Autonomous Defence Systems, State Practice and Opinio Juris: A Worrying Potentiality for International Law

By Fionn Parker

This paper argues that there is a severe risk the actions of autonomous defence systems (ADS) may be taken as representative of state practice, thereby shaping the scope of what we understand as the legitimate use of force in self-defence under international law (the 'self-defence doctrine'). This doctrine is characterised by many ambiguities. It thus requires further development and clarification by state practice and opinio juris. This paper suggests that the conduct of ADS may come to be the source of this state practice, where the necessary opinio juris is provided by subsequent governmental statements endorsing and owning this conduct or by the systems of control deployed over the ADS. The paper concludes by outlining the various risks associated with this potential influence of ADS on the self-defence doctrine. This essay is pressingly relevant in a world where the development of artificial intelligence far outstrips the pace of law reform. Worryingly, it may be the technological developments such as ADS rather than us shaping the law that regulates them.

Keywords: autonomous defence systems; state practice; opinio juris; self-defence; artificial intelligence.

On 7 October 2023, Hamas militants launched at least 2,000 rockets towards Israel. According to the Israeli military, around 90% of these were intercepted by the self-defence system known as the Iron Dome (IDF Editorial Team 2023). The Iron Dome is a fully automated system that uses radar to detect and intercept incoming short-range rockets. Once a missile is detected, the system calculates its flight path and predicts its target. If the inbound rocket is likely to hit a populated area, the system autonomously launches a counter-missile to destroy it. According to the Israeli military, the Iron Dome was used in a way clearly compatible with the self-defence doctrine enshrined in *Article 51* of the UN Charter and Customary International Law (customary international law); that is, the counter-missiles were in response to an 'armed attack,' and their use was both necessary and proportionate. However, as the nature of war inexorably marches towards increased automation, one can imagine a variety of technologically feasible and strategically desirable alterations to the 'Iron Dome' that are much more contentious. What if, rather than intercepting the missiles,

the automated counterstrikes were launched directly at the source of the inbound attack? What if the automated defence system (ADS) immediately executed retaliatory offensives against high-value targets of the enemy (identified by the system)? These responses are not yet functions of the Iron Dome. However, they are real responses of the Israeli military to the October 7 Hamas attack, justified by the Israeli Government as forms of 'self-defence.' Assuming that these actions will continue to remain the sole ambit of human agents seems naïve in the context of the latest 'global arms race': Artificial Intelligence (AI) weaponisation (Simonite 2017). The automation of self-defence in conflict has already begun, with the scope of application of ADS extremely likely to grow over time.

This essay is not uniquely interested in the broader permissibility of automated warfare and self-defence, which have been explored elsewhere (see, for example, Johnson 2022; Kleczkowska 2023). Instead, it focuses on a more discrete aspect of ADS' use – how they might influence the development of international law. Specifically, I argue that there is a serious risk the actions of ADS may be taken as representative of state practice, thereby shaping the scope of permissible application of the self-defence doctrine. The argument progresses in the following way. Firstly, I outline existing and potential future developments in ADS, noting why further proliferation of this technology seems inevitable. In section two of this essay, I provide an overview of the doctrine of self-defence, focusing on its ambiguities: the definition of an 'armed attack,' assessing what is necessary and proportionate, and whether anticipatory self-defence or self-defence against non-state actors is justifiable. Like much international law, self-defence has an obvious 'core' of application around which lies a broad penumbra of uncertain cases in which it is unclear whether self-defence is justified. This 'core' has been slowly developed over time by, among other things, state practice and opinio juris (the perception of states that they are bound by a particular law). Section three explores how, as ADS technology proliferates, the decisions implemented by such systems may be taken as representative of this state practice and opinio juris. Since ADS are tools of the state deployed to achieve state objectives, statements endorsing or justifying their conduct will likely proliferate, providing opinio juris. Opinio juris may also be identified in the 'systems of control' over ADS. This co-existence of opinio juris and state practice means that the decisions of ADS may become seen as legitimate sources of international law, thereby shaping international doctrines such as that of self-defence. I conclude by exploring three risks associated with this potentiality: the unpredictability of ADS leading to uncontrolled development of international law; the worrying lack of

‘considered’ judgment in ADS decisions; and the ‘dystopian concern’ of having our law developed by machines. In sum, this article is intended to provoke contemplation over a hitherto unconsidered risk of using ADS – the risk that these automated decisions might inadvertently shape the scope of international law.

