The broad technical definitions are useful for several reasons. For one, they do not require a system to have specific technical capabilities (such as machine learning) to be considered autonomous. Further, they do not conflate technical autonomy with social science based definitions that involve an agent able to determine its own goals. As McFarland notes, broad technical definitions instead capture the relationship between the autonomous system and its environment, and the relationship between the autonomous system and its operator. A system must have the ability to operate within its environment. This requires a degree of awareness about its environment, capabilities that allow it to make changes or resist changes in its environment, and the ability to make choices about its capabilities that enable it to serve its overriding purpose.<sup>42</sup> As to the relationship between the autonomous system and its operator, once activated the system must be able to operate without external control in pursuit of its goal rather than rely on direct or real-time instructions from human operators.<sup>43</sup>

This paper will adopt a definition of ‘autonomous’ that is consistent with McFarland’s approach. Specifically, autonomy is defined as the ability of a system to operate within an environment in pursuit of its goal without direct or real-time human control. Using this approach creates a sound basis upon which to analyse the relevant law and policy considerations that apply to ACCs. This is because the law is mainly concerned with the outcomes or effects caused by a system with autonomous capabilities.<sup>44</sup> There are three important considerations underpinning this approach to autonomy:

- **‘Autonomy’ refers to a form of control as opposed to an absence of control.** Control is exercised in advance of the capability’s activation rather than occurring on an ongoing basis.

Specifically, while autonomous systems are programmed by human beings, once activated, they can only behave pursuant to the instructions encoded in the software.<sup>45</sup> Even where a system uses machine learning capabilities (so that the exact low-level steps it takes to achieve its high-level goal are not predetermined), the high-level goal or purpose of the system has nevertheless been pre-programmed by human beings. Currently and for the foreseeable future, such systems do not have independence—they cannot define and decide their ultimate goals like human beings can.<sup>46</sup>

- **The concept of autonomy is unrelated to the technical means used to achieve the autonomous effect.** A system can be considered to have autonomous capabilities whether it uses AI techniques such as online machine learning, or whether it simply follows predetermined coded instructions.
- **Autonomy comes in degrees.** A system's autonomous capabilities can vary—some functions may be autonomous but not others. Equally, the degree of autonomous (as opposed to manual) control can vary at different times (for example, autonomous capabilities may only activate when real-time control is impossible).<sup>47</sup> For this reason, attempts to categorise systems with autonomous capabilities (such as automatic, automated, autonomous, semi-autonomous, fully autonomous, and so on) are often simplistic and may not be helpful.<sup>48</sup> Instead, as other scholars have highlighted, what is important is that the system is capable of performing a significant function with a significant degree of autonomy.<sup>49</sup>

## **Autonomous Cyber Capabilities**

The term 'ACCs' refers to software capable of acting within an environment in pursuit of a predetermined goal without direct or real-time human control. This fact, however, does not make ACCs independent of human control; nor can they be considered as separate entities under the law.<sup>50</sup> Instead, an ACC is a tool that can be used by humans to achieve their goals.<sup>51</sup> Unlike AWSs, which are normally electromechanical systems operated by computer software, ACCs involve software agents that can operate across various computer systems. This means ACCs can be centralised within a single system, or can be distributed or duplicated across multiple systems and networks.

Common examples of cyber security software with autonomous capabilities include many firewalls and some intrusion prevention systems (IPSs). Firewalls, for example, use predetermined criteria to decide whether to prevent data traffic from entering into a specific network. Normally this occurs without any interaction from users, or even any knowledge of their presence.<sup>52</sup> Some IPSs may have various sub-systems that can determine whether network activity is malicious, block that activity, and potentially even repair damage caused by the activity (such as removal of virus-infected files).<sup>53</sup> As will be discussed in section four, international law does not prevent the use of these kinds of ACCs to perform cyber security functions. Legal issues arise, however, when ACCs are used in offensive or defensive cyber operations that cause effects in another state.

Stuxnet is illustrative of an ACC used in an offensive cyber operation. Stuxnet is the name given to a piece of malicious software discovered in 2010 that is reported to have been developed by the US and Israel. It was designed to disrupt Iran's uranium enrichment program.<sup>54</sup> The software infected non-networked computers within Iran's Natanz enrichment facility and adjusted the frequency setting that determines the speed at which nuclear centrifuges are spun.<sup>55</sup> To do this, it had a number of features that prevented anti-virus and other security mechanisms from detecting it, and it also had the capacity to make it appear to the human operators of the facility that the infected computers were operating normally.<sup>56</sup> Ultimately Stuxnet is understood to have been responsible for causing physical damage to approximately 1,000 centrifuges at the Natanz facility.<sup>57</sup> While Stuxnet's authors had the technical ability to control it through command-and-control servers, the Natanz facility that it operated within was not networked. So all of Stuxnet's functions were pre-embedded in code enabling it to operate autonomously.<sup>58</sup> Once activated, it was capable of propagating, identifying the appropriate systems to target, and delivering its payload without any direct or real-time human control.

An illustration of the possible ways in which ACCs can be used in a defensive capacity comes from the 'Cyber Grand Challenge' which was organised by the US Defense Advanced Research Projects Agency (DARPA) in 2016. The Cyber Grand Challenge involved the use of 'cyber reasoning systems' (CRSs) to perform cyber security functions without any real-time human intervention. For participating teams, the objective was to score points (and avoid losing them) by protecting the team's software from

adversaries. This was to be achieved by finding and patching vulnerabilities, keeping competitors' own software available, functional and efficient, and exploiting vulnerabilities in adversary software.<sup>59</sup> Once activated, all of this needed to be done by the CRS autonomously with no intervention by humans.<sup>60</sup> 'Mayhem', the CRS that won the competition, had the capability to autonomously discover and patch its own software vulnerabilities, as well as to discover and exploit vulnerabilities in its adversaries' software. Within a changing and unknown environment, Mayhem demonstrated the adaptive ability to make strategic decisions about which vulnerabilities to patch (or leave unpatched), which patches to use, which teams to attack and with what exploits, and how to allocate its resources in performing these functions.<sup>61</sup>

In 2019, the NATO Research Task Group IST-152 on 'Intelligent Autonomous Agents for Cyber Defense and Resilience' released a report providing a reference architecture and technical roadmap for 'intelligent software agents performing active, largely autonomous cyber-defense actions on military networks of computing and communicating devices'.<sup>62</sup> The report focused on an autonomous intelligent cyber defense agent (AICA), which is essentially an autonomous software agent that is able to operate at times when direct or real-time human control is impossible. The report was premised on a scenario in which a particular capability platform such as an unmanned aerial vehicle (UAV) is operating in an environment or at times when communications are being disrupted or are impossible.<sup>63</sup> In the scenario, the UAV was subject to a hostile cyber operation involving malicious software, and the AICA was tasked to keep the platform operational by defeating the hostile malicious software. According to the report, the AICA:

will stealthily monitor the networks, detect the enemy agents while remaining concealed, and then destroy or degrade the enemy malware. The agent will have to do so mostly autonomously, without support or guidance by a human expert.<sup>64</sup>

Developments in AI are expected to have a significant impact on cyber security and cyber operations. This prospect is reflected in, for example, the findings of the US National Security Commission in its 2021 report examining the current and future impacts of AI on national security. Specifically, the report found that the expanding application of AI-enabled cyber capabilities 'will make cyber-attacks more precise and tailored,



further accelerate and automate cyber warfare, enable stealthier and more persistent cyber weapons, and make cyber campaigns more effective on a larger scale'.<sup>65</sup> Similarly, the ASD's *Cyber Threat Report 2022–2023* also notes how '[m]alicious cyber actors could also use AI tools to augment their activities', including through phishing attacks, deepfakes, or 'to help orchestrate cyber intrusions'.<sup>66</sup> Experts predict that developments in machine learning will be a 'game changer'<sup>67</sup> in offensive and defensive cyber operations in the near future. Such developments may enable the development of ACCs that can be assigned a particular goal without the need to provide prior direction as to the specific ways in which the goal is to be achieved. While the ACC would be given the parameters of the environment in which it would be required to operate, it would have the inherent capacity to experiment and develop the strategies through which to achieve the goal itself.<sup>68</sup> This situation can be distinguished from most current cyber capabilities, which require advance knowledge of the target networks and the environment in which they need to operate.<sup>69</sup>

## **Concerns and Risks with Autonomous Systems**

The examples provided in the previous section illustrate the ways in which ACCs may be used in cyber operations either at present or in the future. While the capabilities have considerable potential in offensive and defensive cyber operations, concerns have been raised about the risks associated with the use of autonomous capabilities in the cyber security context. For example, the International Committee of the Red Cross has noted that there is a risk that ACCs incorporating AI and machine learning techniques (to autonomously defend against cyber threats and launch counter-attacks) hold the prospect of increasing the scale, and changing the nature and severity, of cyber-attacks.<sup>70</sup>

Unpredictability is a common concern with autonomous systems (whether ACCs or AWSs). The concern is that these capabilities may operate in an unpredictable way resulting in unintended effects.<sup>71</sup> For example, Paul Scharre notes that AWSs can behave erratically for various reasons, including a malfunction of the system or an unexpected interaction with its environment.<sup>72</sup> Generally it is more difficult to predict the behaviour of complex systems (whether they have autonomous capabilities or not) compared to simple systems.<sup>73</sup> This situation can be further exacerbated by, for example, sophisticated AI systems using neural networks in which

the outputs generated by their internal operations—so-called black boxes—can be unexpected and surprising even to the designers of the systems themselves.<sup>74</sup> Additional questions of predictability arise in relation to cyber systems, compared to physical mechanical systems, since cyber systems are distributed across the internet and can have effects on various systems.

A further concern that relates particularly to ACCs comes from the speed at which they operate. This factor makes it difficult for humans to intervene in time in the event that the software does not operate as intended. If states develop systems capable of automatic hack-backs against the adversary, then there is a risk that conflict will escalate at machine speed.<sup>75</sup>

---

## International Law and Cyber Operations

This section examines key rules of international law relevant to the legality of cyber operations involving the use of ACCs. While it may be lawful for states to employ ACCs to perform cyber security functions on systems and networks within their own territory, issues arise under international law when ACCs are used in cyber operations that cause effects in the territory of another state. In this context, international law prohibiting the use of force (unless in self-defence), the non-intervention principle, and the principle of sovereignty are particularly relevant. This section first outlines these areas of law before discussing ‘intent’ as a relevant notion in international law when determining whether the law has been violated. This concept is particularly important in relation to cyber operations involving ACCs that may cause effects that were unintended by the state using the capability.

As outlined in section two, at meetings of both the UN GGE and the OEWG, states have agreed on the general application of international law to their activities in cyberspace. For example, in the final report of the OEWG, released in 2021, states reaffirmed that international law ‘is applicable and essential to maintaining peace and stability and promoting an open, secure, stable, accessible and peaceful ICT environment’.<sup>76</sup> This sentiment was echoed in the 2021 report of the UN GGE.<sup>77</sup> Despite general agreement about the applicability of international law, however, there continues to be debate about the ways in which specific rules apply in this context. Further, the UN GGE and OEWG reports are not legally binding on states, and there is no legally binding treaty or other instrument applicable specifically to states’ cyber activities. Therefore, it is critical to develop shared understandings about how existing rules of international law apply. To this end, many states have provided official positions on how they consider international law to apply in the cyber context. By virtue of making such assertions, states can contribute to the development of customary international law on the topic.<sup>78</sup> It remains the case, however, that the law is not settled. This means that legal analysis on the use of ACCs is inevitably based on interpretations of the law that remain contested or uncertain.

