26

Artificial Intelligence, Law, and National Security

Ebrahim Afsah

I. INTRODUCTION: KNOWLEDGE IS POWER

The conjecture 'that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it'[1] has motivated scientists for more than half a century, but only recently attracted serious attention from political decision-makers and the general public. This relative lack of attention is perhaps due to the long gestation of the technology necessary for that initial conjecture to become a practical reality. For decades merely an aspiration among a small, highly skilled circle engaged in basic research, the past few years have witnessed the emergence of a dynamic, economically and intellectually vibrant field.

From the beginning, national security needs drove the development of Artificial Intelligence (AI). These security needs were motivated in part by surveillance needs, especially code-breaking, and in part by weapons development, in particular nuclear test simulation. While the utilisation of some machine intelligence has been part of national security for decades, the recent explosive growth in machine capability is likely to transform national and international security, consequently raising important regulatory questions.

Fueled by the confluence of at least five factors – the increase in computational capacity; availability of data and big data; revolution in algorithm and software development; explosion in our knowledge of the human brain; and existence of an affluent and risk-affine technology industry – the initial conjecture is no longer aspirational but has become a reality.[2] The resulting capabilities cannot be ignored by states in a competitive, anarchic international system.[3]

[1] As succinctly put in the project proposal to the 1956 Dartmouth Conference; J McCarthy and others, 'A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955' (2006) 47 *AI Magazine* 12.

[2] NJ Nilsson, *The Quest for Artificial Intelligence: A History of Ideas and Achievements* (2010) (hereafter Nilsson, The Quest for Artificial Intelligence).

[3] The literature is extremely copious, a good point of departure is H Bull, *The Anarchical Society: A Study of Order in World Politics* (1977); KA Oye, 'Explaining Cooperation under Anarchy: Hypotheses and Strategies' (1985) 38 *World Politics* 226. Professor Oye was the convener of the talk by Judge James Baker at MIT on 6 March 2018 that initially got me interested in AI, my intellectual debt to his work is gratefully acknowledged. See JE Baker, 'Artificial Intelligence and National Security Law: A Dangerous Nonchalance' (2018) 18-01 MIT Starr Forum Report (hereafter Baker, 'Artificial Intelligence and National Security Law').

https://doi.org/10.1017/9781009207898.035 Published online by Cambridge University Press

As AI becomes a practical reality, it affects national defensive and offensive capabilities,[4] as well as general technological and economic competitiveness.[5]

There is a tendency to describe intelligence in an anthropomorphic fashion that conflates it with emotion, will, conscience, and other human qualities. While this makes for good television, especially in the field of national security,[6] this seems to be a poor analytical or regulatory guideline.[7] For these purposes, a less anthropocentric definition is preferable, as suggested for instance by *Nils Nilsson*:

> For me, artificial intelligence is that activity devoted to making machines intelligent, and intelligence is that quality that enables an entity to function appropriately and with foresight in its environment. According to that definition, lots of things – humans, animals, and some machines – are intelligent. Machines, such as 'smart cameras,' and many animals are at the primitive end of the extended continuum along which entities with various degrees of intelligence are arrayed. At the other end are humans, who are able to reason, achieve goals, understand and generate language, perceive and respond to sensory inputs, prove mathematical theorems, play challenging games, synthesize and summarize information, create art and music, and even write histories. Because 'functioning appropriately and with foresight' requires so many different capabilities, depending on the environment, we actually have several continua of intelligences with no particularly sharp discontinuities in any of them. For these reasons, I take a rather generous view of what constitutes AI.[8]

[4] M Karlin, 'The Implications of Artificial Intelligence for National Security Strategy' in A Blueprint for the Future of AI (Brookings, 1 November 2018) www.brookings.edu/series/a-blueprint-for-the-future-of-ai/; A Polyakova, 'Weapons of the Weak! Russia and AI-Driven Asymmetric Warfare' in A Blueprint for the Future of AI (Brookings, 15 November 2018) www.brookings.edu/series/a-blueprint-for-the-future-of-ai/; M O'Hanlon, 'The Role of AI in Future Warfare' in A Blueprint for the Future of AI (*Brookings*, 29 November 2018) www.brookings.edu/series/a-blueprint-for-the-future-of-ai/.

[5] Much current attention is given to China's single-minded pursuit of attaining technological competitiveness by 2025 and leadership by 2035, including in the field of AI. The State Council published in July 2017 a 'New Generation Artificial Intelligence Development Plan' that built on the May 2015 'Made in China 2025' plan, which had already listed 'new information technology' as the first of ten strategic fields. The two plans are accessible at https://flia.org/notice-state-council-issuing-new-generation-artificial-intelligence-development-plan/ and http://english.www.gov.cn/2016special/madeinchina2025/. For a discussion see *inter alia* 'AI in China' (*OECD*, 21 February 2020) https://oecd.ai/dashboards/countries/China; 'AI Policy China' (Future of Life Institute, February 2020) <https://futureoflife.org/ai-policy-china/; P Mozur and SL Myers, 'Xi's Gambit: China Plans for a World without American Technology' *New York Times* (11 March 2021) www.nytimes.com/2021/03/10/business/china-us-tech-rivalry.html (hereafter Mozur and Myers, 'Xi's Gambit'); X Yu and J Meng, 'China Aims to Outspend the World in Artificial Intelligence, and XI Jinping Just Green Lit the Plan' *South China Morning Post* (18 October 2017) www.scmp.com/business/china-business/article/2115935/chinas-xi-jinping-highlights-ai-big-data-and-shared-economy.

[6] Perhaps most enduringly in the 1983 movie 'WarGames', where a recently commissioned intelligent central computer is hacked into by a teenager, who inadvertently almost causes nuclear Armageddon. This is only averted when the computer learns, after playing Tic-Tac-Toe with the teenager, that nuclear war cannot have a winner, causing him to rescind the launch command and to comment: 'A strange game. The only winning move is not to play.' There are obvious allusions to the doomsday machine scenario discussed further below. Interestingly, simultaneous to the film but unbeknownst to most until much later, the automated early warning system of the Soviet Union on 26 September 1983, at a time of extreme tension between the two countries, falsely indicated an American nuclear attack, almost triggering a catastrophic retaliatory nuclear attack. This was stopped by Lieutenant Colonel Stanislav Petrov, who disobeyed orders because he intuited that it was a false alarm; M Tegmark, 'A Posthumous Honor for the Man Who Saved the World' (*Bulletin of the Atomic Scientist*, 26 September 2018) https://thebulletin.org/2018/09/a-posthumous-honor-for-the-man-who-saved-the-world/.

[7] A Chayes, 'Cyber Attacks and Cyber Warfare: Framing the Issues' in A Chayes (ed), *Borderless Wars: Civil Military Disorder and Legal Uncertainty* (2015) (hereafter Chayes, 'Cyber Attacks and Cyber Warfare'); L DeNardis, 'The Emerging Field of Internet Governance' (2010) Yale Information Society Project Working Paper Series https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1678343 (hereafter DeNardis, 'The Emerging Field of Internet Governance').

[8] Nilsson, *The Quest for Artificial Intelligence* (n 2) xiii.

https://doi.org/10.1017/9781009207898.035 Published online by Cambridge University Press

The influential *Stanford 100 Year Study on Artificial Intelligence* explicitly endorses this broad approach, stressing that human intelligence has been but the inspiration for an endeavour that is unlikely to actually replicate the brain. It appears that intelligence – whether human, animal, or machine[9] – is not necessarily one of clearly differentiated *kind*, but ultimately a question of *degree* of speed, capability, and adaptability:

> Artificial Intelligence (AI) is a science and a set of computational technologies that are inspired by – but typically operate quite differently from – the ways people use their nervous systems and bodies to sense, learn, reason, and take action. . . . According to this view, the difference between an arithmetic calculator and a human brain is not one of kind, but of scale, speed, degree of autonomy, and generality. The same factors can be used to evaluate every other instance of intelligence – speech recognition software, animal brains, cruise-control systems in cars, Go-playing programs, thermostats – and to place them at some appropriate location in the spectrum.[10]

At its most basic, AI means making sense of data, and can thus be differentiated from cyberspace, which primarily concerns the transmission of data. Collecting data is fairly inconsequential without someone to analyse and make sense of it.[11] If the purpose of a thought or action can be expressed numerically, it can be turned into coded instructions and thereby cause a machine to achieve that purpose. In order to understand the relationship better, it is helpful to differentiate between data, information, knowledge, and intelligence.

Data is raw, unorganised, factual, sensory observation, collected in either analog or digital form, with single data points unrelated to each other. Already in this raw form, data can be used by simple machines to achieve a purpose, for instance temperature or water pressure readings by a thermostat switching a heater on or off, or a torpedo's depth sensor guiding its steering system. Observed and recorded facts can take many forms, such as statistics, satellite surveillance photographs, dialed phone numbers, etc. Such data, whether qualitative or quantitative, stands on its own and is not related to external signifiers. In this form, it is not very informative and fairly meaningless. Where analog storage is logistically limited, the recording of observational data in electronic, machine-readable form is no longer physically limited.

Information, by contrast, depends on an external mental model through which data acquires meaning, context, and significance. Data becomes information through analysis and categorisation; it acquires significance only through the imposition of order and structure. Information is, therefore, data that has been processed, organised according to meaningful criteria, given context, and thereby made useful towards achieving outcomes according to predetermined

[9] Human denial of both intelligence and consciousness in other creatures seems ultimately to be a fairly straightforward case of cognitive dissonance: 'To me, consciousness is the thing that feels like something,' said Carl Safina, an ecologist. 'We're learning that a lot of animals – dogs, elephants, other primates – have it. . . . I think it's because it's easier to hurt them if you think of them as dumb brutes. Not long ago, I was on a boat with some nice people who spear swordfish for a living. They sneak up to swordfish sleeping near the surface of the water and harpoon them, and then the fish just go crazy and kind of explode. When I asked, 'Do the fish feel pain?' the answer was, 'They don't feel anything.' Now, it's been proven experimentally that fish feel pain. I think they feel, at least panic. They clearly are not having a good time when they are hooked. But if you think of yourself as a good person, you don't want to believe you're causing suffering. It's easier to believe that there's no pain.' C Dreifus, 'Carl Safina Is Certain Your Dog Loves You' *New York Times* (21 October 2019) www.nytimes.com/2019/10/21/science/carl-safina-animal-cognition.html.

[10] 'Artificial Intelligence and Life in 2030 – One Hundred Year Study on Artificial Intelligence, Report of the 2015 Study Panel' (*Stanford University*, September 2016) 4, 12 https://ai100.stanford.edu/2016-report.

[11] T Zarsky, '"Mine Your Own Business!": Making the Case for the Implications of the Data Mining of Personal Information in the Forum of Public Opinion' (2003) 5 *Yale J L & Tech* 1, 4 *et seq* (hereafter Zarsky, 'Mine Your Own Business!').

https://doi.org/10.1017/9781009207898.035 Published online by Cambridge University Press

needs. This process is dependent on the existence of conceptional models created in response to these needs.[12] Significance, meaning, and usefulness are, therefore, qualities not inherent in the data, but external impositions to sift, categorise, and 'clean' data from extraneous 'noise'. Data that has been transformed into information has 'useless' elements removed and is given context and significance according to an external yardstick of 'usefulness'. To follow the earlier example, linking temperature readings in different rooms at different times, with occupancy readings and fluctuating electricity prices could be used by a 'smart' thermostat to make 'intelligent' heating choices.