1 Autonomous Defence Systems (ADS)

Whilst AI technology in the military has been explored and deployed in a wide range of contexts, this essay focuses on a relatively narrow class of emerging technology – ADS. ADS are fully autonomous systems, meaning there is no ‘human in the loop’ to make the ‘trigger pull’ (or any other) decision. ADS are deployed in ‘self-defence’ scenarios, meaning that the external stimuli of ‘armed attacks’ trigger their actions. ADS generally have three key elements – 1) a radar or scanning system to identify and assess threats, 2) an internal system capable of calculating the ‘best’ way to respond to such threats, and 3) defensive/offensive capabilities to nullify or respond to such threats. Examples of ADS include missile defence systems such as the ‘Iron Dome’ and cyberwarfare defence technology such as America’s ‘Einstein’ system. Such systems are not as ‘new’ as we might wish to believe. Indeed, former USSR generals have described the existence of a nuclear counterstrike ADS in the Cold War that would become operative in the instance of a decapitation strike (Ryabikjin 2019). As ‘AI-mania’ sweeps the globe, this technology is expected to proliferate.

ADS are desirable for a range of reasons. Firstly, the speed of warfare is exponentially increasing (Erskine and Miller 2024). This increases the allure of ADS, which can identify and respond to threats much more quickly than any human chain of command. The Qassam rocket, known to be used and produced by Hamas, can fly at up to 720 km/h. When fired from short range (often 10 kilometres or less), most attacks find their target within less than a minute. In the case of a nuclear attack, American operators are believed to have approximately three minutes to assess and confirm initial indications from a warning system before a counterstrike is required to be launched (Lewis 2017). This extreme time pressure to evaluate and develop a response is ameliorated by ADS, which can develop and execute a coherent response in a fraction of the time a human or team of humans can. Secondly, ADS are enticing because they lack human cognitive shortcomings such as emotion, heuristics, and groupthink. Whilst this does not guarantee the ‘right’ decision, many see this as improving decision-making effectiveness in high-pressure situations where human cognitive

shortcomings are often exacerbated (Payne 2021). Thirdly, it is argued that ADS can reduce human casualties because of their increased capacity for precision (Weaver 2020). Finally, the 'arms race' psychology that dominates global warfare drives nations to develop and deploy 'cutting edge' technology that gives them an advantage over their opponents, often without fully considering or sufficiently mitigating the associated risks (Geist 2016). With these factors driving growth, the current deployment of limited-scope ADS in the context of missile and cyber defence will likely only be the start of the increased automation of self-defence technology.

2 Self-defence in International Law

To understand the role of ADS in shaping self-defence law, it is necessary first to map the law's parameters. The use of force is prohibited in international law by *Article 2(4)* of the UN Charter. It is a Jus Cogens norm from which no derogation is permitted (Greenwood 2011). However, an exception to this prohibition applies in cases of collective and individual self-defence, per *Article 51* of the UN Charter. For self-defence to be justified, various elements must be satisfied. Firstly, force in self-defence can only be used in response to an 'armed attack,' where the attacking party has the specific, subjective intention of causing harm (*ICJ Oil Platforms Case* 1992). Secondly, the response must be both necessary and proportionate. This principle is customary international law, confirmed in *Nicaragua v United States* (1984). Necessity has been understood to mean that there are no other ways to resist the 'armed attack' except through force (O'Meara 2021). Proportionality means that the use of force must be balanced in relation to the objectives of the response (O'Meara 2021). Necessity and proportionality are complex, interwoven obligations. Finally, according to *Article 51* of the UN Charter, the state must immediately notify the United Nations Security Council of their actions in self-defence.

The 'self-defence doctrine' has a well-defined 'core' of application. For example, responding to a missile attack by launching an isolated counterstrike directly at the attack's source is clearly a use of force legitimated by the self-defence doctrine. However, whether actions are justified under the self-defence doctrine is often less certain. This is because, like all law, there is a 'penumbra of uncertainty' around the meaning and application of the doctrine's elements. In the following paragraphs, I shall explore and expose these penumbral cases, emphasising that the doctrine both requires further elucidation through state practice over

time and sophisticated legal and moral reasoning to evaluate compliance - things it is unclear whether ADS should or can provide.

2.1 Armed attack

The first ambiguity in relation to 'armed attacks' surrounds the magnitude of force required for such an attack to have been deemed to occur. Unfortunately, the UN Charter does not define 'armed attack'. However, since *Article 51* expressly uses a different term to the reference in *Article 2 (4)* to the 'use of force', lawyers have interpreted 'armed attack' as a higher threshold than simply the use of force (Crawford 2019). They are the 'most grave' forms of attack (*ICJ Oil Platforms Case* 1992) reaching a certain scale and having widespread effects (*Nicaragua v United States* 1984). Thus, we know that a 'mere frontier incident' such as a border skirmish is unlikely to meet this threshold (*Nicaragua v United States* 1984). However, in more complex cases, it is exceedingly difficult to determine the exact scale necessary for the use of force to be seen as an 'armed attack.' This unclear threshold is further complicated by the fact that whilst a single incident may be insufficient, the accumulation of incidents over time may be of sufficient gravity to suggest that an 'armed attack' has occurred (*ICJ Oil Platforms Case* 1992).