## Use of Force

Under international law, there is a general prohibition on the use of force. This restriction is contained in article 2(4) of the UN Charter, which provides that:

All Members [of the UN] shall refrain in their international relations from the threat or use of force against the territorial integrity or political independence of any state, or in any other manner inconsistent with the Purposes of the United Nations.

The primary exceptions to this prohibition occur when the use of force is in self-defence, when it has been authorised by the UN Security Council, and when the targeted state consents to it. While article 2(4) was drafted in response to the spectre of 1940s wartime technology, in 1996 the International Court of Justice (ICJ) affirmed that it still applies broadly ‘to any use of force, regardless of the weapons employed’.<sup>79</sup> This means that the threat or use of force can be imposed through a range of kinetic, chemical or biological means and methods, or through the execution of cyber operations. Many states consider that a cyber operation will constitute a use of force where its ‘scale and effects’ are comparable to a use of force by traditional kinetic means.<sup>80</sup> For example, where a cyber operation results in damage or destruction of physical property, or injury or death of human beings, it is likely to be considered as amounting to a use of force in violation of article 2(4) of the UN Charter.

## Self-Defence

Self-defence is one of the key exceptions to the prohibition on the use of force. Article 51 of the UN Charter recognises the right of self-defence under international law. The relevant part provides that:

Nothing in the present Charter shall impair the inherent right of individual or collective self-defence if an armed attack occurs against a Member of the United Nations, until the Security Council has taken measures necessary to maintain international peace and security.<sup>81</sup>

When a state is subject to an ‘armed attack’, it has a right to use force in self-defence provided that its response is necessary and proportionate.<sup>82</sup> According to the ICJ, an armed attack is the most serious use of force.

Whether an incident is considered to be an armed attack is determined with reference to the ‘scale and effects’ of the use of force.<sup>83</sup> Whereas most states consider that an ‘armed attack’ involves a higher degree of violence than a ‘use of force’, the US maintains that the thresholds are the same.<sup>84</sup>

While not explicitly included in the UN Charter, customary international law provides that states also have a right of anticipatory self-defence—i.e. before they have been the victim of an armed attack. The circumstances are limited to when the need to take measures in self-defence is ‘instant, overwhelming, leaving no choice of means, and no moment for deliberation’.<sup>85</sup> In 2002, after the 11 September 2001 terrorist attacks against the US, the US proposed that the scope of the right to self-defence should be expanded to justify the use of force in response to a perceived threat against a state. This proposition was broadly rejected within the international community, however, on the basis that it stretches the concept of anticipatory self-defence too far, to situations where no real or imminent threat of attack exists.<sup>86</sup>

The Tallinn Manual provides that a state that is the victim of a cyber operation constituting an armed attack can exercise its right of self-defence.<sup>87</sup> In this context, the ‘scale and effects’ of the cyber operation will determine whether the threshold of armed attack has been reached.<sup>88</sup> There is a degree of consensus among states that a cyber operation will constitute an ‘armed attack’ where its effects are similar to those that would be achieved by means of an armed attack carried out by kinetic means. For example, according to Australia:

the thresholds and limitations governing the exercise of self-defence under Article 51 apply in respect of cyber operations that constitute an armed attack and in respect of acts of self-defence that are carried out by cyber means. Thus if a cyber operation—alone or in combination with a physical operation—results in, or presents an imminent threat of, damage equivalent to a traditional armed attack, then the inherent right to self-defence is engaged.<sup>89</sup>

Many other states have recognised the application of the international law on self-defence in the cyber context. Most of these countries also endorse a ‘scale and effects’ approach to determining whether the threshold of ‘armed attack’ has been reached. These include Finland,<sup>90</sup> France,<sup>91</sup> Germany,<sup>92</sup> Iran,<sup>93</sup> the Netherlands,<sup>94</sup> the US,<sup>95</sup> the UK,<sup>96</sup> Switzerland,<sup>97</sup> most members

of the Organisation of American States,<sup>98</sup> and NATO.<sup>99</sup> For example, to illustrate an armed attack in the cyber domain, New Zealand's position on how international law applies describes a situation involving a 'cyber activity that disables the cooling process in a nuclear reactor, resulting in serious damage and loss of life'.<sup>100</sup> The UK has stated that an armed attack would occur if cyber operations interfered with a nuclear reactor resulting in 'widespread loss of life', or if cyber means were used to disable air traffic control systems, causing the downing of a civilian aircraft and resulting in lethal effects.<sup>101</sup>

As to anticipatory self-defence, only a few states have outlined their views on its application to cyber operations. Australia is among them.<sup>102</sup> In a lecture given at the University of Queensland, then Attorney-General George Brandis reaffirmed the principle that:

a state may act in anticipatory self-defence against an armed attack when the attacker is clearly committed to launching an armed attack, in circumstances where the victim will lose its last opportunity to effectively defend itself unless it acts.<sup>103</sup>

He then gave the example of 'a threatened armed attack in the form of an offensive cyber operation' that could be launched 'in a split second' and could cause 'large-scale loss of human life and damage to critical infrastructure'.<sup>104</sup> While it can be difficult to determine whether a cyber operation has sufficiently serious and imminent effects, Brandis's view indicates that, as a matter of principle, states have a right to necessary and proportionate anticipatory self-defence in response to a cyber operation.

Despite the existence of clear-cut examples of cyber operations amounting to an 'armed attack', their characterisation as such will inevitably be informed by political and strategic considerations within the framework of international law. Accordingly, while technical information about the nature of the attack will be relevant,<sup>105</sup> the victim state will need to weigh various non-legal factors when deciding whether to use force in self-defence.

## **Intervention**

For cyber operations that fall below the use of force threshold, the non-intervention principle has particular relevance. This principle is closely connected with respect for state sovereignty. It is a rule of customary

international law that prohibits states from using coercive means to intervene in the internal or external affairs of another state, as these are matters which states have the right to decide freely by virtue of their sovereignty.

The non-intervention principle was framed in the 1970 UN General Assembly resolution titled 'The Declaration on Principles of International Law concerning Friendly Relations and Co-operation among States' in the following terms:

No State or group of States has the right to intervene, directly or indirectly, for any reason whatever, in the internal or external affairs of any other State. Consequently, armed intervention and all other forms of interference or attempted threats against the personality of the State or against its political, economic and cultural elements, are in violation of international law.<sup>106</sup>

Historically the non-intervention principle was only considered relevant to 'forcible or dictatorial' interference—that is, interventions involving the use of force.<sup>107</sup> In the modern context, however, a violation of the non-intervention principle is more broadly defined.<sup>108</sup> In the 1986 *Nicaragua* case, the ICJ discussed the principle of non-intervention, stating:

A prohibited intervention must accordingly be one bearing on matters in which each State is permitted, by the principle of State sovereignty, to decide freely. One of these is the choice of a political, economic, social and cultural system, and the formulation of foreign policy. Intervention is wrongful when it uses methods of coercion in regard to such choices, which must remain free ones. The element of coercion, which defines, and indeed forms the very essence of, prohibited intervention, is particularly obvious in the case of an intervention which uses force[.]<sup>109</sup>

On this basis, the non-intervention principle contains two elements. First, there must be coercive interference, and second, it must be directed towards the matters that a sovereign state should be able to decide freely.<sup>110</sup> There is widespread agreement among states that the non-intervention principle is applicable in the cyber context.<sup>111</sup> The Tallinn Manual also provides in Rule 66 that states 'may not intervene, including by cyber means, in the internal or external affairs of another State'.<sup>112</sup>

While it is clear that the non-intervention principle applies to cyber operations, the term coercion is not defined in international law. The majority of the Tallinn Manual experts adopted a narrow approach defining it as:

an affirmative act designed to deprive another State of its freedom of choice, that is, to force that State to act in an involuntary manner or involuntarily refrain from acting in a particular way.<sup>113</sup>

They distinguished coercion from activities that seek to influence (as opposed to factually compel) a state to behave in a particular way, such as public criticism or propaganda.<sup>114</sup> Other commentators, however, have defined coercion more broadly as constituting a form of pressure that seeks to deprive the target state of its free will.<sup>115</sup> This view was shared by the minority of the Tallinn Manual experts, who maintained that it is sufficient that the activity 'has the effect of depriving the State of control over the matter in question'.<sup>116</sup> This broader approach to coercion is also supported by some states in the cyber context. For example, according to the Australian government, coercion involves 'effectively depriv[ing] another state of the ability to control, decide upon or govern matters of an inherently sovereign nature'.<sup>117</sup>

The second element of the non-intervention principle requires that the coercion is directed at matters that states are able to decide freely by virtue of their sovereignty.<sup>118</sup> According to the ICJ, these matters include the 'choice of a political, economic, social and cultural system, and the formulation of foreign policy'.<sup>119</sup> States have also affirmed this in their official positions on how international law applies in cyberspace. Australia's position is that the law requires the coercion to affect 'matters of an inherently sovereign nature',<sup>120</sup> and the UK and New Zealand have adopted a similar approach.<sup>121</sup> Thus there is agreement that the non-intervention principle applies in the cyber context where there is coercion directed at matters that states are able to decide freely by virtue of their sovereignty, though there continues to be debate about what constitutes coercion.

## **Sovereignty**

Cyber operations that do not involve a use of force, or an unlawful intervention, may nevertheless be considered in violation of international law on sovereignty. Sovereignty generally refers to a state's supreme authority



within its territory. It involves three interrelated rights and correlating duties in relation to a state's territorial sovereignty, independence, and sovereign equality.<sup>122</sup> By virtue of the sovereignty principle, states have the right to exercise jurisdiction within their territory, have the freedom to conduct their own affairs independently, and have formal equality under international law.<sup>123</sup> Non-consensual activities within a state's territory are generally prohibited. Such activities may include, for example, flying military aircraft over a state's airspace without its consent.