Knowledge is to make sense of information, being aware of the limitations of the underlying data and theoretical models used to classify it, being able to place that information into a wider context of meaning, purpose, and dynamic interactions, involving experience, prediction, and the malleability of both purpose and model. Knowledge refers to the ability to understand a phenomenon, theoretically or practically, and to use such understanding for a deliberate purpose. It can be defined as 'justified true belief'.[13] This process complements available information with inferences from past experience and intuition, and responds to feedback, including sensory, cognitive, and evaluative.

Intelligence refers to the ability to 'function appropriately and with foresight', thus AI presumes that the act of thinking that turns (sensory) data into information and then into knowledge, and finally into purposeful action is not unique to humans or animals. It posits that the underlying computational process is formally deducible, can be scientifically studied and replicated in a digital computer. Once this is achieved, all the inherent advantages of the computer come to bear: speed, objectivity (absence of bias, emotion, preconceptions, etc.), scalability, permanent operation, etc. In the national security field, some have compared this promise to the mythical figure of the *Centaur*, who combined the intelligence of man with the speed and strength of the horse.[14]

The development of the Internet concerned the distribution of data and information between human and machine users.[15] AI, by contrast, does not primarily refer to the transmission of raw or processed data, the exchange of ideas, or the remote control of machinery (Internet of things, military command and control, etc.), but the ability to detect patterns in data, process data into information, and classify that information in order to predict outcomes and make decisions. *Darrell M. Allen* and *John R. West* suggest three differentiating characteristics of such systems: intentionality, intelligence, and adaptability.[16]

The Internet has already transformed our lives, but the enormous changes portended by AI are just beginning to dawn on us. The difficulty of predicting that change, however, should not serve as an excuse for what *James Baker* deemed 'a dangerous nonchalance' on behalf of decision-makers tasked with managing this transformation.[17] Responsible management of

[12] On this point, see generally E Derman, *Models Behaving Badly, Why Confusing Illusion with Reality Can Lead to Disaster, on Wall Street and in Life* (2011); I Hacking, *Representing and Intervening, Introductory Topics in the Philosophy of Natural Science* (1983).

[13] J Jenkins, M Steup, 'The Analysis of Knowledge' in E N Zalta (ed) *The Stanford Encyclopedia of Philosophy* (Summer 2021 ed.) https://plato.stanford.edu/entries/knowledge-analysis/.

[14] JE Baker, *The Centaur's Dilemma – National Security Law for the Coming AI Revolution* (2021) (hereafter Baker, The Centaur's Dilemma).

[15] See generally, BM Leiner and others, *Brief History of the Internet* (1997) (hereafter Leiner and others, Brief History of the Internet); M Waldrop, 'DARPA and the Internet Revolution' (DARPA, 2015) www.darpa.mil/about-us/timeline/modern-internet (hereafter Waldrop, 'DARPA and the Internet Revolution').

[16] DM West and JR Allen, 'How Artificial Intelligence Is Transforming the World' in A *Blueprint for the Future of AI* (Brookings, 24 April 2018) www.brookings.edu/series/a-blueprint-for-the-future-of-ai/.

[17] See note 3.

https://doi.org/10.1017/9781009207898.035 Published online by Cambridge University Press

national security requires an adequate and realistic assessment of the threats and opportunities presented by new technological developments, especially their effect on the relative balance of power and on global public goods, such as the mitigation of catastrophic risks, arms races, and societal dislocations. In modern administrative states, such management is inevitably done through law, both nationally and internationally.[18]

In this chapter, I will begin by contrasting the challenge posed by AI to the related but distinct emergence of the cyber domain. I then outline six distinct implications for national security: doomsday scenarios, autonomous weapons, existing military capabilities, reconnaissance, economics, and foreign relations. Legal scholarship often proposes new regulation when faced with novel societal or technological challenges. But it appears unlikely that national actors will forego the potential advantages offered by a highly dynamic field through self-restraint by international convention. Still, even if outright bans and arms control-like arrangements are unlikely, the law serves three important functions when dealing with novel challenges: first, as the repository of essential values guiding action; second, offering essential procedural guidance; and third, by establishing authority, institutional mandates, and necessary boundaries for oversight and accountability.

II. CYBERSPACE AND AI

The purpose of this sub-section is not to outline the large literature applying the principles of general international law, and especially the law of armed conflict, to cyber operations. Rather, it seeks to highlight the distinctive elements of the global communication infrastructure, especially how AI is distinct from some of the regulatory and operational[19] challenges that characterise cybersecurity.[20] The mental image conjured by early utopian thinkers and adopted later by realist and military policy-makers rests on the geographical metaphor of 'cyberspace' as a non-corporeal place of opportunity and risk.[21] This place needs to be defended and thus constitutes an appropriate area of military operations.

As technical barriers eventually fell, the complexity of the network receded behind increasingly sophisticated but simple to operate graphical user-interfaces, making networked information-sharing first a mainstream, and eventually a ubiquitous phenomenon, affecting

[18] The literature on the administrative state is too copious to list, disparate discussions that helped guide my own thinking on this matter include S Cassese, 'Administrative Law without the State? The Challenge of Global Regulation' (2005) 37 *NYU Journal of International Law & Politics* 663; PD Feaver, 'The Civil-Military Problematique: Huntington, Janowitz, and the Question of Civilian Control' (1996) 23 *Armed Forces & Society* 149; SJ Kaufman, 'The Fragmentation and Consolidation of International Systems' (1997) 51 IO 173; A Chayes, 'An Inquiry into the Workings of Arms Control Agreements' (1972) 85 *Harvard Law Review* 905; AH Chayes and A Chayes, 'From Law Enforcement to Dispute Settlement: A New Approach to Arms Control Verification and Compliance' (1990) 14 *IS* 147.

[19] Good overviews can be found in GD Brown, 'Commentary on the Law of Cyber Operations and the DoD Law of War Manual' in MA Newton (ed), *The United States Department of Defense Law of War Manual* (2019); WH Boothby, 'Cyber Capabilities' in WH Boothby (ed), *New Technologies and the Law in War and Peace* (2018) (hereafter Boothby, 'Cyber Capabilities'); MN Schmitt (ed), *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations* (2017) 401 *et seq.*; JC Woltag, *Cyber Warfare: Military Cross-Border Computer Network Operations under International Law* (2014) (hereafter Woltag, *Cyber Warfare*); C Droege, 'Get Off My Cloud: Cyber Warfare, International Humanitarian Law, and the Protection of Civilians' (2012) 94 *Int'l Rev of the Red Cross* 533.

[20] With respect to cyber warfare, see also Chayes, 'Cyber Attacks and Cyber Warfare' (n 7); M Finnemore and DB Hollis, 'Constructing Norms for Global Cybersecurity' (2016) 110 *AJIL* 425. Regarding the specific impact of AI, see Baker, *The Centaur's Dilemma* (n 14).

[21] See further J Branch, 'What's in a Name? Metaphors and Cybersecurity' (2021) 75 *IO* 39 (hereafter Branch, 'Metaphors and Cybersecurity').

https://doi.org/10.1017/9781009207898.035 Published online by Cambridge University Press

almost all aspects of human life almost everywhere. This has led to an exponential increase in the availability of information, much of it of a sensitive nature, often voluntarily relinquished. This has created a three-pronged challenge: data protection, information management, and network security.[22]

Much early civilian, especially academic, thinking focused on the dynamic relationship between technology and culture, stressing the emergence of a new, virtual habitat: 'A new universe, a parallel universe created and sustained by the world's computers and communication lines.'[23] But as the novelty wore off while its importance grew, the Internet became 're-territorialised' as nation-states asserted their jurisdiction, including in the hybrid, multi-stakeholder regulatory fora that had developed initially under American governmental patronage.[24] Perhaps more importantly, this non-corporeal realm created by connected computers, came to be seen not as a parallel universe following its own logic and laws, but as an extension of existing jurisdictions and organisational mandates:

> Although it is a man-made domain, cyberspace is now as relevant a domain for DoD [Department of Defence] activities as the naturally occurring domains of land, sea, air, and space. Though the networks and systems that make up cyberspace are man-made, often privately owned, and primarily civilian in use, treating cyberspace as a domain is a critical organizing concept for DoD's national security missions. This allows DoD to organize, train, and equip for cyberspace as we do in air, land, maritime, and space to support national security interests.[25]

This is reflected in the United States (US) National Security Strategy, which observes: 'Cybersecurity threats represent one of the most serious national security, public safety, and economic challenges we face as a nation.'[26] Other countries treat the issue with similar seriousness.[27]

Common to the manner in which diverse nations envisage cybersecurity is the emphasis on information infrastructure, in other words, on the need to keep communication channels operational and protected from unwanted intrusion. This, however, is distinct from the specific challenge of AI, which concerns the creation of actionable knowledge by a machine.

The initial ideas that led to the creation of the Internet sought to solve two distinct problems: the civilian desire to use expensive time-share computing capacity at academic facilities more efficiently by distributing tasks, and the military need to establish secure command and control

[22] See generally GD Solis, *The Law of Armed Conflict. International Humanitarian Law in War* (2016) 673–709 (hereafter Solis, *The Law of Armed Conflict*).

[23] ML Benedikt, *Cyberspace: First Steps* (1991) 1.

[24] See *inter alia* U Kohl, *The Net and the Nation State: Multidisciplinary Perspectives on Internet Governance* (2017) (hereafter Kohl, *The Net and the Nation State*); J Nocetti, 'Contest and Conquest: Russia and Global Internet Governance' (2015) 91 *Int'l Aff* 111; DeNardis, 'The Emerging Field of Internet Governance' (n 7); ML Mueller, *Networks and States: The Global Politics of Internet Governance* (2010).

[25] Department of Defence, 'Strategy for Operating in Cyberspace' (July 2011) https://csrc.nist.gov/CSRC/media/Projects/ISPAB/documents/DOD-Strategy-for-Operating-in-Cyberspace.pdf 5, referring to the 2010 Quadrennial Defence Review. Outer space has been an area of great power competition since the Sputnik satellite, but it has received added impetus in recent years with the creation of dedicated Space Commands in the US and other countries, see WJ Broad, 'How Space Became the Next 'Great Power' Contest between the US and China' *New York Times* (24 January 2021) www.nytimes.com/2021/01/24/us/politics/trump-biden-pentagon-space-missiles-satellite.html.

[26] Department of Defence, 'Strategy for Operating in Cyberspace' (July 2011) https://csrc.nist.gov/CSRC/media/Projects/ISPAB/documents/DOD-Strategy-for-Operating-in-Cyberspace.pdf 1, referring to the 2010 National Security Strategy. Similar language can be found in previous and subsequent national security strategies, both American and others, including the current 2021 interim one issued by the Biden Administration.