The exact 'nature' of an attack that establishes a right to self-defence is also unclear. Whilst it is accepted that an 'armed attack' is to be understood more broadly than 'action by regular armed forces across an international border' (*Nicaragua v United States* 1984 [195]), there is a diversity of types of strikes against a state that may or may not constitute an 'armed attack' capable of enlivening the right to self-defence. Most saliently, cyberattacks are a new form of warfare not considered during the development of the UN Charter. Whilst many authors and countries have suggested that cyberattacks may constitute 'armed attacks' (see, for example, Kesan 2012), others believe that this should be treated with significant caution given the incomparability of cyberattacks with more traditional warfare (see, for example, Germany 2021). This debate is further complexified when cyberattacks have a kinetic result, for example, when an attack shuts down the electricity grid and causes various associated mishaps such as car crashes. Ultimately, whether the 'nature' of the attack that enlivens the right to self-defence extends to other forms of attack, such as cyberattacks, is an unsettled question in international law still to be resolved.

Another complication arises when considering who can commit an 'armed attack.' Whilst a terrorist attack executed by the de jure organs of a State,[1] or for which the State can be attributed responsibility in some other way,[2] is an 'armed attack,' it is not clearly established whether the same threshold for self-defence applies in the case of attack by non-state actors. The International Court of Justice has suggested that terrorist acts unattributable to certain states do not constitute 'armed attacks' even if they are sufficiently 'grave' (for example, in the *Legal Consequences of the Construction of a Wall in the Occupied Palestinian Territory [Advisory Opinion]* 2003). However, this is contradicted by state practice such as that of the US permitted by the United Nations Security Council following the September 2001 attacks, where in Resolutions 1368 and 1373 was agreed that the bombing of the Twin Towers was an 'armed attack' and that military action against Al Qaeda in Afghanistan was justified as self-defence. These competing perspectives suggest that it is an unsettled question in international law to be resolved by further state practice whether a state is justified in using force in response to an 'armed attack' by a non-state actor.

2.2 Necessity and proportionality

Any legitimate action in self-defence must be both necessary and proportionate. Both concepts are forward-looking, meaning that compliance is to be assessed by reference to the goal that the State acting in self-defence is entitled to seek to achieve (Greenwood 2011). The use of force in self-defence is only justified if it is *necessary* to accomplish those justified goals and *proportionate* to achieving them. Thus, what is necessary and proportionate is intrinsically a case-by-case assessment. It requires complex moral assessments with a clear understanding of the state's legally justified goal and how to achieve that goal in the least harmful way. This is an assessment it is widely agreed ADS systems are currently unable, and may never be able, to complete (see, for example, Baggiarini 2024; and Osoba 2024). This line of argument will be picked up in the final section of this response. For now, I simply wish to emphasise the complexities of assessing necessity and proportionality that are not just peripheral but intrinsic to the concepts and the ongoing difficulty of ensuring compliance.

[1]In the *Bosnian Genocide Case (2007)* at [385]*, Article 4* of the International Law Commission's Articles on *Responsibility of States for Internationally Wrongful Acts* (2001), which establishes this principle*,* was affirmed as customary international law.

[2] For example, the International Law Commission's Articles on *Responsibility of States for Internationally Wrongful Acts* (2001) hold that a state may also be held accountable for actions if they are committed by a de facto organ of that state (*Article 8,* affirmed as customary international law in the *Bosnian Genocide Case (2007)* at [41] and [407]), or if those actions are acknowledged and adopted by that state (*Article 11*).

2.3 Anticipatory self-defence

Finally, there is an open question about whether anticipatory self-defence is legally justifiable. Anticipatory self-defence is self-defence in response to an 'imminent' attack. The foundational source of self-defence in international law - the *Caroline* correspondence between America and the UK in 1837, suggests that anticipatory self-defence is likely justified in circumstances where there is 'a necessity of self-defence, instant, overwhelming, leaving no choice of means, and no moment for deliberation.' In this case, the Caroline was an American boat that transported supplies and arms to rebels against the Canadian government, then a territory of Great Britain. The Caroline was camped in American territory on the Canadian border. The Canadian militia, acting under the authority of Great Britain, crept across the river and set the boat on fire. The incident was and is widely accepted as a valid exercise of self-defence despite being 'anticipatory.' However, most cases in the 21st century, particularly those in which ADS are likely to become enlivened, are not as clear-cut as those in 1837. As previously mentioned, the evolving nature and speed of warfare means that a missile fired from Cuba or another island in the Atlantic can strike the United States faster than the Canadian rebels could have crossed the Niagara River. In this context, some have argued that in assessing the 'imminence' which enlivens this potential right to anticipatory self-defence, one should consider the broader circumstances of the threat, including the nature of the harm, capability of the attacking party and nature of the threatened attack (see, for example, Wilmshurst 2006). However, it is not even unanimously accepted that such a right to anticipatory self-defence exists, let alone the context in which this would be justified. Again, this uncertainty in the law of self-defence must be resolved over time.