While states agree that sovereignty extends to the cyber infrastructure within their territory, there is uncertainty about whether (and the threshold at which) remotely conducted cyber operations violate territorial sovereignty. This uncertainty arises primarily from ongoing debate about whether sovereignty operates as a rule of international law independent of principles concerning the use of force and non-intervention. There are two views. According to the 'sovereignty as a rule' approach, sovereignty operates as a primary rule of international law that can be violated. On this view, remotely conducted cyber operations can in some circumstances amount to violations of sovereignty in their own right. The Tallinn Manual advocates this approach<sup>124</sup> and a growing number of states have adopted the same position. These countries include the Netherlands,<sup>125</sup> France,<sup>126</sup> Austria,<sup>127</sup> the Czech Republic,<sup>128</sup> Finland,<sup>129</sup> Iran,<sup>130</sup> New Zealand,<sup>131</sup> Germany,<sup>132</sup> Switzerland,<sup>133</sup> and several states from the Organisation of American States.<sup>134</sup> In contrast, according to the 'sovereignty as a principle' approach, sovereignty does not operate as a standalone primary rule capable of being violated independent of other rules of international law.<sup>135</sup> This approach has been expressly adopted by the UK.<sup>136</sup> It was also a position put forward by a 2017 US Department of Defence memorandum,<sup>137</sup> although the US has since departed from this approach.<sup>138</sup>

Despite these contrasting positions, the core question under debate concerns the threshold at which cyber activities violate international law. There is uncertainty about this even among the states adopting the rule approach to sovereignty. For example, the French and Iranian positions suggest that any unauthorised penetration of their systems or networks would violate sovereignty.<sup>139</sup> In contrast, other states like New Zealand maintain that not 'every unauthorised intrusion' into another state's ICT systems or even all cyber activities that cause effects on another state's territory will be a violation of territorial sovereignty.<sup>140</sup> While there remains

uncertainty about what degree of effects are required for a violation to occur,<sup>141</sup> an increasing number of states nevertheless agree that sovereignty operates as a rule that can be violated in the cyber context. Most also agree that cyber operations causing physical damage or harm to individuals within another state's territory will violate territorial sovereignty, as will cyber operations that undermine a state's governmental functions.<sup>142</sup>

## **Intention and Violations of International Law**

The previous section has demonstrated that cyber operations that are attributable to states are capable of violating international legal principles prohibiting the use of force, the non-intervention principle, and sovereignty. Prior to examining how these rules apply to the use of ACCs, it is necessary to consider the extent to which a state's intention matters for violations of international law. This is important as it informs the question of potential state responsibility for the unintended effects caused by ACCs. The relevant rules concerning intention are detailed in international law on state responsibility. This area of law provides secondary rules that determine, among other things, the circumstances under which a state will be responsible for a violation of international law.

While there has been historical debate on the topic,<sup>143</sup> the International Law Commission (ILC) provides contemporary guidance on the relevance of 'intent' to state responsibility. The ILC is a peak UN body responsible for codifying and developing the law in this context. According to the ILC:

[i]n the absence of any specific requirement of a mental element in terms of the primary obligation, it is only the act of a State that matters, independently of any intention.<sup>144</sup>

Therefore, a state's 'intent' is not uniformly determinative of whether a violation of primary rules of international law has occurred.

Based on the ILC's reasoning, a state will be in violation of international law on the use of force if it designs and uses an ACC that causes effects that are sufficiently serious to constitute a use of force. This is the case even where those effects are unintended and occur, for example, as a result of unforeseen interactions between the ACC and its environment, or due to a programming error. It is true that the ICJ in the *Nicaragua* case maintained that any use of armed force will involve a degree of coercion,<sup>145</sup> and following

this reasoning it could be suggested that the intention of the responsible state is relevant in determining whether there is a violation of article 2(4). Nevertheless, intention is not generally regarded as a constituent element of the prohibition on the use of force.<sup>146</sup> Instead, as outlined above, most states determine whether a cyber operation constitutes a use of force based on considerations of the ‘scale and effects’ of the operation. A similar view is advanced by the Tallinn Manual.<sup>147</sup>

Similarly, intention is not required for a violation of sovereignty.<sup>148</sup> This position was agreed by the Tallinn Manual experts who affirmed that:

a cyber operation by or attributable to a State that is not intended to result in consequences that violate the sovereignty of another State, but that nevertheless generates them, is a violation of sovereignty.<sup>149</sup>

While debate remains about the precise threshold at which effects violate sovereignty, a cyber operation causing the requisite effects can be unlawful even if those effects were unintended.

In contrast to the principles regarding the use of force and sovereignty, the intention of the responsible state is relevant to the application of international law concerning non-intervention.<sup>150</sup> As outlined above, a violation of the principle against non-intervention involves coercive interference, an element that presumes intent. Therefore, to constitute unlawful intervention a cyber operation must be intended to coerce the target state in relation to matters it has the right to decide freely by virtue of its sovereignty.<sup>151</sup> The Tallinn Manual adopts this approach. Indeed, it further asserts that even a cyber operation that fails to produce the desired outcome can nevertheless constitute a prohibited intervention.<sup>152</sup>

To summarise, while there continues to be debate about how international law applies to cyber activities by states, there is general consensus that the prohibition on the use of force, the right to respond in self-defence, and the non-intervention principle do apply. Equally, most states have adopted the position that sovereignty can be violated by cyber operations. However, there is no common position concerning the threshold at which violations of international law will occur. As a generalisation, whether a cyber operation is unlawful depends on its scale and effects. At one end of this spectrum are cyber espionage activities with no or minimal effects, which are generally considered lawful.<sup>153</sup> At the other end of the spectrum are cyber operations with significant effects that rise to the threshold of an armed attack. These

operations violate the prohibition on the use of force and trigger a state's right to respond in self-defence. Between these extremes, there can be various cyber operations that violate sovereignty (for example, cyber operations that cause some effects in the territory of another state, or disrupt the operation of government services); the non-intervention principle (for example, disruptive cyber operations that seek to deprive the victim state of the ability to decide government policy); or the prohibition on the use of force (for example, cyber operations that cause significant physical damage or destruction of hardware in a state). Within this spectrum, there are legal 'grey zones' in which there is uncertainty about the precise way in which the law applies. Finally, cyber operations causing unintended effects can constitute violations of the prohibition on the use of force or sovereignty. By contrast, the non-intervention principle cannot be violated without the requisite intention by the responsible state.

---

## Autonomous Cyber Capabilities and International Law

This section examines how international law on the use of force, self-defence, intervention and sovereignty apply to the employment of ACCs by states. The analysis is organised around the use of ACCs in offensive cyber operations, and as active cyber defence measures (so-called automatic hack-backs). The paper assumes that these operations are conducted by states or are attributable to states.<sup>154</sup>

The rules of international law outlined in the previous section do not prohibit states from using ACCs where their effects are limited to the state's own territory or jurisdiction. For example, ACCs may be developed for domestic law enforcement purposes and be used to disrupt the operation of the computer systems used for criminal activities, or to protect government systems and networks from criminal activities. Even when disruptive effects are caused by these ACCs, they do not violate the sovereignty of another state, constitute an intervention, or constitute a use of force under international law.<sup>155</sup> Autonomous capabilities can also be used lawfully as passive defence measures (such as firewalls) in order to block malicious online traffic originating from foreign actors. This is because no effects are caused in the territory or jurisdiction of another state, and because the activities constitute a lawful exercise of sovereign power over the using state's own cyber infrastructure.<sup>156</sup> It is equally lawful for a state to use ACCs to perform cyber security functions to protect its maritime vehicles, UAVs and other national assets located within another state's offshore jurisdictions.

By contrast to the lawful uses of ACCs within a state's own territory or jurisdiction, their use more broadly can result in violations of international law depending on the nature of the effects caused. For example, a state may develop an offensive cyber capability that, once activated, is capable of locating offshore computer systems that match designated parameters and that can deliver a payload designed to disrupt those systems' operations or shut them down. Where such systems are used to operate, for instance, the target state's critical infrastructure, the effects of the cyber operation may be significant. Similarly, a state may develop and use an automatic hack-back as a defence system that identifies and disrupts the functioning

of an offshore computer system being used to launch the hostile cyber operation.<sup>157</sup> As will be demonstrated in the next section, regardless of the ‘offensive’ or ‘defensive’ nature of the cyber operation in question, its effects on the targeted state are ultimately the most relevant factor for legal analysis.

## **Offensive Cyber Operations**

This section considers the legality of ACCs used in offensive cyber operations with reference to the legal principles outlined in the previous section. Stuxnet provides a useful illustration of several relevant legal considerations. As outlined in section two, Stuxnet is reported to have caused physical damage to approximately 1,000 uranium enrichment centrifuges in Iran. The event involved an offensive cyber operation in which the ACC exercised a significant degree of autonomy in how it spread and delivered its payload.

When subject to legal analysis, the Stuxnet operation is generally considered to have reached the threshold of a ‘use of force’ because of the extent of physical damage caused by its payload within the Natanz facility in Iran.<sup>158</sup> The fact that Stuxnet had the capability for autonomy in how it spread and delivered its payload is not relevant to this assessment. Stuxnet also spread unintentionally to computer systems in third states that were not its intended targets. While it did not cause effects in those states, consider a hypothetical scenario in which it delivered its payload and caused effects constituting a use of force against those states. Stuxnet’s use in this hypothetical scenario would also constitute a use of force given that intention is not an element of the prohibition on the use of force.

Given the debate about when a cyber operation violates sovereignty, there is no consensus as to whether a cyber operation involving an ACC would violate the targeted state’s sovereignty.<sup>159</sup> Applying the approach taken by France and Iran to the Stuxnet example, any penetration of their systems or networks by Stuxnet would have constituted a violation of sovereignty, regardless of the gravity of the effects caused. This view is supported by many academic commentators.<sup>160</sup> By contrast, for states like New Zealand there would have been no violation of sovereignty if Stuxnet had simply spread into systems within its territory.<sup>161</sup> In a similar vein, the UK maintains that an ACC would only violate international law where it caused effects constituting a use of force, or if it was used to coerce a state in

relation to matters it can freely decide, in violation of the non-intervention principle.<sup>162</sup> Accordingly, it is currently unclear what degree of effects would be required for the use of an ACC to violate sovereignty, and this is likely to be considered on a case-by-case basis. But in circumstances where the threshold of effects is reached, the use of an ACC will constitute such a violation.

A state will violate the principle of non-intervention if it uses an ACC in a coercive manner with the intent to affect matters about which the target state has the right to decide freely.<sup>163</sup> In this context, it does not matter whether the use of an ACC causes physical or only disruptive effects. However, in contrast to the law on the use of force and sovereignty, the law on intervention does require an element of *intent*. This means that, if the use of an ACC results in either intended or unintended effects but there is no intent to coerce the targeted state, then the responsible state is not in violation of the non-intervention principle.<sup>164</sup> For example, Stuxnet is considered to have violated the principle of non-intervention as it related to Iran. This is because its use was intended to deprive Iran of the ability to operate its uranium enrichment program (even if Iran was oblivious to it) and this action constituted coercion against a matter that Iran as a sovereign state is permitted to decide.<sup>165</sup> By contrast, if Stuxnet had also delivered its payload and disrupted uranium enrichment processes in other states, this action would not be regarded as violating the non-intervention principle in those states because there was no intention to deprive them of their ability to operate their uranium enrichment programs. Importantly, the focus is on the intention of the state in using the ACC, and not simply the specific technical effects that the state intended or did not intend to cause through its use.

It is lawful to use ACCs in self-defence even when the effects caused constitute a use of force in their own right. Reliance on the justification of self-defence requires the victim state to establish that it has been subject to an armed attack, that it is necessary to use force in response (as opposed to non-forceful means such as peaceful dispute resolution mechanisms), and that its defensive use of the ACC is a proportionate response to the attack.<sup>166</sup> For example, if Stuxnet had been used in response to an armed attack carried out by cyber or non-cyber means, its use could have been justified under international law as an act of self-defence provided that the effects caused were demonstrably necessary and proportionate.