[27] E Afsah, 'Country Report Denmark' in M Kilching and C Sabine (eds), *Economic and Industrial Espionage in Germany and Europe* (2016).

https://doi.org/10.1017/9781009207898.035 Published online by Cambridge University Press

connections between installations, especially to remote nuclear weapons facilities.[28] In both cases, it was discovered that existing circuit switched telephone connections were unreliable. The conceptional breakthrough consisted in the idea of package switched communication, which permitted existing physical networks to be joined non-hierarchically, permitting a non-hierarchical, decentralised architecture that is resilient, scalable, and open.[29]

The Internet is, therefore, not one network, but a set of protocols specifying data formats and rules of transmission, permitting local, physical networks to communicate along dynamically assigned pathways.[30] The technology, the opportunities, and the vulnerabilities it offered came to be condensed in the spatial analogy of cyberspace. This 'foundational metaphor' was politically consequential because the use of certain terminology implied, rather than stated outright, particular understandings of complex issues at the expense of others, thus shaping policy debates and outcomes.[31] Denounced later by himself as merely an 'effective buzzword' chosen because 'it seemed evocative and essentially meaningless', the definition offered by *William Gibson* highlights the problematic yet appealing character of this spatial analogy: 'Cyberspace. A consensual hallucination experienced daily by billions of legitimate operators, in every nation ... A graphic representation of data abstracted from the banks of every computer in the human system. Unthinkable complexity. Lines of light ranged in the nonspace of the mind, clusters and constellations of data. Like city lights, receding.'[32] The term combined the non-physical nature of a world being dynamically created by its denizens in their collective imagination, but relying behind the graphical user-interface on a complex physical infrastructure.[33] The advantages of open communications have eventually led military and

[28] The need to ensure reliable communication after sustaining a devastating first strike was a key ingredient of credible nuclear deterrence. The Soviet 'Dead Hand' system (*Mertvaya Ruka*, officially: *Systema Perimetr*) was an alternative, 'fail-deadly' method of solving that practical problem: meant as a backup to the *Kazbek* communication system, *Perimetr* was to fully automatically trigger nuclear retaliation if it detected an attack, even if command structures and human personnel had been destroyed. US Defence Intelligence Agency, 'Russia Military Power: Building a Military to Support Great Power Aspirations' (2017) https://www.hsdl.org/?view&did=801968 26–28; N Thompson, 'Inside the Apocalyptic Soviet Doomsday Machine' Wired (21 September 2009) www.wired.com/2009/09/mf-deadhand/; WJ Broad, 'Russia Has 'Doomsday' Machine, US Expert Says' *New York Times* (8 October 1993) www.nytimes.com/1993/10/08/world/russia-has-doomsday-machine-us-expert-says.html.

[29] This means that data to be transmitted will be split into several packets, based on various criteria including size. The packets will be sent independently from each other, usually along different pathways, and re-assembled at the destination. They contain the actual data to be sent, destination and source address, and other information necessary for reliable transmission. The idea was simultaneously but independently developed at MIT in Cambridge, Massachusetts (1961–1967), RAND in Santa Monica, California (1962–1965) and the British National Physical Laboratory (NPL) in London (1964–1967). This genesis is well described by several of its key protagonists themselves in Leiner and others, *Brief History of the Internet* (n 15); Waldrop, 'DARPA and the Internet Revolution' (n 15).

[30] This reliance on a conceptional, rather than physical architecture is reflected in the definition laid down in US law: 'The term "Internet" means collectively the myriad of computer and telecommunications facilities, including equipment and operating software, which comprise the interconnected world-wide network of networks that employ the Transmission Control Protocol/Internet Protocol, or any predecessor or successor protocols to such protocol, to communicate information of all kinds by wire or radio.'15 USC § 6501(6), www.law.cornell.edu/uscode/text/15/6501#6.

See also Woltag, *Cyber Warfare* (n 19) 9.

[31] Branch, 'Metaphors and Cybersecurity' (n 21).

[32] W Gibson, *Neuromancer* (1984) 69, emphasis added. Gibson makes the disparaging remarks about his term in the documentary film M Neale, 'No Maps for these Territories' (2000).

[33] 'Gibson's networked artificial environment anticipated the globally internetworked technoculture (and its surveillance) in which we now find ourselves. The term has gone on to revolutionize popular culture and popular science, heralding the power and ubiquity of the information age we now regard as common as iPhones. Since its invention, 'cyberspace' has come to represent everything from computers and information technology to the Internet and "consensual hallucinations" as different as The Matrix, Total Information Awareness, and reality TV.' March 17, 1948: W Gibson, 'Father of Cyberspace' *Wired* (16 March 2009) www.wired.com/2009/03/march-17-1948-william-gibson-father-of-cyberspace-2/.

https://doi.org/10.1017/9781009207898.035 Published online by Cambridge University Press

civilian installations in all nations to become accessible through the Internet, creating unique vulnerabilities due to opportunity costs of communication disruption, physical damage to installations, and interruptions of critical public goods like water or electricity.[34] What the American military defines as its key challenge in this area applies likewise to most other nations:

> US and international businesses trade goods and services in cyberspace, moving assets across the globe in seconds. In addition to facilitating trade in other sectors, cyberspace is itself a key sector of the global economy. Cyberspace has become an incubator for new forms of entrepreneurship, advances in technology, the spread of free speech, and new social networks that drive our economy and reflect our principles. The security and effective operation of US critical infrastructure – including energy, banking and finance, transportation, communication, and the Defense Industrial Base – rely on cyberspace, industrial control systems, and information technology that may be vulnerable to disruption or exploitation.[35]

Some have questioned the definitional appropriation of 'cyberspace' as a 'domain' for military action through 'linguistic and ideational factors [which] are largely overlooked by the prevailing approach to cybersecurity in IR [international relations], which has productively emphasized technical and strategic aspects' at the expense of alternative ways of thinking about security in this field.[36] Without prejudice to the theoretical contributions such investigations could make to political science and international relations,[37] the legal regulation of defensive and offensive networked operations has, perhaps after a period of initial confusion,[38] found traditional concepts to be quite adequate, perhaps because the spatial analogy facilitates the application of existing legal concepts.

The central challenges posed by the increasing and unavoidable dependence on open-architecture communication are both civilian and military. They concern primarily three distinct but related operational tasks: prevent interruptions to the flow of information, especially financial transactions; prevent disruptions to critical command and control of civilian and military infrastructure, especially energy, water, and nuclear installations; and prevent unauthorised access to trade and military secrets.[39] These vulnerabilities have, of course, corresponding opportunities for obtaining strategic information, striking at long distance while maintaining 'plausible deniability',[40] and establishing credible deterrence.[41] Again, how the American military describes its own mandate applies in equal measure to other nations, not least its chief competitors Russia and China:

[34] DE Sanger, 'China Appears to Warn India: Push Too Hard and the Lights Could Go Out' *New York Times* (28 February 2021) www.nytimes.com/2021/02/28/us/politics/china-india-hacking-electricity.html.

[35] US Department of Defence, 'Strategy for Operating in Cyberspace' (July 2011) 1, https://csrc.nist.gov/CSRC/media/Projects/ISPAB/documents/DOD-Strategy-for-Operating-in-Cyberspace.pdf.

[36] Branch, 'Metaphors and Cybersecurity' (n 21) 41.

[37] See for instance M Finnemore and DB Hollis, 'Constructing Norms for Global Cybersecurity' (2016) 110 *AJIL* 425.

[38] A Chayes, 'Implications for Civil-Military Relations in Cyber Attacks and Cyber Warfare' in A Chayes (ed), *Borderless Wars: Civil Military Disorder and Legal Uncertainty* (2015).

[39] Politiets Efterretningstjeneste, 'Trusler mod Danmark: Spionage' (2015), https://pet.dk/spionage; JUO Nielsen, 'Erhvervshemmelighedsværnet i Norden og EU' (2014) *Erhvervsjuridisk Tidsskrift* 1.

[40] See further L Arimatsu, 'The Law of State Responsibility in Relation to Border Crossings: An Ignored Legal Paradigm' (2013) 89 *Int'l L Stud* 21; P Margulies, 'Networks in Non-International Armed Conflicts: Crossing Borders and Defining "Organized Armed Groups"' (2013) 89 *Int'l L Stud* 54.

[41] Y Benkler, 'Degrees of Freedom, Dimensions of Power' (2016) *Daedalus* 18 (hereafter Benkler, 'Degrees of Freedom'). Unlike in classical military spheres, it is important to note that in the cyber-domain effective repulsion and deterrence does not necessarily have to be assumed by the military, see Forsvarsministeriet, 'Center for Cybersikkerhed' (18 September 2020) https://www.fmn.dk/da/arbejdsomraader/cybersikkerhed/center-for-cybersikkerhed/.

https://doi.org/10.1017/9781009207898.035 Published online by Cambridge University Press

> American prosperity, liberty, and security depend upon open and reliable access to information. The Internet empowers us and enriches our lives by providing ever-greater access to new knowledge, businesses, and services. Computers and network technologies underpin US military warfighting superiority by enabling the Joint Force to gain the information advantage, strike at long distance, and exercise global command and control.
>
> The arrival of the digital age has also created challenges for the Department of Defense (DoD) and the Nation. The open, transnational, and decentralized nature of the Internet that we seek to protect creates significant vulnerabilities. Competitors deterred from engaging the US and our allies in an armed conflict are using cyberspace operations to steal our technology, disrupt our government and commerce, challenge our democratic processes, and threaten our critical infrastructure.[42]

Crucially important as these vulnerabilities and opportunities are for national security, defensive and offensive operations occurring on transnational communication networks raise important regulatory questions,[43] including the applicability of the law of armed conflict to so-called cyber-operations.[44] *Yoram Dinstein* dismisses the need for a revolution in the law of armed conflict necessitated by the advent of cyber warfare: 'this is by no means the first time in the history of LOAC that the introduction of a new weapon has created the misleading impression that great legal transmutations are afoot. Let me remind you of what happened upon the introduction of another new weapon, viz., the submarine.'[45] *Dinstein* recounts how the introduction of the submarine in World War I led to frantic calls for international legal regulation. But instead of comprehensive new conventional law, states eventually found the mere restatement that existing rules must also be observed by submarines sufficient. He concludes that were an international convention on cyber warfare to be concluded today, 'it would similarly stipulate in an anodyne fashion that the general rules of LOAC must be conformed with.'[46] *Gary Solis* likewise opens the requisite chapter in his magisterial textbook by stating categorically: 'This discussion is out of date. Cyber warfare policy and strategies evolve so rapidly that is difficult to stay current.' But what is changing are technologies, policies, and strategies, not the law: 'Actually, cyber warfare issues may be resolved in terms of traditional law of war concepts, although there is scant demonstration of its application because, so far, instances of actual cyber warfare have been unusual. Although cyber questions are many, the law of war offers as many answers.'[47] Concrete answers will depend on facts that are difficult to ascertain, due to inherent technical difficulties to forensic analysis in an extremely complex, deliberately heterogeneous network composed of a multitude of actors, both private and public, benign and malign. Legal assessments likewise rely on definitional disputes and normative interpretations that reflect shifting, often short-term, policies and strategies. Given vastly divergent national interests and capabilities, no uniform international understanding, let alone treaty regulation has emerged.[48]

In sum, while AI relies heavily on the same technical infrastructure of an open, global information network, its utilisation in the national security field poses distinct operational and

[42] Department of Defence, 'Cyber Strategy 2018 – Summary' (2018) https://media.defense.gov/2018/Sep/18/2002041658/-1/-1/1/CYBER_STRATEGY_SUMMARY_FINAL.PDF 1.