3. State practice and application to ADS

As previously noted, whilst the right to self-defence is 'treaty law' contained in the UN Charter, it is also widely accepted as customary international law. Consequently, the right to self-defence as an exception against the use of force is given both form and force by state practice and the associated opinio juris. Whilst the force of the law has long been established, its form, as emphasised in the previous section of this essay, remains somewhat amorphous. State practice and opinio juris thus remain ongoingly crucial in shaping the application of the self-defence doctrine in the 21st century, particularly as new challenges created by the evolving nature of conflict arise. If, as I have argued, the key decision-maker in coordinating

a state's response to 'armed attacks' is increasingly likely to be an ADS, is it possible for the decisions executed by these systems to inadvertently shape international law? The rest of this response contemplates this question, ultimately answering it in the affirmative and exposing the considerable risks associated with this.

3.1 Sources of state practice

State practice is any 'instance of conduct by any organ of a state' supporting certain practices as a rule of international law (*Nicaragua v United States* 1984). To become the source of a legal obligation, the relevant opinio juris must accompany this state pract (*The SS Lotus Case* 1927). In practice, state practice and opinio juris are often conflated by courts, academics, and states when assessing whether something is a binding rule of international law. This is because both elements often appear in the same set of sources. According to Crawford (2019), these sources include, but are not limited to:

> 'Diplomatic correspondence, policy statements, press releases, the opinions of government legal advisers, official manuals on legal questions (e.g. manuals of military law), executive decisions and practices, orders to military forces (e.g. rules of engagement), comments by governments on ILC drafts and corresponding commentaries, legislation, international and national judicial decisions, recitals in treaties and other international instruments (especially when in "all states" form), an extensive pattern of treaties in the same terms, the practice of international organs, and resolutions relating to legal questions in UN organs, notably the General Assembly.'

Prima facie, the decisions made in war, automated or otherwise, are not of the character of typical sources of state practice or opinio juris. However, these decisions become representative state practice as they are rationalised, justified and endorsed through later statements, the type of which are listed above. The paradigmatic example is the *Caroline* doctrine, where our understanding of the law of self-defence comes not from the parties' conduct but rather from the discussion and acceptance of this conduct in the following correspondence between the United Kingdom and the United States. In this way, state practice in war becomes combined with the necessary opinio juris and revealed to countries, justices, and legal scholars.

3.2 Transforming ADS decisions into law

The conduct of ADS is directly attributable to the state organ that deploys them. The most accurate understanding of ADS, as argued by Erskine (2024), is that they are 'tools' of the actors that employ them. This is because, in most ways, they are like any other weapon – they are deployed by state militaries (de jure state organs) to achieve state objectives. Of course, there is one glaring distinction between the operation of ADS and traditional weapons – the ADS' immediate actions are not subject to human oversight or control. At the time of initiation, the military (the relevant de jure organ in this analysis) is not in control of, or even necessarily aware of, the executed countermeasures. Nonetheless, sufficient human control still exists over ADS to suggest that they are best conceptualised as 'tools' rather than independent agents. Statements by Australia's expert to the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems demand that meaningful human control is understood across the 'whole life cycle' of an autonomous weapon (Weaver 2019). Australia argues that, even if ADS act independently in the moment of response, they are subject to a 'system of control' that is 'incremental' and 'layered.' Australia's system for the use of Lethal Autonomous Weapons contains legal and policy foundations, design and development oversight, controls in testing, evaluation and review, the acceptance of use and certification of systems, pre-deployment measures, and controls over when to deploy and use the system. All these involvements in the life cycle of the ADS mean that whilst the military does not execute the final decision, they have sufficient involvement in the decision to be attributed responsibility. The ADS is not a 'rogue,' independent agent, but rather a tool designed and deployed by states to assist in protecting their interests. This conceptualisation ensures the attributability of conduct to de jure organs of states. This is extremely important as the autonomisation of weaponry proliferates. As emphasised by the Group of Governmental Experts on the use of Lethal Autonomous weapons (2019), 'human responsibility for decisions on the use of weapons systems must be retained since accountability cannot be transferred onto machines.' If the conduct of ADS is thus directly attributable to the organ that deploys them, it is, by extension, attributable to the state to which that organ belongs.

Importantly, there is a distinction between states being *responsible* for attributable actions and those actions being a *source of state practice*. Most distinctly, the Laws of State Responsibility apply when a nation breaches international law, whereas state practice is

meant to illustrate compliance with the law. For example, in *Caire v United States* (1929), two Mexican police officers kidnapped a French national, tried to extort him, and then shot him. Since they used their official position as police members (a de jure state organ) to commit the crime, Mexico was liable for these internationally wrongful acts even though the officers 'were deemed to have acted outside their competence.' Of course, the officers' actions would not be deemed 'state practice' capable of influencing customary international law. To become state practice capable of shaping international law, the conduct must be accompanied by the relevant opinio juris. This is why statements made by states, rather than solely the conduct, are important. For example, when discussing the military coalition against ISIS in 2017, then Minister for Foreign Affairs, the Hon. Julie Bishop MP, made the following statement at the Global Coalition Summit:

> 'The Coalition's objectives have been to inflict serious damage on ISIS; to destroy the caliphate; and to end the group's ability to conduct and inspire terrorist attacks inside and outside Iraq and Syria. We have made significant progress, and we should remain focused on these objectives. Australia continues to play a major part in this effort, with over 1000 military personnel conducting and supporting air operations and training. We have undertaken over 2,000 sorties and trained 20,000 Iraqi security force personnel.'