While this section has considered the legality of ACCs with reference to the Stuxnet example, it is important to consider the extent to which the use of AI-enabled ACCs affects the legal analysis. Stuxnet operated with a significant degree of autonomy and had the ability to deliver its payload without any direct or real-time human control, but it nevertheless operated according to predetermined rules. Specifically, it was programmed based on analysts' precise knowledge of the environment in which it would operate, and the hardware and software configuration of the computers that it targeted. Based on these calculations, Stuxnet was designed and tested so that its payload operated in a predictable way against specific computer configurations known to be in use at Iran's Natanz facility. While Stuxnet inadvertently spread to computer systems in other countries, it was unable to deliver its payload beyond Iran. However, unlike Stuxnet and similar capabilities that operate according to predetermined rules, AI-enabled ACCs such as those using machine learning can operate differently. As outlined in section two, AI-enabled ACCs may simply be given the parameters of the environment and assigned a goal (such as disrupting the operation of specific computer systems within a network), and the AI will then be able to test and develop strategies through which to achieve the goal. The operation of these ACCs may also be opaque. While Stuxnet could be reverse engineered so that its precise operation could be understood, more sophisticated AI-enabled ACCs can behave in ways unknown to their developers. The use of AI in this context increases the likelihood of the ACC operating in unpredictable ways, and this in turn increases the risk of causing unintended effects which may have legal consequences.

For example, consider a hypothetical 'AI Stuxnet' with machine learning capabilities that is assigned the goal of disrupting the operation of the computers responsible for the operation of the centrifuge machines in Iran's Natanz facility. In this example, the ACC is not pre-programmed with the steps it will take to achieve this outcome. If this hypothetical AI Stuxnet caused the same effects as the real Stuxnet, the conclusions of the above legal analysis would be the same. That is, whether the mere propagation of AI Stuxnet would violate sovereignty would be uncertain given the current debate about the threshold at which this occurs; its use would constitute a violation of the non-intervention principle as it was intended to coerce Iran into changing matters it can freely determine; and its use would constitute a use of force given the effects caused by its payload. Further, even where the hypothetical AI Stuxnet operated in an unpredictable way and caused



unintended effects in Iran or in another state that constituted a use of force or a violation of sovereignty, the state that designed and activated it would be responsible. And if the state responsible for the AI Stuxnet had the intent to coerce Iran, its use would constitute a violation of the non-intervention principle (even if the specific technical effects it caused were unintended). As this example demonstrates, where a state is responsible for the development and use of an ACC, it is the effects caused by the use of the ACC that are most relevant for legal analysis, and not the technical means used to achieve autonomy.

The Stuxnet example therefore usefully illustrates how international law applies to the use of ACCs by states in an offensive cyber operation that causes effects in another state. The autonomy of the cyber capability in this context does not raise novel legal issues different to those relating to cyber operations generally. To an extent, this is because ACCs used in offensive cyber operations often involve capabilities developed for a particular operation,<sup>167</sup> and this means the circumstances in which they will be used and the effects they are intended or expected to cause will be known in advance.<sup>168</sup> However, the use of AI-enabled ACCs in this context does increase the risk of legal violations, because the capabilities' complex AI systems are more likely to operate in unpredictable ways and cause unintended effects. These risks must be taken into account by states developing and using these capabilities to ensure their lawful use.

## **Active Cyber Defence Measures**

Legal issues concerning the use of ACCs as active cyber defence measures are not novel compared to cyber operations generally. Nevertheless, application of the law is highly complex in this context. This is particularly the case where the ACC is designed to automatically respond to malicious cyber operations and needs the capability to make a range of assessments about what the relevant legal frameworks are and to ensure that its response accords with those frameworks. Achieving such capability requires the ACC to make a range of contextual judgments involving a variety of legal and non-legal factors based on both technical and non-technical information.

A hypothetical scenario inspired by Mayhem illustrates the legal complexity of using ACCs for active cyber defence in response to cyber threats. As outlined earlier in this paper, Mayhem is the system that won the DARPA

Cyber Grand Challenge in 2016. In reality, Mayhem provided a proof of concept of the utility of ACCs in cyber defence, and its effects were only virtual in nature. For the purposes of this example, a purely hypothetical ACC called 'Chaos' will be used instead. The analysis will be based on the following scenario.

Chaos is AI-enabled software designed as a defensive cyber capability by state A to protect the cyber security of its government systems, networks, and critical infrastructure. Once activated, Chaos is able to identify and respond to malicious cyber operations against state A without direct or real-time human control. In addition to passive defence measures (such as detecting and blocking malicious traffic), Chaos is capable of taking active measures such as tracing the source of a malicious cyber operation and disrupting the operation of the computer system and networks being used to undertake it. Chaos is active in state A when it is subject to hostile cyber operations from state B. Passive defence measures are insufficient to stop the cyber operation, so Chaos adopts active defence measures causing effects in state B.

The legal analysis of this scenario raises a number of relevant issues. Firstly, the question of whether the use of Chaos by state A to cause effects in state B would violate international law would largely depend on the nature and extent of the effects that Chaos caused. In relation to sovereignty, for example, consider a situation in which Chaos only traced the source of the malicious cyber operation in order to collect intelligence about who is responsible for it in state B (for example, to help establish attribution). Currently, few states would consider this action as violating the sovereignty of state B.<sup>169</sup> As a form of espionage, it would be lawful for state A to use Chaos in this way.<sup>170</sup> By contrast, if Chaos went beyond intelligence collection to conduct disruptive or limited physical effects in state B, then its use would more likely constitute a violation of sovereignty. Further, if in these efforts Chaos caused significant physical effects in state B, then its use would likely also constitute an unlawful use of force. Were these effects to be unintended (for example, as a result of Chaos operating in an unpredictable way), state A would nevertheless be in violation of the law, because intention is not a requisite element in assessing a use of force or a breach of sovereignty. By contrast, the use of Chaos would only violate the principle of non-intervention where its deployment was coupled with the intent to coercively affect matters that state B has the right to freely decide.

Determining legality in this context would depend on an assessment of the intention of state A in designing and using Chaos, and not only the nature and extent of effects caused.

The circumstances in which Chaos is used are relevant to assessing issues of legality. For example, depending on the scale and effects of state B's cyber operation that it responded to, state A could potentially use Chaos in a lawful act of self-defence.<sup>171</sup> This would be the case, for example, if state A were to have been subject to a cyber operation constituting an armed attack by state B, and if state A's use of Chaos in response, while constituting a use of force, were deemed necessary and proportionate. Such a response, however, would need to be premised on a number of complex judgments based on a predetermined threshold of effects coded into Chaos by state A.<sup>172</sup> Specifically, Chaos would need to be capable of attributing the cyber operation to state B both factually and under international law, determining that state B's cyber operation amounted to an 'armed attack' to which it was necessary for state A to respond with force, and determining that the force used was proportionate. Determining what constitutes an 'armed attack' by state B would require an assessment of the level of destruction, injury and/or death caused. Determining that the scale and effects of state B's cyber operation reached this threshold would require quantification of both the virtual effects of the cyber operation and its real-world effects.<sup>173</sup> Determining whether the use of force by Chaos in self-defence was necessary and proportionate would require a determination that passive defence measures (such as blocking the malicious traffic) were insufficient to terminate state B's cyber-attack, and that other non-forcible measures (including peaceful dispute resolution mechanisms with state B) were also unavailable.<sup>174</sup> While some elements of these threshold judgments could be pre-programmed into Chaos based on technical considerations, some involve complex political and strategic factors that are not amenable to coding.

It is a feature of international law that judgments around legality are highly contextual. Politics and international law are inextricably linked, meaning that judgments concerning, for example, the gravity of 'scale and effects'<sup>175</sup> constitutes 'a political decision taken in the framework of international law'<sup>176</sup> that also involves considerations of strategic context. Equally, determining what constitutes an 'armed attack' involves highly complex political considerations that are not capable of being pre-programmed into an

ACC.<sup>177</sup> As a matter of principle, if Chaos was authorised by state A to make these assessments, was technically capable of determining that the cyber operation was attributable to state B, and could be programmed to assess the circumstances as constituting an armed attack, then the necessary and proportionate use of the ACC in response could be assessed as a lawful exercise of self-defence.

Additional complexity would arise from the use of Chaos if it occurred in anticipatory self-defence. For example, consider a scenario in which Chaos discovers malicious software in state A's systems attributable to state B at a time of increased tension between the two states. This malicious software provides state B with backdoor access to the critical infrastructure of state A, allowing it to, for example, shut down power grids or to open the flood gates of dams. While some states, including Australia, maintain that states may act in anticipatory self-defence in the cyber context, developing ACCs with the capability to do so automatically would be extremely problematic. In addition to establishing attribution, Chaos would need to be capable of assessing the likelihood of particular effects occurring; it would need to take into account a range of non-technical factors such as the relations between the states and other political and strategic factors; and it would need the capacity to assess the requirements of necessity and proportionality before responding with a use of force. Establishing 'necessity' would be particularly problematic, especially where non-forceful means were available such as passive defence measures (e.g. deleting the malicious software). These complexities increase the risk of ACCs used as active cyber defence measures causing effects that are in violation of the law. The variables are far greater than those which arise in offensive cyber operations where the capability is designed for a particular operation with advance knowledge of its intended or expected effects.

---

## Conclusion

Various states are engaging in malicious cyber activities in pursuit of their national security objectives.<sup>178</sup> AI techniques are increasingly being used to enable and support states to defend against these activities, and to support their offensive cyber operations. Among them, Australia is investing in its cyber capabilities, including ACCs. This measure supports the DSR's recommendation that Defence strengthen its cyber capabilities so it can deliver broad and responsive capabilities to support ADF operations.<sup>179</sup> Given Australia's investment in ACCs, it must ensure these technologies are used in a lawful and responsible way so it can maintain its commitment to upholding international law and advancing responsible state behaviour in cyberspace.<sup>180</sup>

This paper has examined ACCs under international law to demonstrate the legal parameters for the use of these technologies. It first showed how states have agreed on the general application of international law to their conduct in cyberspace. There is agreement that cyber operations constituting a use of force will violate article 2(4) of the UN Charter unless conducted in self-defence, and that coercive cyber operations directed at matters that states have the sovereign right to decide freely will violate the non-intervention principle. Further, most states agree that cyber operations can violate sovereignty independent of these rules. But the thresholds at which cyber operations violate international law remain contested. These issues will not be resolved until more states develop settled positions on how they consider the law applies to their cyber activities, supported by evidence of general practice among states that will inform the emergence of customary international law on this topic.

The core legal issues surrounding the use of ACCs under international law on the use of force, self-defence, sovereignty and intervention are not novel. However, the use of ACCs in cyber operations does increase the risk of violations of international law, particularly where AI-enabled ACCs are used. The paper has illustrated this point with reference to the use of ACCs in offensive cyber operations and as active cyber defence measures.

In relation to ACCs used in offensive cyber operations, these capabilities can be designed to operate consistently with the law. This is because the law

is primarily concerned with the effects of a cyber operation, and because these ACCs involve capabilities developed for a particular operation with advance knowledge of its intended or expected effects.<sup>181</sup> However, where ACCs used in this context have AI capabilities, then their operation may be less predictable and this increases the risk of unintended effects. These unintended effects may have legal consequences where they occur in the territory of another state and have the potential to result in violations of international law. These risks are exacerbated by the development of ACCs capable of automatic hack-backs. These systems carry a heightened risk of violating international law, due to their need to be capable of determining what the relevant legal frameworks are, and making assessments involving a range of legal and non-legal factors based on technical and non-technical information. For these reasons their use involves a higher risk of violations of the law compared to ACCs used in offensive cyber operations.