[43] Kohl, *The Net and the Nation State* (n 24); DeNardis, 'The Emerging Field of Internet Governance' (n 7).

[44] Boothby, 'Cyber Capabilities' (n 19); WH Boothby, 'Methods and Means of Cyber Warfare' (2013) 89 *Int'l L Stud* 387.

[45] RN Chesney, 'Computer Network Operations and US Domestic Law: An Overview' (2013) 87 *International Law Studies* 218, 286.

[46] Ibid, 287.

[47] Solis, *The Law of Armed Conflict* (n 21) 673.

[48] But note the highly representative Tallinn Manual, see W Heintschel von Heinegg, 'Chapter 1: The Tallinn Manual and International Cyber Security Law' (2012) 15 *YBIHL* 3.

https://doi.org/10.1017/9781009207898.035 Published online by Cambridge University Press

legal challenges not fully encompassed by the law of 'cyber warfare'.[49] That area of law presents the lawyer primarily with the challenge of applying traditional legal concepts to novel technical situations, especially the evidentiary challenges of defining and determining an armed attack, establishing attribution, the scope of the right to self-defence and proportionality, as well as thorny questions of the treatment of non-state or quasi-state actors, the classification of conflicts, and not least the threshold of the 'use of force'.[50] AI sharpens many of the same regulatory *conundra*, while creating novel operational risks and opportunities.[51]

III. CATASTROPHIC RISK: DOOMSDAY MACHINES

In the latest instalment of the popular Star Wars movie franchise, there is a key scene where the capabilities of truly terrible robotic fighting machines are presented. The franchise's new hero, the eponymous *Mandalorian*, manages only with considerable difficulty to defeat but one of these robots, of which, however, an entire battalion is waiting in the wings. The designers of the series have been praised for giving audiences 'finally an interesting stormtrooper', that is a machine capable of instilling fear and respect in the viewer.[52]

Whatever the cineastic value of these stormtroopers, in a remarkable coincidence a real robotics company simultaneously released a promotional video of actual robots that made these supposedly frightening machines set in a far distant future look like crude, unsophisticated toys. The dance video released by Boston Dynamics in early 2021 to show off several of its tactical robots jumping, dancing, pirouetting elegantly to music put everything Hollywood had come up with to shame: these were no prototypes, but robots that had already been deployed to police departments[53] and the military,[54] doing things that one previously could only have imagined in computer generated imagery.[55] Impressive and fearsome as these images are, these robots do exhibit motional 'intelligence' in the sense that they are able to make sense of their surroundings and act purposefully in it, but they are hardly able to replicate, let alone compete with human action, yet.

The impressive, even elegant capabilities showcased by these robots show that AI has made dramatic strides in recent years, bringing to mind ominous fears. In an early paper written in 1965, one of the British Bletchley Park cryptographers, the pioneering computer scientist and friend of *Alan Turing*, *Irving John 'Jack' Good* warned that an 'ultra-intelligent machine' would be built in the near future that could prove to be mankind's 'last invention' because it would lead to an 'intelligence explosion', that is an exponential increase in self-generating machine

[49] An excellent overview is provided by Solis, *The Law of Armed Conflict* (n 22) 673–709.

[50] MN Schmitt, 'The Law of Cyber Warfare: Quo Vadis?' (2014) 25 *Stanford Law & Policy Review* 269, 279.

[51] See in Baker, *The Centaur's Dilemma* (n 14) 69–94.

[52] J Hellerman, '"The Mandalorian" Finally Gives Us an Interesting Stormtrooper' (*No Film School Blog*, 18 December 2020) https://nofilmschool.com/storm-troopers-dumb.

[53] A Olla, 'A Dystopian Robo-Dog Now Patrols New York City. That's the Last Thing We Need' *The Guardian* (2 March 2021) www.theguardian.com/commentisfree/2021/mar/02/nypd-police-robodog-patrols.

[54] The humanoid Russian FEDOR tactical robot has already been deployed to the International Space Station, L Grush, 'Russia's Humanoid Robot Skybot Is on Its Way Home After a Two-Week Stay in Space' (*The Verge*, 6 September 2019) www.theverge.com/2019/9/6/20852602/russia-skybot-fedor-robot-international-space-station-soyuz.

[55] The video carried a note that these were not digital images but real footage of actual robots. See also E Ackerman, 'How Boston Dynamics Taught Its Robots to Dance' (*IEEE Spectrum*, 7 January 2021) https://spectrum.ieee.org/automaton/robotics/humanoids/how-boston-dynamics-taught-its-robots-to-dance; B Gilbert, 'Watch a Rare Video of Robots Jumping and Dancing Inside One of America's Leading Robotics Firms' *Business Insider* (29 March 2021) www.businessinsider.com/video-robots-jumping-and-dancing-inside-boston-dynamics-2021-3.

https://doi.org/10.1017/9781009207898.035 Published online by Cambridge University Press

intelligence.[56] While highly agile tactical robots conjure tropes of dangerous machines enslaving humanity, the potential risk posed by the emergence of super-intelligence is unlikely to take either humanoid form or motive but constitutes both incredible opportunity and existential risk, as *Good* pointed out half a century ago:

> The survival of man depends on the early construction of an ultra-intelligent machine. . . . Let an ultra-intelligent machine be defined as a machine that can far surpass all the intellectual activities of any man however clever. Since the design of machines is one of these intellectual activities, an ultra-intelligent machine could design even better machines; there would then unquestionably be an 'intelligence explosion,' and the intelligence of man would be left far behind. Thus, the first ultra-intelligent machine is the last invention that man need ever make, provided that the machine is docile enough to tell us how to keep it under control. It is curious that this point is made so seldom outside of science fiction. It is sometimes worthwhile to take science fiction seriously.[57]

Good would have been pleased to learn that both the promise and premonition of AI are no longer the preserve of science fiction, but taken seriously at the highest level of political decision-making. In a well-reported speech, President Vladimir Putin of Russia declared in 2017 that leadership in AI: 'is the future, not only for Russia, but for all humankind. It comes with colossal opportunities, but also threats that are difficult to predict. Whoever becomes the leader in this sphere will become the ruler of the world.'[58] Very similar statements guide official policy in all great powers, raising the spectre of what has been termed an 'arms race' in AI,[59] as a result of which 'super-intelligent' machines (i.e. those with capabilities higher than humans across the board), might endanger mankind.[60]

It is interesting to note that the tone of the debate has changed significantly. Writing in a popular scientific magazine in 2013, *Seth Baum* asked rhetorically whether his readers should even take the topic seriously: 'After all, it is essentially never in the news, and most AI researchers don't even worry. (AGI today is a small branch of the broader AI field.) It's easy to imagine this to be a fringe issue only taken seriously by a few gullible eccentrics.'[61] Today, these statements are no longer true. As Artificial General Intelligence, and thus the prospect of super-intelligence, is becoming a prominent research field, worrying about its eventual security implications is no longer the preserve of 'a few gullible eccentrics'. *Baum* correctly predicted that the relative lack of public and elite attention did not mean that the issue was unimportant.

Comparing it to the issue of climate change that likewise took several decades to evolve from a specialist concern to an all-consuming danger, he predicted that the trend was clear that given the exponential development of technology, the issue would soon become headline news.

[56] IJ Good, 'Speculations Concerning the First Ultraintelligent Machine' (1966) 6 *Advances in Computers* 31.

[57] Ibid, 31, 33, references omitted.

[58] J Vincent, 'Putin says the nation that leads in AI "will be the ruler of the world"', The Verge (4 September 2017) https://www.theverge.com/2017/9/4/16251226/russia-ai-putin-rule-the-world.

[59] The comprehensive study commissioned by the European Parliament on this topic lists existential risk only as the last item of twelve 'ethical harms and concerns' currently tackled by national and international regulatory efforts; E Bird and others, 'The Ethics of Artificial Intelligence: Issues and Initiatives' (*European Parliament*, March 2020) www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)634452_EN.pdf> 42–43 (hereafter Bird and others, 'The Ethics of Artificial Intelligence').

[60] See also the section on 'Safety and Beneficence of Artificial General Intelligence (AGI) and Artificial Superintelligence (ASI)' in M Bourgon and R Mallah, 'Ethically Aligned Design – A Vision for Prioritizing Human Well-Being with Autonomous and Intelligent Systems, (1st ed.)' (*IEEE*, 2019) https://ethicsinaction.ieee.org (hereafter Bourgon and Mallah, 'Ethically Aligned Design').

[61] S Baum, 'Our Final Invention: Is AI the Defining Issue for Humanity?' *Scientific American* (11 October 2013) https://blogs.scientificamerican.com/guest-blog/our-final-invention-is-ai-the-defining-issue-for-humanity/.

https://doi.org/10.1017/9781009207898.035 Published online by Cambridge University Press

The same point was made roughly at the same time by the co-founder of the Centre for the Study of Existential Risk (CSER) at the University of Cambridge, *Huw Price*. Summing up the challenge accurately, Price acknowledged that some of these concerns might seem far-fetched, the stuff of science fiction, which is exactly part of the problem:

> The basic philosophy is that we should be taking seriously the fact that we are getting to the point where our technologies have the potential to threaten our own existence – in a way that they simply haven't up to now, in human history. We should be investing a little of our intellectual resources in shifting some probability from bad outcomes to good ones. To the extent – presently poorly understood – that there are significant risks, it's an additional danger if they remain for these sociological reasons outside the scope of 'serious' investigation.[62]

There are two basic options: either to design safe AI with appropriate standards of transparency and ethical grounding as inherent design features, or not to design dangerous AI.[63] Given the attendant opportunities and the competitive international and commercial landscape, this latter option remains unattainable. Consequently, there has been much scientific thinking on devising ethical standards to guide responsible further technological development.[64] International legal regulation, in contrast, has so far proven elusive, and national efforts remain embryonic.[65]

Some serious thinkers and entrepreneurs argue that the development of super-intelligence must be abandoned due to inherent, incalculable, and existential risks.[66] Prudence would indicate that even a remote risk of a catastrophic outcome should keep all of us vigilant. Whatever the merits of these assessments, it appears unlikely that an international ban of such research is likely. Moreover, as *Ryan Calo* and others have pointed out, there is a real opportunity cost in focusing too much on such remote but highly imaginative risks.[67]

While the risks of artificial super-intelligence, which is defined as machine intelligence that surpasses the brightest human minds, are still remote, they are real and may quickly threaten human existence by design or indifference. Likewise, general AI or human-level machine intelligence remains largely aspirational, referring to machines that can emulate human beings at a range of tasks, switching fluidly between them, training themselves on data and their own past performance, and re-writing their operating code. In contrast, concrete policy and regulatory challenges need to be addressed now as a result of the exponential development of the less fearsome but concrete narrow AI, defined as machines that are as good or better than humans at particular tasks, such as interpreting x-ray or satellite images.