Such a statement clarified Australia's belief that the use of force against ISIS to achieve the specified objectives was justified. This statement rendered the associated conduct, such as the over 2000 sorties executed by Australian troops, a relevant form of state practice for developing the doctrine on the use of force. In this way, we see how military conduct can be taken as 'state practice' to shape international law when accompanied by the relevant opinio juris through things such as public statements.

In the context of ADS, one can imagine a similar process: the ADS executes an action, the state releases statements endorsing or justifying such action, thus providing the necessary opinio juris, and the conduct becomes a source of state practice. Let's consider a hypothetical scenario in the near future, where a nation's prime minister makes the following statement:

> 'We have struck back at the infidels in response to their brazen and unprovoked armed attack against our people. The bases of their key military assets located near our borders have been

> targeted and destroyed. This response was necessary to ensure the ongoing safety and security of our great country and proportionate to the achievement of this goal.'

This statement seems innocuous – it reads like many others we have heard throughout history. However, the difference is that this fictional statement refers to ADS conduct rather than more traditional warfare executed intentionally by human military generals. It describes an *automated* response to a perceived 'armed attack' in which the action, calculated and implemented by the system, was the immediate nullification of important enemy military bases. Whether such action would be 'legitimate' is a potentially complex determination that is contingent on the facts of this hypothetical scenario. Whether it is legitimate or not, the fictional statement performs the crucial function of providing opinio juris. It recognises the constraints of the self-defence doctrine and indicates that the state believes the actions of the ADS were compliant with these constraints. In this way, we see how governmental statements may transform the ADS' conduct from something the state is simply *responsible* for into *state practice* capable of shaping international law.

Moreover, government statements are not the only thing capable of transforming military conduct into state practice capable of shaping international law. Emergingly, the 'systems of control' over ADS may provide the necessary opinio juris. As states design and develop ADS, they impute limits and controls on the type and extent of actions the ADS can implement. For example, they may prevent action from exceeding a certain gravity, thereby recognising a potential limit on what conduct is proportionate. Similarly, in deciding when and where to deploy the system, they recognise bounds on their capacity to act. For example, Israel may decide to implement ADS technology on the border of Palestine, a country with whom they are at war, but not on the border with Egypt. This implicitly recognises the limits of their capacity to act in self-defence, whereby an 'armed attack' has been recognised by the government, and, therefore, they feel justified in deploying self-defence measures in this context, but not others. These constraints may be seen as 'opinio juris' – they show that states feel they are bound to act within certain restrictions. Since these restrictions are placed on the ADS conduct directly, it may, in turn, legitimate the conduct ostensibly pursued within these constraints as 'state practice.' Consequently, this conduct, for example, counterstrikes launched directly at the sources of missile attacks in heavily populated civilian areas, might be seen as 'necessary' and 'proportionate' – shaping our understanding of the self-defence doctrine across time.

In sum, this section has argued the following: Firstly, ADS are tools of de jure state organs, not independent agents. States are thus responsible for their conduct. To be transformed into state practice capable of shaping the scope of the self-defence doctrine, this conduct must be accompanied by the relevant opinio juris. Opinio juris can be discerned from the various statements made by governments about the decisions executed by the system. It might also be discernible from the system of control developed over and around the ADS. Like any state practice, assessing whether the conduct of the ADS is a legitimate source is a nuanced task that requires consideration of various factors. However, the core point I have intended to establish is that there is a high possibility that ADS conduct will be explicitly or implicitly endorsed by the state, and thus may become seen as state practice capable of shaping the scope of the self-defence doctrine in international law. The final section of this essay will explore some of the risks associated with this potentiality.

4 Risks

In this context, ADS technology poses various easily overlooked risks for the future direction of self-defence law. These risks arise as states generally accept the ADS' executed decision. I speak here of cases in which the human agent would possibly decide differently but nonetheless accepts and 'owns' the conduct of the ADS. This dynamic may arise in contexts where the human agent would have acted similarly but would have made slight changes to the exact course of action taken. However, it may also arise in more extreme cases where the human agents expressly disagree with the ADS conduct but nonetheless feel obligated to accept and own the course of action given that it is their weapon being deployed. This dynamic is common in military settings, where 'higher-ups' or the executive government internally disapprove of their subordinates' or officers' actions but publicly support them due to reputational and cultural motivations. States also have a vested interest in representing potentially illegitimate conduct as legitimate, since this mitigates their culpability. In such cases, the 'state practice' of the ADS is accompanied by ostensible opinio juris; that is, the state expresses a belief that the actions of the ADS were justified. However, the direction of the state practice differs from the direction that would have been taken if a human agent were in charge. These cases pose significant, insidious dangers for the development of self-defence law that I shall develop further in the final section of this essay.