States developing or deploying ACCs must ensure they are used consistently with international law. While the law is primarily focused on the effects caused by ACCs, to minimise the risk of legal violation, states must also consider the expected effects and foreseeable unintended effects of these systems. There should likewise be proper safeguards in place, such as rigorous testing and verification,<sup>182</sup> to ensure the effects of these capabilities will be consistent with international law. Where the law is uncertain, states should promote the responsible use of ACCs. Measures could include temporal constraints, such as the ACC 'erasing' itself after a period of time, and spatial constraints to limit unnecessary proliferation and indiscriminate effects.<sup>183</sup> States should also advance the responsible use of ACCs in their cyber strategies and national positions on how they consider international law to apply in the cyber context.

Finally, it is important to acknowledge that international law is most concerned with the nature of the expected or anticipated effects caused by the use of ACCs, and not the technical means through which those effects are achieved. ACCs are tools programmed by human beings; they implement decisions made by human beings; it is human beings who decide to use these capabilities; and it is those human beings who must ensure their use is in accordance with international law.<sup>184</sup>

---

## About the Author

Dr Samuli Haataja is a Senior Lecturer at Griffith Law School, Griffith University. His research explores international law and cyber security, with a focus on state-sponsored cyber operations under public international law. His book *Cyber Attacks and International Law on the Use of Force: The Turn to Information Ethics* was published by Routledge in 2019, and he has published in various international law and technology journals.

---

## Endnotes

- 1 Australian Government, *National Defence: Defence Strategic Review* (Canberra: Commonwealth of Australia, 2023), p. 64.
- 2 Ibid, p. 64.
- 3 Australian Signals Directorate, 'REDSPICE: Resilience—Effects—Defence—Space—Intelligence—Cyber—Enablers', at: <https://www.asd.gov.au/about/what-we-do/redspice>.
- 4 Australian Signals Directorate, 'REDSPICE: A Blueprint for Growing ASD's Capabilities', p. 13, at: <https://www.asd.gov.au/sites/default/files/2022-05/ASD-REDSPICE-Blueprint.pdf>.
- 5 Pat Conroy, 'The Advanced Strategic Capabilities Accelerator Is Up and Running to Drive Defence Innovation', press release, 1 July 2023, Minister for Defence, at: <https://www.minister.defence.gov.au/media-releases/2023-07-01/advanced-strategic-capabilities-accelerator-and-running-drive-defence-innovation>.
- 6 Australian Government, Defence Science and Technology Group, 'Opportunity: Call for Submissions: Cyber Autonomy Gym for Experimentation (CAGE) Challenge 3', at: <https://www.dst.defence.gov.au/opportunity/call-submissions-cyber-autonomy-gym-experimentation-cage-challenge-3>.
- 7 Australian Government, *2023–2030 Australian Cyber Security Strategy* (Canberra: Commonwealth of Australia, 2023), p. 56.
- 8 United Kingdom Ministry of Defence, *Joint Doctrine Note 2/11: The UK Approach to Unmanned Aircraft Systems*, JDN 2-11 (Swindon: United Kingdom Ministry of Defence, 2011); United States Department of Defence, *Unmanned Systems Integrated Roadmap FY2011-2036*, 11-S-3613 (Washington: United States Department of Defence, 2011).
- 9 Human Rights Watch, *Losing Humanity: The Case against Killer Robots* (Washington: International Human Rights Clinic, 2012); Human Rights Watch, *Making the Case: The Dangers of Killer Robots and the Need for a Preemptive Ban* (Washington: International Human Rights Clinic, 2016).
- 10 Human Rights Watch, *Losing Humanity*, pp. 30–36.
- 11 *Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects (and Protocols)* (as Amended on 21 December 2001), opened for signature 10 October 1980, 1342 UNTS 137 (entered into force 2 December 1983).
- 12 *Report of the 2014 Informal Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS)*, UN Doc CCW/MSP/2014/3 (10 June 2014); *Report of the 2015 Informal Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS)*, UN Doc CCW/MSP/2015/3 (1 June 2015); *Report of the 2016 Informal Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS)*, UN Doc CCW/CONF.V/2 (10 June 2016).
- 13 *Report of the 2023 Session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems*, UN Doc CCW/GGE.1/2023/2 (24 May 2023), p. 4.
- 14 United Nations General Assembly, *Developments in the Field of Information and Telecommunications in the Context of International Security*, GA Res 53/70, UN Doc A/RES/53/70 (4 January 1999).



- 15 United Nations Secretary-General, *Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security: Report of the Secretary-General*, UN Doc A/60/202 (5 August 2005), p. 12.
- 16 United Nations Secretary-General, *Report of the Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security*, UN Doc A/68/98 (24 June 2013), p. 2; UN Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security, *Report of the Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security*, UN Doc A/70/174 (22 July 2015); UN Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security, *Report of the Group of Governmental Experts on Advancing Responsible State Behaviour in Cyberspace in the Context of International Security* (28 May 2021, advance copy), pp. 13–14, at: <https://front.un-arm.org/wp-content/uploads/2021/06/final-report-2019-2021-gge-1-advance-copy.pdf>.
- 17 *Resolution Adopted by the General Assembly on 5 December 2018*, UNGA Res 73/27, UN Doc A/RES/73/27 (11 December 2018), p. 5.
- 18 Open-ended Working Group on Developments in the Field of Information and Telecommunications in the Context of International Security, *Final Substantive Report*, UN Doc A/AC.290/2021/CRP.2 (10 March 2021).
- 19 United Nations Institute for Disarmament Research (UNIDIR), *The Weaponization of Increasingly Autonomous Technologies: Autonomous Weapon Systems and Cyber Operations*, UNIDIR Report No. 7 (Geneva: UNIDIR, 2017), pp. 1–3, at: <https://unidir.org/publication/weaponization-increasingly-autonomous-technologies-autonomous-weapon-systems-and-cyber/>.
- 20 United States Department of Defence, *Directive Number 3000.09: Autonomy in Weapon Systems*, Homeland Security Digital Library, 21 November 2012), p. 2, at: <https://www.hsdl.org/?abstract&did=726163>.
- 21 Though, as Paul Scharre notes, this was largely because of the added complexity that discussion of cyber operations would have added to efforts to formulate this policy document. See Paul Scharre, *Army of None* (New York: W.W. Norton, 2018), p. 228.
- 22 US National Security Commission on Artificial Intelligence, *Final Report* (Arlington: US National Security Commission on Artificial Intelligence, 2021), p. 283, at: <https://www.nscai.gov/wp-content/uploads/2021/03/Full-Report-Digital-1.pdf>.
- 23 See, however, Chair of the Open-ended Working Group, *OEWG Chair's Letter on the Summary Report of the Informal Intersessional Consultative Meeting from 2–4 December 2019*, Permanent Mission of Switzerland to the United Nations, 28 January 2020, p. 8, at: <https://unoda-web.s3.amazonaws.com/wp-content/uploads/2020/01/200128-OEWG-Chairs-letter-on-the-summary-report-of-the-informal-intersessional-consultative-meeting-from-2-4-December-2019.pdf>. Here one of the issues raised in relation to trust and confidence building measures is that 'developments in new technologies, such as increasingly autonomous cyber operations, would significantly reduce the predictability of the technology itself and thus constitute a source for anxiety and mistrust'.
- 24 See, for example, Enn Tyugu, 'Command and Control of Cyber Weapons', in Christian Czosseck, Rain Ottis and Katharina Ziolkowski (eds), *2012 4th International Conference on Cyber Conflict* (Tallinn: NATO CCD COE Publications, 2012); Caitríona H Heintz, 'Artificial (Intelligent) Agents and Active Cyber Defence: Policy Implications', in Pascal Brangetto, Markus Maybaum and Jan Stinissen (eds), *2014 6th International Conference on Cyber Conflict* (Tallinn: NATO CCD COE, 2014).

- 25 Alessandro Guarino, 'Autonomous Intelligent Agents in Cyber Offence', in K Podins, J Stinissen and M Maybaum (eds), *2013 5th International Conference on Cyber Conflict* (Tallinn: NATO CCD COE Publications, 2013); Jarrod Stuard and James McGhee, 'Is Skynet the Answer? Rules for Autonomous Cyber Response Capabilities', in Misty Blowers (ed.), *Evolution of Cyber Technologies and Operations to 2035* (New York: Springer, 2015), pp. 151–162.
- 26 Rain Liivoja, Maarja Naagel and Ann Väljataga, *Autonomous Cyber Capabilities under International Law*, NATO Cooperative Cyber Defence Centre of Excellence Working Paper (Tallinn: NATO CCDCOE, 2020), pp. 1–49, at: <https://ccdcoe.org/library/publications/autonomous-cyber-capabilities-under-international-law>; Rain Liivoja and Ann Väljataga (eds), *Autonomous Cyber Capabilities under International Law* (Tallinn: NATO CCDCOE, 2021). See also François Delerue, *Cyber Operations and International Law* (Cambridge: Cambridge University Press, 2020), pp. 157–65; Nicholas Tsagourias and Russell Buchan, 'Automatic Cyber Defence and the Laws of War', *German Yearbook of International Law* 60 (2018): 203; Francis Grimal and Jae Sundaram, 'Cyber Warfare and Autonomous Self-Defence', *Journal on the Use of Force and International Law* 4, no. 2 (2017): 312.
- 27 Michael Schmitt, *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations*, 2nd edn (Cambridge: Cambridge University Press, 2017).
- 28 See Dan Efrony and Yuval Shany, 'A Rule Book on the Shelf? Tallinn Manual 2.0 on Cyberoperations and Subsequent State Practice', *American Journal of International Law* 112, no. 4 (2018): 583.
- 29 Ann Väljataga and Rain Liivoja, 'Cyber Autonomy and International Law: An Introduction', in Rain Liivoja and Ann Väljataga (eds), *Autonomous Cyber Capabilities under International Law* (Tallinn: NATO CCDCOE, 2021), p. 2.
- 30 See *Chair's Summary—First 2024 Session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons System*, UN Doc CCW/GGE.1/2024/WP.6 (3 April 2024) pp. 2–3.
- 31 *A Concise Oxford Dictionary of Politics and International Relations*, 4th edn (Oxford University Press, 2018), s.v. 'autonomy'.
- 32 *APA Dictionary of Psychology*, s.v. 'autonomy', accessed November 2021, at: <https://dictionary.apa.org/autonomy>.
- 33 *The Oxford Dictionary of Philosophy*, 3rd edn (Oxford University Press, 2016), s.v. 'autonomy/heteronomy'.
- 34 Tim Smithers, 'Autonomy in Robots and Other Agents', *Brain & Cognition* 34, no. 1 (1997): 88–106.
- 35 Peter Norvig and Stuart J Russell, *Artificial Intelligence: A Modern Approach*, 3rd edn (New Jersey: Prentice Hall, 2010), p. 39.
- 36 *Ibid.*, p. 39.
- 37 Philip Brey and Johnny Hartz Søraker, 'Philosophy of Computing and Information Technology', in Anthonie Meijers (ed.), *Philosophy of Technology and Engineering Sciences* (Oxford: Elsevier Science and Technology, 2009), p. 1373.
- 38 Willem Haselager, 'Robotics, Philosophy and the Problems of Autonomy', in Itiel E Dror (ed.), *Cognitive Technologies and the Pragmatics of Cognition* (Philadelphia: John Benjamins Publishing Company, 2007), p. 74.
- 39 Patrick Lin, George Bekey and Keith Abney, *Autonomous Military Robotics: Risk, Ethics, and Design*, US Department of Navy, Office of Naval Research Report (San Luis Obispo: US Department of Navy, 2008), p. 4, at: [https://digitalcommons.calpoly.edu/cgi/viewcontent.cgi?article=1001&context=phil\\_fac](https://digitalcommons.calpoly.edu/cgi/viewcontent.cgi?article=1001&context=phil_fac).