These more mundane systems are already operational and rapidly increase in importance, especially in the military field. Here, perhaps even more than in purely civilian domains, *Pedro Domingos*' often quoted adage seems fitting: 'People worry that computers will get too smart and

[62] F Lewsey, 'Humanity's Last Invention and Our Uncertain Future' (*University of Cambridge*, 25 November 2012) www.cam.ac.uk/research/news/humanitys-last-invention-and-our-uncertain-future.

[63] To some extent, this debate is already moot because automated strategic nuclear defence systems have existed – and likely remain operational – in both Russia and the United States, see n 27.

[64] The evolving scientific, industry, and governmental consensus about the principles necessary to ensure responsible and safe AI have been outlined *inter alia* in Bourgon and Mallah, 'Ethically Aligned Design' (n 60): 'Asilomar Principles on Intelligent Machines and Smart Policies – Research Issues, Ethics and Values, Longer-Term Values' (*Future of Life Institute*, 2017) futureoflife.org/ai-principles.

[65] For an overview of national efforts, see Bird and others, 'The Ethics of Artificial Intelligence' (n 59).

[66] For an approving summary of these arguments, see J Barratt, *Our Final Invention* (2013).

[67] R Calo, 'Artificial Intelligence Policy: A Primer and Roadmap' (2017) 51 *University of California Davis Law Review* 399–435 (hereafter Calo, 'Artificial Intelligence Policy').

https://doi.org/10.1017/9781009207898.035 Published online by Cambridge University Press

take over the world, but the real problem is that they're too stupid and they've already taken over the world.'[68] Without belittling the risk of artificial general or super-intelligence, *Calo* is thus correct to stress that focusing too much attention on this remote risk will reduce necessary attention from pressing societal needs and thereby risk 'an AI Policy Winter' in which necessary regulation limps behind rapid technical development.[69]

IV. AUTONOMOUS WEAPONS SYSTEM

Automated weapons have been in use for a long time; how long depends largely on the degree of automation informing one's definition. A broad definition of a robot, under which we can subsume autonomous weapons systems, is a physical system that senses, processes, and acts upon the world. We can thus differentiate between 'disembodied AI' which collects, processes, and outputs data and information, but whose effect in the physical world is mediated; and robotics which leverage AI to itself physically act upon the world.[70]

In order to ascertain the likely impact of AI on autonomous weapons systems, it is helpful to conceive of them and the regulatory challenges they pose as a spectrum of capabilities rather than sharply differentiated categories, with booby traps and mines on one end; improvised explosive devices (IEDs), torpedoes, and self-guided rockets somewhere in the middle; drones and loitering munition further towards the other end; and automated air defence and strategic nuclear control systems at or beyond the other polar end. It appears that two qualitative elements are crucial: the degree of processing undertaken by the system,[71] and the amount of human involvement before the system acts.[72]

It follows that the definition of 'autonomous' is not clear-cut, nor is it likely to become so. Analytically, one can distinguish four distinct levels of autonomy: human operated, human delegated, human supervised, and fully autonomous.[73] These classifications, however, erroneously 'imply that there are discrete levels of intelligence and autonomous systems',[74] downplaying the importance of human–machine collaboration.[75] Many militaries, most prominently that of the US, insist that a human operator must remain involved, including 'fail safe' security precautions:

[68] P Domingos, *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World* (2015) 286.

[69] Calo, 'Artificial Intelligence Policy' (n 67) 435.

[70] Calo, 'Artificial Intelligence Policy' (n 67) 407. Calo argues that the respective legal assessment is likely to be different; see also HY Liu, 'Refining Responsibility: Differentiating Two Types of Responsibility Issues Raised by Autonomous Weapons Systems' in C Kreß and others (eds), *Autonomous Weapons Systems: Law, Ethics, Policy* (2016) (hereafter Liu, 'Refining Responsibility').

[71] See further HY Liu, 'Categorization and Legality of Autonomous and Remote Weapons Systems' (2012) 94 *Int'l Rev of the Red Cross* 627; G Sartor and O Andrea, 'The Autonomy of Technological Systems and Responsibilities for their Use' in C Kreß and others (eds), *Autonomous Weapons Systems: Law, Ethics, Policy* (2016).

[72] See further N Sharkey, 'Staying in the Loop: Human Supervisory Control of Weapons' in C Kreß and others (eds), *Autonomous Weapons Systems: Law, Ethics, Policy* (2016) (hereafter Sharkey, 'Staying in the Loop'); GS Corn, 'Autonomous Weapons Systems: Managing the Inevitability of 'Taking the Man Out of the Loop'' in C Kreß and others (eds), *Autonomous Weapons Systems: Law, Ethics, Policy* (2016) (hereafter Corn, 'Autonomous Weapons Systems'); D Saxon, 'A Human Touch: Autonomous Weapons, DoD Directive 3000.09 and the Interpretation of 'Appropriate Levels of Human Judgment over the Use of Force'' in C Kreß and others (eds), *Autonomous Weapons Systems: Law, Ethics, Policy* (2016) (hereafter Saxon, 'A Human Touch').

[73] G Galdorisi, 'Keeping Humans in the Loop' (2015) 141/2/1,344 US Naval Institute Proceedings 36, 38.

[74] Department of Defense, Defense Science Board, 'Task Force Report: The Role of Autonomy in DoD Systems' (*US Department of Defense*, July 2012) 4 https://fas.org/irp/agency/dod/dsb/autonomy.pdf.

[75] G Galdorisi, 'Keeping Humans in the Loop' (2015) 141/2/1,344 US Naval Institute Proceedings 36; Sharkey, 'Staying in the Loop' (n 72).

https://doi.org/10.1017/9781009207898.035 Published online by Cambridge University Press

> Semi-autonomous weapons systems that are onboard or integrated with unmanned platforms must be designed such that, *in the event of degraded or lost communications*, the system does *not autonomously select and engage* individual targets or specific target groups that have not been previously selected by an authorized human operator. It is DoD policy that . . . autonomous and semi-autonomous weapons systems shall be designed to allow commanders and operators to exercise appropriate levels of human judgment of the use of force.[76]

In contrast to the assumptions underlying the discussion in the previous section, even fully autonomous systems currently always involve a human being who 'makes, approves, or overrides a fire/don't fire decision'.[77] Furthermore, such systems have been designed by humans, who have programmed them within specified parameters, which include the need to observe the existing law of armed conflict.[78] These systems are deployed into battle by human operators and their commanders,[79] who thus carry command responsibility,[80] including the possible application of strict liability standards known from civil law.[81]

Given the apparent military benefits of increased automation and an extremely dynamic, easily transferable civilian field, outright bans of autonomous weapon systems, robotics, and unmanned vehicles appear 'insupportable as a matter of law, policy, and operational good sense'.[82] To be sure, some claim that the principles of distinction, proportionality, military necessity, and the avoidance of unnecessary suffering, which form the basis of the law of armed conflict,[83] in conjunction with general human rights law,[84] somehow impose a 'duty upon individuals and states in peacetime, as well as combatants, military organizations, and states in armed conflict situations, not to delegate to a machine or automated process the authority or capability to initiate the use of lethal force independently of human determinations of its moral and legal legitimacy in each and every case.'[85] Without restating the copious literature on this topic, it is respectfully suggested that such a duty for human determination cannot be found in existing international, and only occasionally in national,[86] law. *Solis*' textbook begins discussing the war crime liability of autonomous weapons by stating the

[76] Directive 3000.09: Autonomy in Weapons Systems, Unmanned Systems Integrated Roadmap, FY 2013–2038, US Department of Defence, Washington D.C. (21 November 2012); Solis, *The Law of Armed Conflict* (n 22) 537, my emphasis.

[77] Solis, *The Law of Armed Conflict* (n 22) 537.

[78] P Asaro, 'On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanization of Lethal Decision-Making' (2012) 94 *Int'l Rev of the Red Cross* 687, 691 (hereafetr Asaro, 'On Banning Autonomous Weapon Systems').

[79] MN Schmitt and JS Thurnher, ''Out of the Loop': Autonomous Weapons Systems and the Law of Armed Conflict' (2013) 4 *Harvard National Security Journal* 231, 235 (hereafter Schmitt and Thurnher, 'Out of the Loop').

[80] Sharkey, 'Staying in the Loop' (n 72); Liu, 'Refining Responsibility' (n 70).

[81] Calo, 'Artificial Intelligence Policy' (n 67) 418; R Calo, 'Robotics and the Lessons of Cyberlaw' (2015) 103 *California Law Review* 513, 538–545 (hereafter Calo, 'Robotics and the Lessons of Cyberlaw').

[82] Schmitt and Thurnher, 'Out of the Loop' (n 79) 233.

[83] Solis, *The Law of Armed Conflict* (n 22) 268–327, 539–541, 551–552.

[84] M Milanovic, 'The Lost Origins of Lex Specialis: Rethinking the Relationship between Human Rights and International Humanitarian Law' in JD Ohlin (ed), *Theoretical Boundaries of Armed Conflict and Human Rights* (2015); G Pinzauti, 'Good Time for a Change: Recognizing Individuals' Rights under the Rules of International Humanitarian Law on the Conduct of Hostilities' in A Cassese (ed), *Realizing Utopia: The Future of International Law* (2012); T Meron, 'On the Inadequate Reach of Humanitarian and Human Rights Law and the Need for a New Instrument' (1983) 77 *AJIL* 589–606; T Meron, *Human Rights and Humanitarian Norms as Customary Law* (1991).

[85] Asaro, 'On Banning Autonomous Weapon Systems' (n 78) 687; E Lieblich and B Eyal, 'The Obligation to Exercise Discretion in Warfare: Why Autonomous Weapons Systems Are Unlawful' in C Kreß and others (eds), *Autonomous Weapons Systems: Law, Ethics, Policy* (2016).

[86] The American military, it is remembered, formally maintains that it is bound by such a duty, as a matter of internal policy. Whether this amounts to a legal obligation under domestic law remains a matter of some dispute; see further Department of Defence, *Directive 3000.09: Autonomy in Weapons Systems* (n 77); Saxon, 'A Human Touch' (n 72).

https://doi.org/10.1017/9781009207898.035 Published online by Cambridge University Press

obvious counter-factual: ‘Any lawful weapon can be employed unlawfully.’ He proceeds to devise a number of hypothetical scenarios in which autonomous weapons could indeed be used or deliberately designed unlawfully, to conclude:

> The likelihood of an autonomous weapon system being unlawful in and of itself is very remote; it would not meet Article 36 testing requirements and thus would not be put into use. And the foregoing four scenarios involving possible unlawful acts by operators or manufacturers are so unlikely, so phantasmagorical, that they are easily lampooned. . . . While acts such as described in the four scenarios are unlikely, they are possible.[87]

As stated, Article 36 of the 1977 Additional Protocol I to the Geneva Conventions imposes on the contracting parties the obligation to determine prior to the deployment of any new weapon that it conforms with the existing law of armed conflict and ‘any other rule of international law applicable’. For states developing new weapons, this obligation entails a continuous review process from conception and design, through its technological development and prototyping, to production and deployment.[88]

Given the complexity and rapid continuous development of autonomous weapons systems, especially those relying on increasingly sophisticated AI, such a legally mandatory review will have to be continuous, rigorous, and overcome inherent technical difficulties, given the large number of sub-systems from a large number of providers. Such complexity notwithstanding, autonomous weapons, including those relying on AI, are not unlawful in and of themselves.