4.1 Unpredictability

The most obvious danger of such cases is that the unpredictability of ADS systems may lead to courses of action becoming representative of state practice that could not have been anticipated by human developers or officers in charge of deployment, thus undermining the law's legitimacy. This connects with the paradigmatic case outlined above, where states 'own' the ADS' actions despite not reflecting what they would have chosen – or hoped the ADS would have selected. The unpredictability of AI systems is well documented. Indeed, it was a key reason that Deepmind's 'Alpha-Go' defeated the world champion Lee Sedol at the complex game 'Go' – making a series of moves that 'no human would ever do (Metz 2016). In certain contexts, such as the innovative tactics of AlphaGo, unpredictability is a benefit. However, unpredictability is far less desirable in the development of legal principles, particularly customary international law. Thinkers such as Fuller (1965) and Raz (1979) argue that consistency and predictability are crucial to a law's legitimacy because unpredictability erodes trust in the law and legal institutions. As argued, if ADS systems become a source of state practice shaping the law of self-defence, then this law may develop in unpredictable ways. These developments may seem arbitrary to the uncomprehending human mind. Consequently, this unpredictability of ADS conduct may undermine the legitimacy of the law of self-defence. As well as being intrinsically problematic, this decreased legitimacy may have practical consequences of reducing compliance with international law (see, for example, Weber 1946). Therefore, in the context of ADS systems becoming a potential source of state practice, the unpredictability of AI systems poses dangers for the legitimacy of the self-defence doctrine.

4.2 Computation without compassion or comprehension

The second key risk associated with ADS' decisions becoming accepted as state practice is even more fundamental to the operation of AI-driven systems – how they calculate their course of action. Contrary to popular belief, AI systems do not 'think' in a way analogous to humans. Instead, their decisions derive from complex mathematical calculations that even their developers often cannot explain or understand (Pasquale 2015; Amoore 2020). Importantly, these calculations lack the critical thinking, creativity, emotion, and intuition intrinsic to human mentation (Johnson 2022). This is particularly problematic in the context of decisions to resort to force, which have potentially profound consequences for the lives of

many. When humans make these evaluations, we factor in our fears, emotions, and intuitions. These are not 'weaknesses' that limit our capacity to act objectively and correctly but rather essential features of the decision-making process that both offer and limit 'reasons to act' (Pfaff 2019). For example, Zala (2024) discusses how, at the height of the Cuban Missile Crisis, Soviet naval officer Vasily Arkhipov refused to authorise the launch of a nuclear-armed torpedo despite all the external factors indicating that an 'armed attack' had occurred and that he should deploy the weapon in 'self-defence.' It was the role of human doubt, along with the taboo associated with deploying nuclear force, that prevented an almost certain 'full-scale' exchange of nuclear weaponry between the United States and the USSR. Unfortunately, these considerations are beyond the capability of AI as it currently exists (Baggiarini 2024; and Osoba 2024). Their decisions are judgments without understanding, form without meaning (Davis 2024). This is problematic for two reasons. Firstly, there is the risk of causing escalation that human intuitions and emotions might inhibit or prevent (Johnson 2022). In the context of my argument, this escalation may become legitimated over time by ADS acceptance as a source of state practice, reducing the effectiveness of international law to constrain the scope of conflict. Secondly, and more philosophically, there is the risk that the law will be developed without an understanding of its nature and purpose, since the ADS lacks the capacity for 'comprehension' that characterises humanity's cerebration. This danger of 'unthinking' agents shaping the scope of the law leads me into the final risk of ADS systems contemplated by this essay.

4.3 The dystopian concern – controlled by machines?

The final concern this paper raises is the concern we are all most familiar with – the risk we will lose control of our society to machines. For generations, science fiction has conjured images of algorithmic authoritarianism, humankind a mere appendage to the machine whose mathematical mentation demands our devotion to an incomprehensible grand plan. These visions distract from a subtler subjugation, one less dramatic but far closer to home – the willing deference to the dictates of AI in the belief it is in our best interests. Already, the fluctuation of global markets is orchestrated by a web of interconnected algorithms that conduct trades faster than humans can read headlines (Wintermeyer 2023),[3] election results are shaped by news stories conjured up by generative AI (Felstein 2023), and sentencing

[3] 60-75% of trading on all major stock markets is algorithmic.

decisions are made upon the opaque algorithmic assessment of an offender's likelihood of recidivism (Pasquale 2017). If the provocative argument of this essay is accepted, then we may also begin to lose control over the law itself as it is given scope by the autonomous decisions of non-human entities. We live in a world where Johnson (2019) argues there are real fears that:

> 'Military AI could potentially push the pace of combat to a point where the actions of machines surpass the cognitive and physical ability of human decision-makers to control (or even fully understand) future warfare.'