- 40 Stan Franklin and Art Graesser, 'Is It an Agent, or Just a Program? A Taxonomy for Autonomous Agents', in Jörg P Müller, Michael J Wooldridge and Nicholas R Jennings (eds), *Intelligent Agents III: Agent Theories, Architectures, and Languages* (Budapest: Springer 1997), p. 25. This is a broad definition which captures biological agents (like human beings and animals), robotic agents and software agents. See Franklin and Graesser, 'Is It an Agent, or Just a Program? A Taxonomy for Autonomous Agents', pp. 25–27, 31.
- 41 Tim McFarland, *Autonomous Weapon Systems and the Law of Armed Conflict* (Oxford: Cambridge University Press, 2020), p. 35.
- 42 Tim McFarland, 'The Concept of Autonomy', in Rain Liivoja and Ann Väljataga (eds), *Autonomous Cyber Capabilities under International Law* (Tallinn: NATO CCDCOE, 2021), pp. 12–36.
- 43 McFarland, 'The Concept of Autonomy', p. 18.
- 44 Ibid., pp. 27–28.
- 45 McFarland, *Autonomous Weapon Systems and the Law of Armed Conflict*, p. 35.
- 46 Estonia and Finland, *Categorizing Lethal Autonomous Weapons Systems: A Technical and Legal Perspective to Understanding LAWS (GGE LAWS Working Paper)*, UN Doc CCW/GGE.2/2018/WP.2 (24 August 2018), p. 3. ('[Independence] is undesirable but also highly unlikely in the foreseeable future as it would require human-like or superhuman AI, available beyond the singularity point'.) Scharre likens predicting if and when software that is capable of creating new code that can work toward a goal or modifying their goal will be invented to predicting if and when time travel will be invented. See Scharre, *Army of None*, p. 226. Guarino writes that 'for an agent to be truly intelligent, the internal knowledge and the utility function itself should change over time responding to the experience acquired, or, in other words, an autonomous agent should learn from experience. This can even include modifications of the goal—the target—and can have deep ramifications for autonomous agents employed in cyber offence operations'. Guarino, 'Autonomous Intelligent Agents in Cyber Offence', p. 380.
- 47 See McFarland, *Autonomous Weapon Systems and the Law of Armed Conflict*, pp. 41–47.
- 48 Ibid., p. 43.
- 49 Liivoja, Naagel and Väljataga, *Autonomous Cyber Capabilities under International Law*, p. 11.
- 50 For a legal analysis based on this possibility, see Samuli Haataja, 'Autonomous Cyber Capabilities and Attribution in the Law of State Responsibility', in Rain Liivoja and Ann Väljataga (eds), *Autonomous Cyber Capabilities under International Law* (Tallinn: NATO CCDCOE, 2021), pp. 280–290.
- 51 McFarland, 'The Concept of Autonomy', p. 20.
- 52 Ibid., p. 21.
- 53 Ibid., p. 22. See also Tanel Tammet, 'Autonomous Cyber Defence Capabilities', in Rain Liivoja and Ann Väljataga (eds), *Autonomous Cyber Capabilities under International Law* (Tallinn: NATO CCDCOE, 2021), pp. 40–41.
- 54 Josh Halliday and Julian Borger, 'Nuclear Plants Likely Target of Foiled Cyber Sabotage', *The Guardian*, 25 September 2010, at: <https://www.theguardian.com/world/2010/sep/25/iran-cyber-hacking-nuclear-plants>; David Sanger, 'Obama Order Sped up Wave of Cyberattacks against Iran', *The New York Times*, 1 June 2012, at: <http://www.nytimes.com/2012/06/01/world/middleeast/obama-ordered-wave-of-cyberattacks-against-iran.html>.

- 55 Nicolas Falliere, Liam Murchu and Eric Chien, *W32.Stuxnet Dossier: Version 1.4* (Cupertino: Symantec, 2011), pp. 41–43, at: [https://www.symantec.com/content/en/us/enterprise/media/security\\_response/whitepapers/w32\\_stuxnet\\_dossier.pdf](https://www.symantec.com/content/en/us/enterprise/media/security_response/whitepapers/w32_stuxnet_dossier.pdf).
- 56 Ibid., pp. 14, 48–49.
- 57 Joby Warrick, 'Iran's Natanz Nuclear Facility Recovered Quickly from Stuxnet Cyberattack', *The Washington Post*, 16 February 2011, at: <http://www.washingtonpost.com/wp-dyn/content/article/2011/02/15/AR2011021506501.html>.
- 58 Falliere, O'Murchu and Chien, *W32.Stuxnet Dossier*, p. 3.
- 59 Thanassis Avgerinos et al., 'The Mayhem Cyber Reasoning System', *IEEE Security & Privacy* 16, no. 2 (2018): 52, 53. See also Steve Lohr, 'Stepping up Security for an Internet-of-Things World', *The New York Times*, 16 October 2016, at: <https://www.nytimes.com/2016/10/17/technology/security-internet.html>.
- 60 Avgerinos et al., 'The Mayhem Cyber Reasoning System', p. 58.
- 61 Ibid., pp. 56–57.
- 62 Alexander Kott et al., *Autonomous Intelligent Cyber-Defense Agent (AICA) Reference Architecture Release 2.0*, ARL-SR-0421 (Adelphi: Combat Capabilities Development Command Army Research Laboratory, 2019), p. 1, at: <https://arxiv.org/ftp/arxiv/papers/1803/1803.10664.pdf>.
- 63 Kott et al., *Autonomous Intelligent Cyber-Defense Agent (AICA) Reference Architecture Release 2.0*, p. 2.
- 64 Ibid., p. 3. This is a hypothetical scenario and, as the report notes, an AICA of this kind is subject to 'very substantial' research challenges. Kott et al., *Autonomous Intelligent Cyber-Defense Agent (AICA) Reference Architecture Release 2.0*, p. 79.
- 65 US National Security Commission on Artificial Intelligence, *Final Report*, p. 51.
- 66 Australian Signals Directorate, *ASD Cyber Threat Report 2022–2023* (Canberra: Australian Signals Directorate, 2023), p. 57.
- 67 Ben Buchanan et al., *Automating Cyber Attacks: Hype and Reality* (Washington: Centre for Security and Emerging Technology, 2020), p. 25, at: <https://cset.georgetown.edu/publication/automating-cyber-attacks>.
- 68 Ibid., pp. 17–18.
- 69 Ibid., p. 18.
- 70 International Committee of the Red Cross, *Artificial Intelligence and Machine Learning in Armed Conflict: A Human-Centred Approach* (Geneva: International Committee of the Red Cross, 2016), p. 3, at: <https://www.icrc.org/en/document/artificial-intelligence-and-machine-learning-armed-conflict-human-centred-approach>. Others have also raised concerns about 'independent' ACCs – see Tyugu, 'Command and Control of Cyber Weapons', p. 334: 'In the long run, there exists the danger that intelligent agents may become too independent and they will perform unexpected and unwanted (harmful) actions'. See also Guarino, 'Autonomous Intelligent Agents in Cyber Offence', p. 387.
- 71 It is important to note, however, that even complex systems without autonomous capabilities can operate in unpredictable ways. See Estonia and Finland, *Categorizing Lethal Autonomous Weapons Systems*, p. 2.
- 72 Paul Scharre, *Autonomous Weapons and Operational Risk*, Ethical Autonomy Project (Washington: Center for a New American Security, 2016), p. 8.
- 73 Ibid., p. 11.
- 74 Ibid., pp. 13–15.

- 75 Scharre, *Army of None*, pp. 229–230.
- 76 Open-ended Working Group on Developments in the Field of Information and Telecommunications in the Context of International Security, *Final Substantive Report*, pp. 5–6.
- 77 UN Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security, *Report of the Group of Governmental Experts on Advancing Responsible State Behaviour in Cyberspace in the Context of International Security*, pp. 13–14.
- 78 See *Resolution Adopted by the General Assembly on 20 December 2018*, UNGA Res 73/203, UN Doc A/RES/73/203 (20 December 2018).
- 79 *Legality of the Threat or Use of Nuclear Weapons (Advisory Opinion)* [1996] ICJ Rep, 226, 244.
- 80 Department of Foreign Affairs and Trade, *Australia's International Cyber Engagement Strategy* (Canberra: Commonwealth of Australia, 2017), p. 90; New Zealand Foreign Affairs and Trade, *The Application of International Law to State Activity in Cyberspace*, Department of the Prime Minister and Cabinet: Te Tari o Pirimia me te Komiti Matua, 1 December 2020, at: <https://dpmc.govt.nz/publications/application-international-law-state-activity-cyberspace>; Government of the Netherlands, 'Appendix: International Law in Cyberspace', in *Letter to the Parliament on the International Legal Order in Cyberspace* (Amsterdam: Government of the Netherlands, 2019), pp. 3–4, at: <https://www.government.nl/documents/parliamentary-documents/2019/09/26/letter-to-the-parliament-on-the-international-legal-order-in-cyberspace>; The Federal Government of Germany, *On the Application of International Law in Cyberspace* (Berlin: German Federal Foreign Office and German Federal Ministry of Defence, 2021), p. 6, at: <https://www.auswaertiges-amt.de/blob/2446304/2ae17233b62966a4b7f16d50ca3c6802/on-the-application-of-international-law-in-cyberspace-data.pdf>; Switzerland Federal Department of Foreign Affairs, *Switzerland's Position Paper on the Application of International Law in Cyberspace* (Bern: Federal Department of Foreign Affairs, 2021), p. 4, at: [https://www.eda.admin.ch/dam/eda/en/documents/aussenpolitik/voelkerrecht/20210527-Schweiz-Annex-UN-GGE-Cybersecurity-2019-2021\\_EN.pdf](https://www.eda.admin.ch/dam/eda/en/documents/aussenpolitik/voelkerrecht/20210527-Schweiz-Annex-UN-GGE-Cybersecurity-2019-2021_EN.pdf).
- 81 *Charter of the United Nations* (adopted 26 June 1945, entered into force 24 October 1945) (1945) 1 UNTS XVI, art 51.
- 82 *Oil Platforms (Iran v United States of America)* (2003) ICJ Rep 161, 198.
- 83 *Case Concerning Military and Paramilitary Activities in and Against Nicaragua (Nicaragua v United States of America)* (Merits) [1986] ICJ Rep 14, 101.
- 84 Harold Hongju Koh, 'International Law in Cyberspace', *Harvard International Law Journal Online* 54 (2012): 1, 7–8.
- 85 See Michael Wood, 'The Caroline Incident – 1837', in Tom Ruys, Olivier Corten, Alexandra Hofer (eds), *The Use of Force in International Law: A Case-Based Approach* (Oxford: Oxford University Press, 2018), p. 5.
- 86 James Crawford, *Brownlie's Principles of Public International Law*, 9th edn (Oxford: Oxford University Press, 2019), p. 725.
- 87 Schmitt, *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations*, p. 339.
- 88 *Ibid.*, p. 339.
- 89 Department of Foreign Affairs and Trade, *Australia's International Cyber Engagement Strategy (2019 International Law Supplement)* (Canberra: Commonwealth of Australia, 2017).