In principle, the underlying ethical *conundra* and proportional balancing of competing values that need to inform responsible robotics generally,[89] need to inform the conception, design, deployment, and use of autonomous weapons system, whether or not powered by AI: ‘I reject the idea that IHL [international humanitarian law] is inadequate to regulate autonomous weapons. . . . However far we go into the future and no matter how artificial intelligence will work, there will always be a human being at the starting point . . . This human being is bound by the law.’[90] The most likely use scenarios encompass so-called narrow AI where machines have already surpassed human capabilities. The superior ability to detect patterns in vast amounts of unstructured (sensory) data has for many years proven indispensable for certain advanced automated weapons systems. Anti-missile defence systems, like the American maritime Aegis and land-based Patriot, the Russian S300 and S400 or the Israeli ‘Iron Dome’, all rely on the collection and processing of large amounts of radar and similar sensor data, and the ability to respond independently and automatically. This has created unique vulnerabilities: their susceptibility to cyber-attacks ‘blinding’ them,[91] the dramatic shortening of warning and reaction time even where human operators remain ‘in the loop’,[92] and the possibility to render these

[87] Solis, *The Law of Armed Conflict* (n 22) 543.

[88] International Committee of the Red Cross (ICRC), *A Guide to the Legal Review of New Weapons, Means and Methods of Warfare* (2006) 23.

[89] See generally Calo, ‘Robotics and the Lessons of Cyberlaw’ (n 81).

[90] See further M Sassòli, ‘Autonomous Weapons and International Law: Advantages, Open Technical Questions and Legal Issues to be Clarified’ (2014) 90 *International Law Studies* 308, 323; likewise, Schmitt and Thurnher, ‘Out of the Loop’ (n 79) 277.

[91] Note for instance the Israeli electronic disabling of Syria’s expensive, Russian-made air defence system prior to their bombing of a half-constructed nuclear power reactor in 2007, discussed in Solis, *The Law of Armed Conflict* (n 22) 677.

[92] This has been the main excuse offered by the Captain of the US warship *Vincennes* for shooting down an Iranian civilian airliner in 1988. With AI, this problem is likely to become much more acute. For a discussion of the former, see ibid, 563–566. For the latter, see Baker, ‘Artificial Intelligence and National Security Law’ (n 3).

https://doi.org/10.1017/9781009207898.035 Published online by Cambridge University Press

expensive, highly sophisticated systems economically unviable by targeting them with unconventional countermeasures, such as very cheap, fairly simple commercial drones.[93]

V. EXISTING MILITARY CAPABILITIES

Irrespective of the legal and ethical questions raised, AI is having a transformative effect on the operational and economic viability of many sophisticated weapons systems. The existing military technology perhaps most immediately affected by the rise of AI are unmanned vehicles of various kinds, so-called drones and 'loitering munitions'.[94] Currently relying on remote guidance by human operators or relatively 'dumb' automation, their importance and power is likely to increase enormously if combined with AI. Simultaneously, certain important legacy systems, for instance large surface ships such as aircraft carriers, can become vulnerable and perhaps obsolete due to neurally linked and (narrowly) artificially intelligent 'swarms' of very small robots.[95]

The ready availability of capable and affordable remotely operated vehicles, plus commercial satellite imagery and similar information sources has put long-range power-projection capabilities in the hands of a far larger group of state and non-state actors. This equalisation of relative power is further accelerated by new technology rendering existing weapon systems vulnerable or ineffective. Important examples include distributed, swarm-like attacks on ships or permeating expensive air defence systems with cheap, easily replaceable commercial drones.[96]

The recent war over Nagorno-Karabakh exposed some of these general vulnerabilities, not least the inability of both Armenia and Azerbaijan's short-range air defense (SHORAD) arsenals, which admittedly were limited in size and quality, to protect effectively against sophisticated drones. While major powers like the US, China, and Russia are developing and deploying their own drone countermeasures,[97] certain existing systems, for instance aircraft carriers, have become vulnerable. This portends potential realignments in relative power where large numbers of low-cost expendable machines can be used to overwhelm an otherwise superior adversary.[98]

There has been much academic speculation about the perceived novelty of drone technology and the suggested need to update existing legal regulations.[99] It needs to be stated from the outset that remotely piloted land-, air-, or sea-crafts have been used since the 1920s,[100] and thus

[93] Note for instance the successful use by Houthi militias in Yemen and by Hamas in Gaza of very cheap commercial drones as deliberate targets for very expensive Israeli, Emirati, and Saudi Patriot air defence systems; see A Kurth Cronin, *Power to the People: How Drones, Data and Dynamite Empower and Imperil Our Security* (2019) 213.

[94] These are sometimes called 'suicide drones.' For an excellent technical overview, see D Gettinger and HM Arthur, 'Loitering Munitions' (*CSD Bard*, 2017) https://dronecenter.bard.edu/files/2017/02/CSD-Loitering-Munitions.pdf.

[95] T McMullan, 'How Swarming Drones Will Change Warfare' (*BBC News*, 16 March 2019) www.bbc.com/news/technology-47555588 (hereafter McMullan, 'How Swarming Drones Will Change Warfare; SM Williams, 'Swarm Weapons: Demonstrating a Swarm Intelligent Algorithm for Parallel Attack' (2018) https://apps.dtic.mil/sti/pdfs/AD1071535.pdf (hereafter Williams, 'Swarm Weapons').

[96] R Martinage, 'Toward a New Offset Strategy – Exploiting US Long-Term Advantages to Restore US Global Power Projection Capability' (CSBA, 2014) https://csbaonline.org/uploads/documents/Offset-Strategy-Web.pdf 23–28 (hereafter Martinage, 'Toward a New Offset Strategy').

[97] S Shaikh and R Wes, 'The Air and Missile War in Nagorno-Karabakh: Lessons for the Future of Strike and Defense' (CSIC, 8 December 2020) www.csis.org/analysis/air-and-missile-war-nagorno-karabakh-lessons-future-strike-and-defense (hereafter Shaikh and Wes, 'Lessons for the Future of Strike and Defense').

[98] McMullan, 'How Swarming Drones Will Change Warfare' (n 95); Williams, 'Swarm Weapons' (n 95).

[99] For an overview, see PL Bergen and D Rothenberg (eds), *Drone Wars: Transforming Conflict, Law, and Policy* (2015) (hereafter Bergen and Rothenberg, Drone Wars).

[100] K Kakaes, 'From Orville Wright to September 11: What the History of Drone Technology Says about Its Future' in Bergen and Rothenberg, *Drone Wars: Transforming Conflict, Law, and Policy* (2015) (hereafter Kakaes, 'From Orville Wright to September 11').

https://doi.org/10.1017/9781009207898.035 Published online by Cambridge University Press

cannot be considered either new or unanticipated by the existing law of armed conflict.[101] Likewise, it is difficult to draw a sharp technical distinction between certain drones and some self-guided missiles, which belong to a well-established area of military operations and regulation.[102]

The novelty lies less in the legal or ethical assessment, than in the operational challenge of the dispersal of a previously highly exclusive military capability. The US has twice before responded to such a loss of its superior competitive edge by embarking on an 'offset' strategy meant to avoid having to match capabilities, instead seeking to regain superiority through an asymmetric technological advantage.[103]

The 'First Offset' strategy successfully sought to counter Soviet conventional superiority through the development and deployment of, especially tactical, nuclear weapons.[104] The 'Second Offset' strategy was begun towards the end of the Vietnam War and reached its successful conclusion during the Iraq War of 1991. It meant to counter the quantitative equalisation of conventional assets, especially airpower, not by increasing the number of assets but their quality. Mustering American socio-economic advantages in technological sophistication, the key to the strategy was the development of previously unimaginable strike precision. As with any other military technology, it was anticipated that the opponent would eventually catch up, at some point neutralising this advantage. Given the economic near-collapse of the Soviet Union and its successor Russia, the slow rise of China, and the relative absence of other serious competitors, the technological superiority the US had achieved in precision strike capability surprisingly endured far longer than anticipated:

> Perhaps the most striking feature of the evolution of non-nuclear (or conventional) precision strike since the Cold War ended in 1991 has been what has not happened. In the early 1990s, there was growing anticipation that for major powers such as the United States and Russia, 'long-range precision strike' would become 'the dominant operational approach.' The rate at which

[101] For a good overview, see Solis, *The Law of Armed Conflict* (n 22) 545–554. The claim that the existing law of armed conflict is inadequate for the actual conflict at hand is probably as old as the truism that this body of law is 'always one war behind.' While there is some truth in the latter observation, the first is usually little more than exculpatory. Both discussions are as old as humanitarian law itself and it is unlikely that the rise of either drone technology or AI will do much to affect its basis parameters, namely the basic adequacy of existing legal principles. For the debate as such, see *inter alia* T Meron, 'Customary Humanitarian Law Today: From the Academy to the Courtroom' in A Clapham and P Gaeta (eds), *The Oxford Handbook of International Law in Armed Conflict* (2014); MN Schmitt and S Watts, 'State Opinio Juris and International Humanitarian Law Pluralism' (2015) 91 *International Law Studies* 171–215; G Best, *Humanity in Warfare: The Modern History of the International Law of Armed Conflicts* (1980); T Meron, 'Humanization of Humanitarian Law' (2000) 94 *AJIL* 239–278.

[102] See generally Y Dinstein, 'International Humanitarian Law Research Initiative: IHL in Air and Missile Warfare' (2006) www.ihlresearch.org/amw/; Y Dinstein, 'The Laws of Air, Missile and Nuclear Warfare' (1997) 27 *Isr Y B Hum Rts* 1–16.

[103] O Manea and RO Work, 'The Role of Offset Strategies in Restoring Conventional Deterrence' (2018) *Small Wars Journal* https://smallwarsjournal.com/jrnl/art/role-offset-strategies-restoring-conventional-deterrence (hereafter Manea and Work, 'The Role of Offset Strategies'); RR Tomes, 'The Cold War Offset Strategy: Assault Breaker and the Beginning of the RSTA Revolution' (*War on the Rocks*, 20 November 2014) https://warontherocks.com/2014/11/the-cold-war-offset-strategy-assault-breaker-and-the-beginning-of-the-rsta-revolution/ (hereafter Tomes, 'The Cold War Strategy').