In this context, where war is dominated by automation, it seems imperative that the human agents place moral and legal constraints on the operation of these autonomous systems. It is a long-recognised paradox of customary international law that state practice justifies future state practice. In the context of automated warfare and defence, the dystopian fear is thus – ADS conduct legally justifies such conduct in the future.

5. Conclusion

This paper has argued that there is a serious risk the actions of ADS may be taken as representative of state practice, thereby shaping the scope of permissible application of the self-defence doctrine. The doctrine's ambiguities mean that, over time, it requires further development and clarification by state practice and opinio juris. Given the trend towards automation, the source of this state practice is likely to include the conduct of ADS. The relevant opinio juris will be provided by the subsequent endorsement by states, even if they do not necessarily understand or fully agree with it. Alternatively, opinio juris may also be integrated into the ADS by the systems of control over it. This poses significant risks given the unpredictability of ADS decisions, ADS' inability to comprehend the law they are inadvertently shaping, and the removal of important human intuitions and emotions from resort-to-force decision making. My argument also connects with a broader concern of losing control over the social fabric we hold dear, as the states' direct grip loosens not just over the law of self-defence but over customary international law more broadly as 'state practice' becomes the practice of AI-driven systems. It is frightening enough to live in a world where algorithms fight algorithms – we must be vigilant to prevent these algorithms from also setting the rules of this conflict.

References:

Amoore, Louise. 2020. *Cloud Ethics: Algorithms and the Attributes of Ourselves and Others.* Durham: Duke University Press.

Baggiarini, Bianca. 2024. "Algorithmic War and the Dangers of In-visibility, Anonymity, and Fragmentation." *Australian Journal of International Affairs* 78 (2) (Special Issue) *Anticipating the Future of War: AI, Automated Systems, and Resort-to-Force Decision Making*: 257-265.

Bishop, Julie. 2017. "The Military Coalition against Daesh." Statement to the Global Coalition Summit, Washington DC, March 22, 2017.

Crawford, James. 2019. *Brownlie's Principles of Public International Law*. 9th ed. Oxford: Oxford University Press.

Davis, Jenny. 2024. "Elevating Humanism in High-Stakes Automation: Experts-in-the-Loop and Resort-to-Force Decision Making." *Australian Journal of International Affairs* 78 (2) (Special Issue) *Anticipating the Future of War: AI, Automated Systems, and Resort-to-Force Decision Making*: 200-209.

Erskine, Toni. 2024. "AI and the Future of IR: Disentangling Flesh-and-Blood, Institutional, and Synthetic Moral Agency in World Politics." *Review of International Studies* 50 (3) (Special Issue) *Global Politics: The Next Fifty Years*: 534-559.

Erskine, Toni, and Steve Miller. 2024. "The Possibilities and Perils of AI in Resort-to-Force Decision Making." *Australian Journal of International Affairs* 78 (2) (Special Issue) *Anticipating the Future of War: AI, Automated Systems, and Resort-to-Force Decision Making*: 135-147.

Feldstein, Steven. 2023. "The Consequences of Generative AI for Democracy." *Survival* 65 (5): 117-142.

Frankel, Julia. 2023. "As Hamas Fires Thousands of Rockets at Israel, Here's What to Know About the Iron Dome." *Time*. https://time.com/6327831/israel-iron-dome-rocket-defense-gaza/. Accessed December 12, 2023.

Fuller, Lon. 1965. *The Morality of Law*. New Haven: Yale University Press.

Geist, Edward M. 2016. "It's Already Too Late to Stop the AI Arms Race – We Must Manage It Instead." *Bulletin of the Atomic Scientists* 72 (5): 318–321.

Germany. 2021. "National Position of Germany on Cyber Operations." https://cyberlaw.ccdcoe.org/wiki/National_position_of_Germany_(2021). Accessed December 9, 2023.

Greenwood, Chris. 2011. "Self Defence." *Max Planck Encyclopedias of International Law*. https://opil-ouplaw-com.virtual.anu.edu.au/display/10.1093/law:epil/9780199231690/law-9780199231690-e401?rskey=m92N0r&result=1&prd=MPIL. Accessed October 21, 2023.

Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems. 2019. *Report of the 2019 Session.*

IDF Editorial Team. 2023. "Israel's Life Saving Invention: The Iron Dome." *Israeli Defence Force.* https://www.idf.il/en/mini-sites/the-iron-dome/israel-s-life-saving-invention-the-iron-dome/. Accessed December 12, 2023.

Johnson, James. 2019. "Artificial Intelligence: A Threat to Strategic Stability." *Strategic Studies Quarterly* 14 (1): 16–39.