- 90 Ministry for Foreign Affairs, *Finland Published Its Positions on Public International Law in Cyberspace* (Helsinki: Finnish Government, 2020), at: <https://valtioneuvosto.fi/en/-/finland-published-its-positions-on-public-international-law-in-cyberspace> ('it is widely recognized that such a qualification depends on the consequences of a cyberattack. Most commentators agree that a cyberattack which is comparable to an armed attack in terms of its extent and impacts equates to an armed attack, and self-defence is justified as response').
- 91 *Developments in the Field of Information and Telecommunications in the Context of International Security: Report of the Secretary-General*, UN Doc A/74/120 (24 June 2019), p. 22. ('In the view of France, if the scope or impact of a major cyberattack perpetrated by a State, or by non-State actors acting under the supervision or instruction of a State, reaches a sufficient threshold (such as substantial loss of life, significant material damage, insufficient critical infrastructure with significant consequences), and is attributable to a State, that could constitute 'armed aggression' under article 51 of the Charter and thus justify a claim of self-defence'.)
- 92 The Federal Government of Germany, *On the Application of International Law in Cyberspace*, p. 15.
- 93 Islamic Republic of Iran, *Declaration of General Staff of the Armed Forces of the Islamic Republic of Iran Regarding International Law Applicable to the Cyberspace* (Tehran: Islamic Republic of Iran, 2020), at: <https://nournews.ir/En/News/53144/General-Staff-of-Iranian-Armed-Forces-Warns-of-Tough-Reaction-to-Any-Cyber-Threat>.
- 94 Government of the Netherlands, *Appendix: International Law in Cyberspace*, pp. 8–9.
- 95 Koh, 'International Law in Cyberspace', p. 4.
- 96 Jeremy Wright, 'Cyber and International Law in the 21st Century' (speech at Chatham House, the Royal Institute of International Affairs), Gov.UK, 23 May 2018, at: <https://www.gov.uk/government/speeches/cyber-and-international-law-in-the-21st-century>.
- 97 Switzerland Federal Department of Foreign Affairs, *Switzerland's Position Paper on the Application of International Law in Cyberspace*, p. 4.
- 98 Inter-American Judicial Committee, *International Law and State Cyber Operations* (Inter-American Judicial Committee, 2020), pp. 34–35, at: [http://www.oas.org/en/sla/iajc/docs/International\\_Law\\_and\\_State\\_Cyber\\_Operations\\_publication.pdf](http://www.oas.org/en/sla/iajc/docs/International_Law_and_State_Cyber_Operations_publication.pdf).
- 99 NATO, *Allied Joint Doctrine for Cyberspace Operations*, Allied Joint Publication-3.20 (Brussels: NATO Standardisation Office, 2020), p. 20, at: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/899678/doctrine\\_nato\\_cyberspace\\_operations\\_ajp\\_3\\_20\\_1\\_.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/899678/doctrine_nato_cyberspace_operations_ajp_3_20_1_.pdf).
- 100 New Zealand Foreign Affairs and Trade, *The Application of International Law to State Activity in Cyberspace*.
- 101 Wright, *Cyber and International Law in the 21st Century*.
- 102 George Brandis, 'The Right of Self-Defence against Imminent Armed Attack in International Law' (lecture delivered at the University of Queensland, Brisbane, 11 April 2017), p. 8, at: <https://law.uq.edu.au/blog/2017/05/developments-international-law-self-defence-against-imminent-armed-attack>. See also French Ministry of the Armies, *International Law Applied to Operations in Cyberspace* (French Ministry of the Armies, 2019), p. 9, at: <https://www.defense.gouv.fr/content/download/567648/9770527/file/international-law+applied+to+operations+in+cyberspace.pdf>.
- 103 Brandis, 'The Right of Self-Defence against Imminent Armed Attack in International Law', p. 8.

- 104 Ibid., p. 8. Similarly, according to NATO's Allied Joint Doctrine for Cyberspace Operations, non-destructive cyber operations, such as those causing temporary disruptions of service or involving information gathering where they are conducted to enable a conventional threat—including an imminent threat of one—could be considered to trigger the right of self-defence. See NATO, *Allied Joint Doctrine for Cyberspace Operations*, p. 20.
- 105 See, for example, Federal Government of Germany, *On the Application of International Law in Cyberspace*, p. 15.
- 106 *The Declaration on Principles of International Law concerning Friendly Relations and Co-operation among States*, UNGA Res 2625 (XXV), UN Doc A/RES/2625(XXV), p. 123 (24 October 1970).
- 107 Robert Y Jennings and Arthur Watts (eds), *Oppenheim's International Law, Volume 1, Peace*, 9th edn (Essex: Longman, 1992), p. 430, quoted in David J Harris, *Cases and Materials on International Law*, 7th edn (Thomson Reuters, 2010), p. 743. See also Lori F Damrosch, 'Politics across Borders: Nonintervention and Nonforcible Influence over Domestic Affairs', *American Journal of International Law* 83, no. 1 (1989): 1, 3.
- 108 According to the ICJ activities involving the use of force would be 'wrongful in the light of both the principle of non-use of force, and that of non-intervention'. *Case Concerning Military and Paramilitary Activities in and Against Nicaragua (Nicaragua v United States of America)* (Merits) [1986] ICJ Rep 14, 108. This was affirmed by the ICJ in *Armed Activities on the Territory of the Congo (Democratic Republic of the Congo v Uganda)* [2005] ICJ Rep 168, 227.
- 109 *Case Concerning Military and Paramilitary Activities in and Against Nicaragua (Nicaragua v United States of America)* (Merits) [1986] ICJ Rep 14, 108.
- 110 Maziar Jamnejad and Michael Wood, 'The Principle of Non-Intervention', *Leiden Journal of International Law* 22, no. 2 (2012): 345, 347.
- 111 UN Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security, *Report of the Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security*, p. 12.
- 112 Schmitt, *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations*, p. 312.
- 113 Robert Y Jennings and Arthur Watts (eds), *Oppenheim's International Law, Volume 1, Peace*, pp. 430–431, quoted in Schmitt, *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations*, p. 317.
- 114 Schmitt, *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations*, p. 318.
- 115 Harriet Moynihan, *The Application of International Law to State Cyberattacks: Sovereignty and Non-Intervention*, Chatham House Research Paper (Chatham House, The Royal Institute of International Affairs, 2019), p. 30, at: <https://www.chathamhouse.org/2019/12/application-international-law-state-cyberattacks>.
- 116 Schmitt, *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations*, p. 318.
- 117 Department of Foreign Affairs and Trade, *Australia's International Cyber Engagement (Strategy 2019 International Law Supplement)* (Canberra: Commonwealth of Australia, 2017). See also Federal Government of Germany, *On the Application of International Law in Cyberspace*, p. 5.
- 118 See Moynihan, *The Application of International Law to State Cyberattacks*, p. 34.

- 119 *Case Concerning Military and Paramilitary Activities in and Against Nicaragua (Nicaragua v United States of America)* (Merits) [1986] ICJ Rep 14, 108.
- 120 Department of Foreign Affairs and Trade, *Australia's International Cyber Engagement Strategy (2019 International Law Supplement)*.
- 121 Wright, *Cyber and International Law in the 21st Century*; New Zealand Foreign Affairs and Trade, *The Application of International Law to State Activity in Cyberspace*.
- 122 Moynihan, *The Application of International Law to State Cyberattacks*, p. 11.
- 123 *Ibid.*, pp. 12–13.
- 124 Schmitt, *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations*, p. 17.
- 125 Government of the Netherlands, *Appendix: International Law in Cyberspace*, p. 2.
- 126 *Developments in the Field of Information and Telecommunications in the Context of International Security: Report of the Secretary-General*, UN Doc A/74/120 (24 June 2019), p. 22.
- 127 Government of Austria, *Comments on Pre-Draft Report of the OEWG—ICT* (UN Office of Disarmament Affairs, 2020), at: <https://front.un-arm.org/wp-content/uploads/2020/04/comments-by-austria.pdf>.
- 128 Government of Czech Republic, *Statement by Richard Kadlčák at Open-ended Working Group on Developments in the Field of Information and Telecommunications in the Context of International Security of the First Committee of the General Assembly of the United Nations* (New York, 11 February 2020), at: [https://www.nukib.cz/download/publications\\_en/CZ%20Statement%20-%20OEWG%20-%20International%20Law%2011.02.2020.pdf](https://www.nukib.cz/download/publications_en/CZ%20Statement%20-%20OEWG%20-%20International%20Law%2011.02.2020.pdf).
- 129 Finnish Government, Ministry for Foreign Affairs, *Finland Published Its Positions on Public International Law in Cyberspace*.
- 130 Islamic Republic of Iran, 'Declaration of General Staff of the Armed Forces of the Islamic Republic of Iran Regarding International Law Applicable to the Cyberspace', *Nour News*, July 2020, at: <https://hournews.ir/En/News/53144/General-Staff-of-Iranian-Armed-Forces-Warns-of-Tough-Reaction-to-Any-Cyber-Threat>.
- 131 New Zealand Foreign Affairs and Trade, *The Application of International Law to State Activity in Cyberspace*.
- 132 The Federal Government of Germany, *On the Application of International Law in Cyberspace*, p. 2.
- 133 Switzerland Federal Department of Foreign Affairs, *Switzerland's Position Paper on the Application of International Law in Cyberspace*, p. 2.
- 134 Inter-American Judicial Committee, *International Law and State Cyber Operations*, p. 52.
- 135 Gary P Corn and Robert Taylor, 'Sovereignty in the Age of Cyber', *AJIL Unbound* 111 (2017): 209.
- 136 Wright, *Cyber and International Law in the 21st Century*; United Kingdom of Great Britain and Northern Ireland, *Non-Paper on Efforts to Implement Norms of Responsible State Behaviour in Cyberspace, as Agreed in UN Group of Government Expert Reports of 2010, 2013 and 2015* (UN Office of Disarmament Affairs, 2019), p. 11, at: <https://unoda-web.s3.amazonaws.com/wp-content/uploads/2019/09/uk-un-norms-non-paper-oewg-submission-final.pdf>.
- 137 Jennifer M O'Connor, *International Law Framework for Employing Cyber Capabilities in Military Operations*, Memorandum (2017). This approach has subsequently been



detailed further by Gary Corn (former Staff Judge Advocate (General Counsel) to the US Cyber Command) and Robert Taylor (former Principal Deputy General Counsel of the US Department of Defence) in a non-official capacity. See Corn and Taylor, 'Sovereignty in the Age of Cyber'. While the US Department of Defence memo was initially made publicly available, its distribution since then was restricted. See Michael Schmitt and Liis Vihul, 'Respect for Sovereignty in Cyberspace', *Texas Law Review* 95 (2017): 1639, 1641.