[104] 'Since we cannot keep the United States an armed camp or a garrison state, we must make plans to use the atom bomb if we become involved in a war.' President Eisenhower in 1953, quoted in Martinage, 'Toward a New Offset Strategy' (n 96) 8. I have provided a brief history of the dynamic development of US nuclear strategy in E Afsah, 'Creed, Cabal, or Conspiracy: Origins of the Current Neo-Conservative Revolution in US Strategic Thinking' (2003) *GLJ* 902, 907–910; a fuller, accessible account can be found in DM Lawson and DB Kunsman, 'A Primer on US Strategic Nuclear Policy' (*OSTI*, 1 January 2001) www.osti.gov/servlets/purl/776355/ (hereafter Lawson and Kunsman, US Strategic Nuclear Policy').

https://doi.org/10.1017/9781009207898.035 Published online by Cambridge University Press

> this transformation might occur was anyone's guess but many American observers presumed that this emerging form of warfare would proliferate rather quickly. Not widely foreseen in the mid-1990s was that nearly two decades later long-range precision strike would still be a virtual monopoly of the US military.[105]

Written in 2013, this assessment is no longer accurate. Today, a number of states have caught up and dramatically improved both the precision and range of their power projection. The gradual loss of its relative monopoly with respect to precision strike capability, remote sensing, and stealth, while simultaneously exclusive assets like aircraft carrier groups are becoming vulnerable, ineffective, or fiscally unsustainable,[106] led the US to declare its intention to respond with a 'Third Offset' strategy. It announced in 2014 that it would counter potential adversaries asymmetrically, rather than system by system:

> Trying to counter emerging threats symmetrically with active defenses or competing 'fighter for fighter' is both impractical and unaffordable over the long run. A third offset strategy, however, could offset adversarial investments in A2/AD [anti-access/area denial] capabilities in general – and ever-expanding missile inventories in particular – by leveraging US core competencies in unmanned systems and automation, extended-range and low-observable air operations, undersea warfare, and complex system engineering and integration. A GSS [global surveillance and strike] network could take advantage of the interrelationships among these areas of enduring advantage to provide a balanced, resilient, globally responsive power projection capability.[107]

The underlying developments have been apparent for some time, 'disruptive technologies and destructive weapons once solely possessed by advanced nations' have proliferated and are now easily and cheaply available to a large number of state and non-state opponents, threatening the effectiveness of many extremely expensive weapon systems on which power-projection by advanced nations, especially the US, had relied.[108] One of these disruptive technologies has been unmanned vehicles, especially airborne 'drones'. While these have been used for a century and have been militarily effective for half a century,[109] the explosion in surveillance and reconnaissance capability afforded by AI, and the dramatic miniaturisation and commercialisation of many of the underlying key components have transformed the global security landscape by making these capabilities far more accessible.[110]

Drones have proven their transformative battlefield impact since the 1973 Yom Kippur War and 1982 Israeli invasion of Lebanon.[111] Whatever their many operational and strategic benefits, unmanned aircraft were initially not cheaper to operate than conventional ones: 'higher costs for personnel needed to monitor and analyze data streams that do not exist on manned platforms, as well as the costs for hardware and software that go into the sensor packages,'[112] to say nothing of

[105] BD Watts, 'The Evolution of Precision Strike' (*CSBA*, 2013) https://csbaonline.org/uploads/documents/Evolution-of-Precision-Strike-final-v15.pdf 1–2, references omitted.

[106] Martinage, 'Toward a New Offset Strategy' (n 96) 17–20, 72.

[107] Martinage, 'Toward a New Offset Strategy' (n 96) 72.

[108] US Defence Secretary Chuck Hagel outlined these threats in a programmatic speech on 3 September 2014, which explicitly drew an analogy to Eisenhower's 'first offset' strategy and committed the country to invest in asymmetric, high-technology counter-measures, including AI, see *inter alia* Martinage, 'Toward a New Offset Strategy' (n 96) i.

[109] Their history is well summarised in Kakaes, 'From Orville Wright to September 11' (n 100).

[110] See generally Bergen and Rothenberg, *Drone Wars* (n 99).

[111] Kakaes, 'From Orville Wright to September 11' (n 100) 375.

[112] J Abizaid and R Brooks, 'Recommendations and Report of the Task Force on US Drone Policy' (*Stimson*, April 2015) www.stimson.org/wp-content/files/file-attachments/recommendations_and_report_of_the_task_force_on_us_drone_policy_second_edition.pdf 23.

https://doi.org/10.1017/9781009207898.035 Published online by Cambridge University Press

the considerable expense of training their pilots,[113] left drones and the long-range precision targeting capability they conferred out of the reach of most armies, primarily due to economic costs, skilled manpower shortages, and technological complexity.

The recent conflict between Azerbaijan and Armenia has decisively shown that these conditions no longer hold. Both are relatively poor nations with fairly unsophisticated armed forces, with the crucial suppliers being the medium powers of Turkey and Israel. This highlighted the dramatic availability and affordability of such technology,[114] much of it off-the-shelf and available through a number of new entrants in the market, raising important questions of export controls and procurement.[115] Drone technology and their transformational impact on the battlefield are no longer the prerogative of rich industrial nations. While AI does not appear to have played a large role in this conflict yet,[116] the decisiveness of the precision afforded by long-range loitering munition, unmanned vehicles, and drastically better reconnaissance,[117] has not been lost on more traditional great powers.[118]

This proliferation of precision long-range weaponry portends the end of the enormous advantages enjoyed by the US as a result of its 'Second Offset' strategy. Following the Vietnam War, the US successfully sought to counteract the perceived[119] numerical superiority of the Soviet Union[120] in air and missile power by investing in superior high-precision weaponry, harnessing the country's broad technological edge.[121] These investments paid off and conferred a surprisingly long-lasting dominance. The loss of its main adversary and the inability of other adversaries to match its technological capabilities, meant that the unique advantages conferred to the US – primarily the ability to essentially eliminate risk to one's own personnel by striking remotely and to reduce political risk from 'collateral damage' by striking precisely – created an enduring willingness to deploy relatively unopposed in a vast number of unconventional conflict scenarios, sometimes dubbed a 'New American Way of War'.[122]

In principle, 'combat drones and their weapons systems are lawful weapons'.[123] Moreover, given inherent technical differences, especially their drastically higher loitering ability, lack of risk to personnel and higher precision, can actually improve observance of the law of armed

[113] Since 2009, the US Air Force has trained more drone than conventional pilots and the US Navy has announced in 2015 that the current F-35 will be the last manned strike fighter aircraft they will buy and operate, discussed in Solis, *The Law of Armed Conflict* (n 22) 547.

[114] The Turkish Bayraktar TB2 drone relies heavily on commercial civilian components, such as generic Garmin navigation systems. The UK defence minister remarked with respect to Turkey's new role as a supplier of weaponry, training, and intelligence that 'other countries are now leading the way' and that, therefore, the UK would itself begin to invest in such new, much cheaper drone technology; D Sabbagh, 'UK Wants New Drones in Wake of Azerbaijan Military Success' *The Guardian* (29 December 2020) www.theguardian.com/world/2020/dec/29/uk-defence-secretary-hails-azerbaijans-use-of-drones-in-conflict (hereafter Sabbagh, 'UK Wants New Drones').

[115] J Detsch, 'The US Army Goes to School on Nagorno-Karabakh Conflict – Off-the-Shelf Air Power Changes the Battlefield of the Future' Foreign Policy (30 March 2021) https://foreignpolicy.com/2021/03/30/army-pentagon-nagorno-karabakh-drones/.

[116] Ibid.

[117] Shaikh and Wes, 'Lessons for the Future of Strike and Defense' (n 97).

[118] Sabbagh, 'UK Wants New Drones' (n 114).

[119] There is good reason to doubt that this perceived inferiority actually existed, see Martinage, 'Toward a New Offset Strategy' (n 96) 11 *et seq*; Lawson and Kunsman, 'US Strategic Nuclear Policy' (n 104) 51–64.

[120] Manea and Work, 'The Role of Offset Strategies' (n 103).

[121] R Grant, 'The Second Offset' Air Force Magazine (24 June 2016) www.airforcemag.com/article/the-second-offset/; Tomes, 'The Cold War Offset Strategy' (n 103).

[122] RR Tomes, *US Defence Strategy from Vietnam to Operation Iraqi Freedom: Military Innovation and the New American Way of War, 1973–2003* (2006).

[123] Solis, The Law of Armed Conflict (n 22) 551.

https://doi.org/10.1017/9781009207898.035 Published online by Cambridge University Press

conflict by making it easier to distinguish and reduce 'collateral damage',[124] having led some to claim that not to use drones would actually be unethical.[125] Given vastly better target reconnaissance and the possibility for much more deliberate strike decisions, convincing arguments can be made that remotely operated combat vehicles are not only perfectly lawful weapons but have the potential to increase compliance with humanitarian objectives: 'While you can make mistakes with drones, you can make bigger mistakes with big bombers, which can take out whole neighborhoods. A B-2 [manned bomber] pilot has no idea who he is hitting; a drone pilot should know exactly who he is targeting.'[126] These very characteristics – the absence of risk to military personnel and vastly better information about battlefield conditions – have also made drone warfare controversial, aspects that are heightened but not created by the addition of AI. The relative absence of operational and political risk led to a greater willingness to use armed force as a tool of statecraft, in the process bending or breaking traditional notions of international law and territorial integrity.[127] Some have argued that remote warfare with little to no risk to the operator of the weapon is somehow unethical, somehow incompatible with the warrior code of honour, concerns that should, if anything, apply even more forcefully to machines killing autonomously.[128] Whatever the merits of the conception of fairness underlying such conceptions, such 'romantic and unrealistic views of modern warfare' do not reflect a legal obligation to expose oneself to risk.[129]

There is a legal obligation, however, to adequately balance risks resulting from obtaining military advantages, which include reducing exposing service-members to risk, and the principle of distinction meant to protect innocent civilians. Many years ago, *Stanley Hoffmann* denounced the perverse doctrine of 'combatant immunity' in the context of high altitude bombing by manned aircraft staying above the range of air defences despite the obvious costs in precision and thus civilian casualties this would entail.[130] In some respects, the concerns *Hoffmann* expressed have been addressed by unmanned aircraft, which today permit unprecedented levels of precision, deliberation, and thus observance of the principle of distinction:

> Drones are superior to manned aircraft, or artillery, in several ways. Drones can gather photographic intelligence from geographic areas too dangerous for manned aircraft. Drones carry no risk of friendly personnel death or capture. Drones have an operational reach greater than that of aircraft, allowing them to project force from afar in targets far in excess of manned aircraft. The accuracy of drone-fired munitions is greater than that of most manned aircraft, and that accuracy allows them to employ munitions with a kinetic energy far less than artillery or close air support require, thus reducing collateral damage.[131]

[124] Ibid, 551–553.

[125] B Wittes, 'Drones and Democracy: A Response to Firmin DeBrabander' (*Lawfare Blog*, 15 September 2014) www.lawfareblog.com/drones-and-democracy-response-firmin-debrabander.

[126] DE Sanger, Confront and Conceal: Obama's Secret Wars and Surprising Use of American Power (2012) 257 (hereafter Sanger, Confront and Conceal), quoted in Solis, *The Law of Armed Conflict* (n 22) 554.