Johnson, James. 2022. "Delegating Strategic Decision-Making to Machines: Dr. Strangelove Redux?" *Journal of Strategic Studies* 45 (3): 439-477.

Kesan, Jay. 2012. "Mitigative Counterstriking: Self-defence and Deterrence in Cyberspace." *Harvard Journal of Law and Technology* 25 (2): 430-543.

Lewis, Jeffrey. 2017. "Is Launch Under Attack Feasible?" *Nuclear Threat Initiative*. https://www.nti.org/analysis/articles/is-launch-under-attack-feasible/. Accessed November 28, 2023.

Metz, Cade. 2016. "In Two Moves, AlphaGo and Lee Sedol Redefined the Future." *Wired*. https://www.wired.com/2016/03/two-moves-alphago-lee-sedol-redefined-future/. Accessed December 19, 2023.

O'Meara, Chris. 2021. *Necessity and Proportionality and the Right of Self-Defence in International Law*. Oxford: Oxford University Press.

Pasquale, Frank. 2015. *The Black Box Society: The Secret Algorithms that Control Money and Information.* Cambridge: Harvard University Press.

Pasquale, Frank. 2017. "Secret Algorithms Threaten the Rule of Law." *MIT Technology Review.* https://www.technologyreview.com/2017/06/01/151447/secret-algorithms-threaten-the-rule-of-law/. Accessed November 22, 2023.

Payne, Kenneth. 2021. *I, Warbot: The Dawn of Artificially Intelligent Conflict.* London: Hurst and Company.

Pfaff, Anthony. 2019. "The Ethics of Acquiring Disruptive Technologies." *PRISM* 8 (3): 128-145.

Raz, Joseph. 1979. "The Rule of Law and Its Virtue." In *The Authority of Law: Essays on Law and Morality*, edited by Raz, Joseph. Oxford: Clarendon Press.

Ryabikhin, Leonid. 2019. *Russia's NC3 and Early Warning Systems.* Technology for Global Security (special report).

Simonite, Tom. 2017. "For Superpowers, Artificial Intelligence Fuels New Global Arms Race." *Wired*. https://www.wired.com/story/for-superpowers-artificial-intelligence-fuels-new-global-arms-race/. Accessed December 8, 2023.

Weaver, Johanna on behalf of Australia. 2019. "Australia's System of Control and Applications for Autonomous Weapons Systems." Statement to the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems.

Weaver, Johanna on behalf of Australia. 2020. "Lethal Autonomous Weapons System." National Commentary on the Convention on Certain Conventional Weapons to the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems.

Weber, Max. 1946. "Politics as Vocation" in *Max Weber: Essays in Sociology* edited by Gerth, H.H. and Mills, C.W., Oxford: Oxford University Press.

Wilshurst, Elizabeth. 2006. "The Chatham House Principles of International Law on the Use of Force by States in Self-defence." *International and Comparative Law Quarterly* 55: 963-72.

Wintermeyer, Lawrence. 2023. "You Can't Spell Blockchain Without AI: AI Will Dominate Crypto Trading." *Forbes*. https://www.forbes.com/sites/lawrencewintermeyer/2023/09/14/you-cant-spell-blockchain-without-ai-ai-will-dominate-crypto-trading/?sh=31a103a66ede. Accessed December 23, 2023.

Zala, Benjamin. 2024. "Should AI Stay or Should AI Go? First Strike Incentives & Deterrence Stability in the Third Nuclear Age." Australian Journal of International Affairs 78 (2) (Special Issue) *Anticipating the Future of War: AI, Automated Systems, and Resort-to-Force Decision Making*: 154-163.

Legislation and Case Law:

Bosnia and Herzegovina v Serbia and Montenegro [2007] ICJ 2 ('*Bosnian Genocide Case'*).

Caroline correspondence between America and the UK in 1837, as outlined in Ruys, Tom, Corten, Olivier and Hofer, Alexander. 2018. *The Use of Force in International Law: A Case-Based Approach*. Oxford University Press, Oxford.

Charter of the United Nations (1945)

France v Turkey (1927) PCIJ Ser A. No 10 *('SS Lotus Case').*

Gustav Caire (France) v United Mexican States (1929) 5 RIAA 516.

International Law Commission, *Responsibility of States for Internationally Wrongful Acts* (2001). As in United Nations General Assembly (UNGA) Res 56/83 (12 December 2001) A/56/49(Vol. I)/Corr.4.

Islamic Republic of Iran v The United States of America (1992) ICJ 4 (*'Oil Platforms Case'*).

Legal Consequences of the Construction of a Wall in the Occupied Palestinian Territory, Advisory Opinion, [2004] ICJ Rep 136.

The Republic of Nicaragua v The United States of America (1986) ICJ 14 (*'Nicaragua v United States'*).

United Nations Security Council (UNSC) Res 1368 (12 September 2001) S/RES/1368.

United Nations Security Council (UNSC) Res 1373 (28 September 2001) S/RES/1373.