- 138 United Nations General Assembly, *Official Compendium of Voluntary National Contributions on the Subject of How International Law Applies to the Use of Information and Communications Technologies by States Submitted by Participating Governmental Experts in the Group of Governmental Experts on Advancing Responsible State Behaviour in Cyberspace in the Context of International Security Established Pursuant to General Assembly Resolution 73/266, GA Res 76/136, UN Doc A/76/136\** (13 July 2021), p. 140.
- 139 United Nations Secretary-General, *Developments in the Field of Information and Telecommunications in the Context of International Security*, p. 22; Islamic Republic of Iran, 'Declaration of General Staff of the Armed Forces of the Islamic Republic of Iran Regarding International Law Applicable to the Cyberspace'. A similar position has been adopted by some academic commentators who argue that cyber operations amount to violations of sovereignty when they penetrate another state's computers or networks even without physical damage or effects in the territory of the state. For example, Russell Buchan adopts this position and maintains that a violation occurs on the basis of the act of 'the unauthorised intrusion into a domain protected by state sovereignty'. Russell Buchan, *Cyber Espionage and International Law* (Oxford: Hart Publishing, 2018), p. 54. François Delerue also argues that '[i]f a cyber operation penetrates the cyber infrastructures in the territory of a foreign State, this would irrefutably constitute a violation of territorial sovereignty'. Delerue, *Cyber Operations and International Law*, p. 214.
- 140 New Zealand Foreign Affairs and Trade, *The Application of International Law to State Activity in Cyberspace*.
- 141 The Tallinn Manual experts were also divided on this. See Schmitt, *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations*, pp. 20–21.
- 142 See Dan Svantesson et al., 'On Sovereignty', *Masaryk University Journal of Law and Technology* 17, no. 1 (2023): 33, 54–56.
- 143 On the debate about the relevance of fault in state responsibility, and the objective and subjective approaches to determining this, see James Crawford and Simon Olleson, 'The Nature and Forms of International Responsibility', in Malcolm D Evans (ed.), *International Law*, 2nd edn (Oxford: Oxford University Press, 2010), pp. 464–465; Sandra Szurek, 'The Notion of Circumstances Precluding Wrongfulness', in James Crawford et al. (eds), *The Law of International Responsibility* (Oxford: Oxford University Press, 2010), p. 433; Martti Koskeniemi, 'Doctrines of State Responsibility', in James Crawford et al. (eds), *The Law of International Responsibility* (Oxford: Oxford University Press, 2010), p. 4951; Brigitte Stern, 'The Elements of an Internationally Wrongful Act', in James Crawford et al. (eds), *The Law of International Responsibility* (Oxford: Oxford University Press, 2010), pp. 209–210.
- 144 James Crawford, *The International Law Commission's Articles on State Responsibility: Introduction, Text and Commentaries* (Cambridge: Cambridge University Press, 2002), p. 84.
- 145 *Case Concerning Military and Paramilitary Activities in and Against Nicaragua (Nicaragua v United States of America)* (Merits) [1986] ICJ Rep 14, 108. ('The element of coercion ... is particularly obvious in the case of an intervention which uses force'.)

- 146 See Olivier Corten, *The Law Against War: The Prohibition on the Use of Force in Contemporary International Law* (London: Bloomsbury Publishing, 2010), pp. 78–80.
- 147 Schmitt, *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations*, p. 331.
- 148 Corten, *The Law Against War*, p. 83.
- 149 Schmitt, *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations*, p. 24.
- 150 Jamnejad and Wood, 'The Principle of Non-Intervention', *Leiden Journal of International Law* 22, no. 2 (2012): 368.
- 151 Nicholas Tsagourias maintains that, in establishing a violation of the non-intervention principle, 'intent is critical, particularly in cyberspace, where operations are often factually indistinguishable, and their effects permeate borders unintentionally'. Nicholas Tsagourias, 'Electoral Cyber Interference, Self-Determination, and the Principle of Non-intervention in Cyberspace', in Dennis Broeders and Bibi van den Berg (eds), *Governing Cyberspace: Behaviour, Power, Diplomacy* (Lanham: Rowman & Littlefield, 2020), p. 55. In slight contrast, Moynihan maintains that, based on the ICJ's decision in the *Nicaragua* case, the intention of the responsible state is 'of little relevance' to establishing a violation of the non-intervention principle. However, she nonetheless provides that 'the coercive behaviour on the part of the perpetrating state will, by its nature, be intentional, and thus the description of coercion discussed above necessarily involves an intention to compel an outcome or conduct'. Moynihan, *The Application of International Law to State Cyberattacks*, p. 32. See also Sean Watts, 'Low-Intensity Cyber Operations and the Principle of Non-Intervention', *Baltic Yearbook of International Law* 14, no. 1 (2014): 137, 158.
- 152 Schmitt, *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations*, pp. 331–332.
- 153 Public international law is largely silent on peacetime espionage conducted by states, except in circumstances in which there is specific prohibition on the activity in question (for example, under the law protecting diplomats and consulates). See Katharina Ziolkowski, 'Peacetime Cyber Espionage—New Tendencies in Public International Law', in Katharina Ziolkowski (ed.), *Peacetime Regime for State Activities in Cyberspace: International Law, International Relations and Diplomacy* (Tallinn: NATO CCDCOE, 2013), pp. 430–445; Russell Buchan, *Cyber Espionage and International Law*, pp. 70–94.
- 154 Attribution is outside the scope of this paper. On attribution of conduct by states using ACCs under international law on state responsibility, see Haataja, 'Autonomous Cyber Capabilities and Attribution in the Law of State Responsibility'. On the complexities of legal and technical attribution in the cyber context generally, see Nicholas Tsagourias and Michael Farrell, 'Cyber Attribution: Technical and Legal Approaches and Challenges', *European Journal of International Law* 31 no. 3 (2020): 941. On legal responsibilities relating to AWS, see Eve Massingham and Simon McKenzie, 'Testing Knowledge: Weapons Reviews of Autonomous Weapons Systems and the International Criminal Trial', in Emma Palmer, Edwin Bikundo, Susan Harris Rimmer and Martin Clark (eds), *Futures of International Criminal Justice* (Abingdon: Routledge, 2022).
- 155 In addition to potential issues under a state's national laws, the use of ACCs in this context may also have implications under international human rights law. For human rights implications around cyberspace generally, see David Fidler, 'Cyberspace and Human Rights', in Nicholas Tsagourias and Russell Buchan (eds), *Research Handbook on International Law and Cyberspace* (Northampton: Edward Elgar, 2015).

- 156 Marjolein Busstra and Wieteke Theeuwen, 'International Law in the Context of Cyber Operations', in Marjolein Busstra et al. (eds), *International Law for a Digitalised World* (Leiden: Asser Press, 2020), pp. 9–10.
- 157 Jay Kesan and Carol Hayes, 'Mitigative Counterstriking: Self-Defense and Deterrence in Cyberspace', *Harvard Journal of Law & Technology* 25, no. 2 (2012): 429, 475.
- 158 See Samuli Haataja, *Cyber Attacks and International Law on the Use of Force: The Turn to Information Ethics* (New York: Routledge, 2019), pp. 146–148; Samuli Haataja and Afshin Akhtar-Khavari, 'Stuxnet and International Law on the Use of Force: An Informational Approach', *Cambridge International Law Journal* 7, no. 1 (2018): 99, 109–111.
- 159 See also Michael Schmitt, 'Autonomous Cyber Capabilities and the International Law of Sovereignty and Intervention', in Rain Liivoja and Ann Väljataga (eds), *Autonomous Cyber Capabilities under International Law* (Tallinn: NATO CCDCOE, 2021), p. 132.
- 160 See Russell Buchan, *Cyber Espionage and International Law*, p. 54; Delerue, *Cyber Operations and International Law*, p. 214.
- 161 New Zealand Foreign Affairs and Trade, *The Application of International Law to State Activity in Cyberspace*.
- 162 Wright, *Cyber and International Law in the 21st Century*.
- 163 Schmitt, 'Autonomous Cyber Capabilities and the International Law of Sovereignty and Intervention', pp. 136–137.
- 164 However, depending on the nature of the effects, the state could be in violation of sovereignty or the use of force.
- 165 See also Delerue, *Cyber Operations and International Law*, pp. 240–241.
- 166 See also Schmitt, 'Autonomous Cyber Capabilities and the International Law of Sovereignty and Intervention', pp. 147–150.
- 167 Alec Tattersall and Damian Copeland, 'Reviewing Autonomous Cyber Capabilities', in Rain Liivoja and Ann Väljataga (eds), *Autonomous Cyber Capabilities under International Law* (Tallinn: NATO CCDCOE, 2021), p. 217.
- 168 Väljataga and Liivoja, 'Cyber Autonomy and International Law: An Introduction', p. 5.
- 169 Only France and Iran have proposed that cyber operations penetrating their systems could constitute violations of sovereignty. See section three of this paper.
- 170 International law is largely silent on espionage in peacetime, meaning that unless there is a prohibition on a specific activity, then states are permitted to freely engage in espionage. See generally Buchan, *Cyber Espionage and International Law*.
- 171 These include countermeasures and necessity; however, these are outside the scope of the paper. On these in relation to cyber operations generally, see Henning Lahmann, *Unilateral Remedies to Cyber Operations: Self-Defence, Countermeasures, Necessity, and the Question of Attribution* (Cambridge: Cambridge University Press, 2020). See also Samuli Haataja, 'Cyber Operations and Collective Countermeasures Under International Law', *Journal of Conflict and Security Law* 25, no. 1 (2020): 33. For an overview of countermeasures and necessity in relation to ACCs, see Schmitt, 'Autonomous Cyber Capabilities and the International Law of Sovereignty and Intervention', pp. 143–146.
- 172 Liivoja, Naagel and Väljataga, *Autonomous Cyber Capabilities under International Law*, p. 24.
- 173 Tsagourias and Buchan, 'Automatic Cyber Defence and the Laws of War', p. 209. See also Tattersall and Copeland, 'Reviewing Autonomous Cyber Capabilities', pp. 247–248.

- 174 Liivoja, Naagel and Väljataga, *Autonomous Cyber Capabilities under International Law*, p. 24. See also Tsagourias and Buchan, 'Automatic Cyber Defence and the Laws of War', pp. 212–218; Schmitt, 'Autonomous Cyber Capabilities and the International Law of Sovereignty and Intervention', p. 150.
- 175 Liivoja, Naagel and Väljataga, *Autonomous Cyber Capabilities under International Law*, p. 24.
- 176 The Federal Government of Germany, *On the Application of International Law in Cyberspace*, p. 15.
- 177 Liivoja, Naagel and Väljataga, *Autonomous Cyber Capabilities under International Law*, p. 24.
- 178 See, for example, Samuli Haataja, 'Cyber operations against critical infrastructure under norms of responsible state behaviour and international law', *International Journal of Law and Information Technology* 30, no. 4 (2022): 423, 424–426.
- 179 Australian Government, *National Defence: Defence Strategic Review*, p. 64.
- 180 Australian Government, *2023–2030 Australian Cyber Security Strategy*, p. 56.
- 181 Väljataga and Liivoja, 'Cyber Autonomy and International Law: An Introduction', p. 5.
- 182 US National Security Commission on Artificial Intelligence, *Final Report*, p. 97.
- 183 See Monica Kaminska, Dennis Broeders, Fabio Cristiano, 'Limiting Viral Spread: Automated Cyber Operations and the Principles of Distinction and Discrimination in the Grey Zone', in T Jančárková, L Lindström, M Signoretti, I Tolga and G Visky (eds), *13th International Conference on Cyber Conflict: Going Viral Proceedings 2021* (Tallinn: NATO CCDCOE, 2021); *Report of the 2023 Session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems*, UN Doc CCW/GGE.1/2023/2 (24 May 2023), p. 4.
- 184 See also Schmitt, 'Autonomous Cyber Capabilities and the International Law of Sovereignty and Intervention', p. 138.





[researchcentre.army.gov.au](https://researchcentre.army.gov.au)