[127] From the copious literature, see *inter alia* Y Dinstein, 'Concluding Observations: The Influence of the Conflict in Iraq on International Law' in RA Pedrozo (ed), *The War in Iraq: A Legal Analysis* (2010); M Sassòli, 'Ius ad Bellum and Ius in Bello: The Separation between the Legality of the Use of Force and Humanitarian Rules to be Respected in Warfare: Crucial or Outdated' in MN Schmitt and J Pejic (eds), *International Law and Armed Conflict: Exploring the Faultlines* (2007).

[128] See generally C Heyns, 'Autonomous Weapons Systems: Living a Dignified Life and Dying a Dignified Death' in C Kreß and others (eds), *Autonomous Weapons Systems: Law, Ethics, Policy* (2016); GS Corn, 'Autonomous Weapons Systems' (n 72).

[129] Solis, *The Law of Armed Conflict* (n 22) 553.

[130] S Hoffmann, 'The Politics and Ethics of Military Intervention' (1995) 37 *Survival* 29.

[131] Solis, *The Law of Armed Conflict* (n 22) 550.

https://doi.org/10.1017/9781009207898.035 Published online by Cambridge University Press

At the same time, however, the complete removal of risk to one's own personnel has reduced traditional inhibitions to engage in violence abroad,[132] including controversial policies of 'targeted killings'.[133] Many of the ethical and legal *conundra*, as well as operational advantages that ensured are heightened if the capability of remotely operated vehicles is married with AI, which can improve independent or pre-authorised targeting by machines.[134]

VI. RECONNAISSANCE

The previous section showed that the rapid development of AI is transforming existing military capabilities, leading to considerable adjustments in relative strength. As in the civilian field, the main driver is the removal of a key resource constraint, namely the substitution of skilled, thus expensive and often rare, manpower by machines no longer constrained by time, availability, emotions, loyalty, alertness, etc. The area where these inherent advantages are having the largest national security impact is reconnaissance and intelligence collection.[135]

It is not always easy to distinguish these activities clearly from electronic espionage, sabotage, and intellectual property theft discussed above, but it is apparent that the capabilities conferred by automated analysis and interpretation of vast amounts of sensor data is raising important regulatory questions related to privacy, territorial integrity, and the interpretation of classical ius in bello principles on distinction, proportionality, and military necessity.

The advantages of drones outlined just above[136] have conferred unprecedented abilities to pierce the 'fog of war' by giving the entire chain of command, from platoon to commander in chief, access to information of breathtaking accuracy, granularity, and actuality.[137] Such drone-supplied information is supplemented by enormous advances in 'signal and electronic intelligence', that is eavesdropping into communication networks to obtain information relevant for tactical operations and to make strategic threat assessments. But all this available information would be meaningless without someone to make sense of it. Just like in civilian surveillance,[138] the limiting factor has long been the human being needed to watch and interpret the video or

[132] Sanger, *Confront and Conceal* (n 126).

[133] A Barak, 'International Humanitarian Law and the Israeli Supreme Court' (2014) *Isr L Rev* 181; N Melzer, *Targeted Killing in International Law* (2008); J Ulrich, 'The Gloves Were Never On: Defining the President's Authority to Order Targeted Killing in the War against Terrorism' (2005) *Va J Int'l L* 1029; D Kretzmer, 'Targeted Killings of Suspected Terrorists: Extra-Judicial Execution or Legitimate Means of Defence?' (2005) 16 *EJIL* 171.

[134] P Kalmanovitz, 'Judgment, Liability and the Risks of Riskless Warfare' in C Kreß and others (eds), *Autonomous Weapons Systems: Law, Ethics, Policy* (2016); Saxon, 'A Human Touch' (n 72).

[135] See also G Allen and T Chan, 'Artificial Intelligence and National Security' (*Belfer Center*, July 2017) www.belfercenter.org/sites/default/files/files/publication/AI%20NatSec%20-%20final.pdf 27–35 (hereafter Allen and Chan, 'Artificial Intelligence and National Security').

[136] Solis, *The Law of Armed Conflict* (n 22) 550.

[137] See further S Smagh, 'Intelligence, Surveillance, and Reconnaissance Design for Great Power Competition' (*Congressional Research Service*, 4 June 2020) https://crsreports.congress.gov/product/pdf/R/R46389.

[138] The human factor is not only expensive and rare, it is also susceptible to bias, emotional attachment, and similar factors, which limit systemic reliability as a whole. The enormous human cost in both effort and emotional distortion in classical surveillance has been described with great artistic sensibility in the film *The Lives of Others* about the East German surveillance system. The film's great impact and merit lay in its humanisation of those charged with actually listening to the data feed; C Dueck, 'The Humanization of the Stasi in 'Das Leben der Anderen'' (2008) *German Studies Review* 599; S Schmeidl, 'The Lives of Others: Living Under East Germany's "Big Brother" or the Quest for Good Men (Das Leben der Anderen) (review)' (2009) *HRQ* 557.

https://doi.org/10.1017/9781009207898.035 Published online by Cambridge University Press

data feed.[139] As this limiting factor is increasingly being removed by computing power and algorithms, real-time surveillance at hitherto impractical levels becomes possible.[140]

Whether the raw data is battlefield reconnaissance, satellite surveillance, signal intelligence, or similar sensor data, the functional challenge, regulatory difficulty, and corresponding strategic opportunity are the same: mere observation is relatively inconsequential – from both a regulatory and operational point of view – unless the information is recorded, classified, interpreted, and thereby made 'useful'.[141] This reflects a basic insight made already some forty years ago by *Herbert Simon*:

> in an information-rich world, the wealth of information means a dearth of something else: a scarcity of whatever it is that information consumes. What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention and a need to allocate that attention efficiently among the overabundance of information sources that might consume it.[142]

In systems design, whether military or civilian, the main design problem is often seen as acquiring and presenting more information, following the traditional mental model that information scarcity is the chief constraint. As *Simon* and others correctly pointed out, however, these design parameters fundamentally mistake the underlying transformation brought about by technological change that is the ever-decreasing cost of collecting and transmitting data leading to the potential for 'information overload'. In other words, the real limiting factor was attention, defined as 'focused mental engagement on a particular item of information. Items come into our awareness, we attend to a particular item, and then we decide whether to act.'[143]

The true distinguishing, competitive ability is, therefore, to design systems that filter out irrelevant or unimportant information and to identify among a vast amount of data those patterns likely to require action. AI is able to automatise this difficult, taxing, and time-consuming process, by spotting patterns of activity in raw data and bringing it to the attention of humans. The key to understanding the transformation wielded by AI, especially machine learning, is the revolutionary reversal of the role of information. For most of human history, information was a scarce resource, which had to be obtained and transmitted at great material and human cost. Technological advances during the latter half of the twentieth century reversed that historic trajectory, making information suddenly over-abundant. Today, the limiting factor is no longer the availability of information as such, but our ability to make sense of its sheer amount. The ability to use computing power to sift through that sudden information abundance thus becomes a chief competitive ability, in business just as on the battlefield: 'Data mining is correctly defined as the nontrivial process of identifying valid, novel, potentially useful and ultimately understandable patterns in data.'[144] The key to performance, whether military or economic, is to derive

[139] 'The Intelligence Agencies of the United States each day collect more raw intelligence data than their entire workforce could effectively analyze in their combined lifetimes.' Allen and Chan, 'Artificial Intelligence and National Security' (n 134) 27, referring to P Domingos, *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World* (2015) 19.

[140] This early realisation was made by Joseph Weizenbaum, the creator of ELIZA, one of the earliest natural language processing softwares. It ran on ordinary personal computers and, despite its simplicity, yielded important insights about computers themselves as social objects. The insight about surveillance was expressed in J Weizenbaum, *Computer Power and Human Reason: From Calculation to Judgment* (1976) 272.

[141] R Calo, 'Peeping HALs: Making Sense of Artificial Intelligence and Privacy' (2010) *European Journal of Legal Studies* 168, 171–174.

[142] HA Simon, *Designing Organizations for an Information-Rich World* (1971) 40–41.

[143] T Davenport and J Beck, *The Attention Economy: Understanding the New Currency of Business* (2001) 20.

[144] Zarsky, 'Mine Your Own Business!' (n 11) 4, 6.

https://doi.org/10.1017/9781009207898.035 Published online by Cambridge University Press

knowledge from data, that is the ability to search for answers in complex and dynamic environments, to spot patterns of sensitive activity among often unrelated, seemingly innocuous information and to bring it to the attention of human decision-makers or initiate automated responses. Drastic advances in AI, made possible by the triple collapse in the price of sensor data collection, data storage, and processing power,[145] finally seem to offer a solution to the problem of information over-abundance by substituting machine attention for increasingly scarce human mental energy.

These long-gestating technological capabilities have suddenly aligned to bring about the maturation of AI. As we saw with respect to unmanned vehicles, one of their key structural advantages consists in their ability to deliver large amounts of sensor data, just like signal intelligence. Traditionally, one of the key constraints consisted in the highly skilled, thus rare and expensive, manpower necessary to make sense of that data: interpreting photographic intelligence, listening in on air control communications in foreign languages, etc.[146] Most of these tasks can already successfully be carried out by narrow AI, offering three game-changing advantages: first, the complete removal of manpower constraint in classifying and interpreting data, detecting patterns and predicting outcomes; second, machine intelligence is quicker than humans, it doesn't tire, it isn't biased,[147] but perhaps most importantly, it can detect patterns humans wouldn't be able to see; and third, AI permits disparate data to be fused, permitting otherwise invisible security-relevant connections to be identified.[148]

VII. FOREIGN RELATIONS

Perhaps more important than the ability to lift the 'fog of war' through better reconnaissance might be the transformation of the role of information and trust in the conduct of foreign relations. Again, this aspect of AI overlaps but is distinct from the Internet. To highlight the enormity of the challenges posed by AI, it might be useful to recall the early years of the Internet. The first time I surfed the web was in the autumn of 1995. Email was known to exist but it was not used by anyone I knew; my own first email was only sent two years later in graduate school. That autumn, I had to call and book a time-slot at the central library of the University of London, the websites I managed to find were crude, took a god-awful time to load and one had to know their addresses or look them up in a physical, printed book.[149]

My conclusion after that initial experience seemed clear: this thing would not catch on. I did not use it again for several years. After all, who would want to read a newspaper on a computer, waiting forever and scrambling through terrible layout? In a now-hilarious appearance on an American late-night show that year, the Microsoft founder *Bill Gates* responded to the host's

[145] Allen and Chan, 'Artificial Intelligence and National Security' (n 135) 14.

[146] During my graduate training at the Kennedy School of Government's specialisation in international security, my tutorial group consisted largely of seconded military officers, many of whom had been trained to do precisely these very difficult, very taxing, and fairly boring intelligence tasks. Especially the need to do this in difficult foreign languages was a very serious limiting factor. The promise of AI and especially machine learning in voice recognition etc. here is apparent.

[147] The issue of bias in the underlying algorithms is itself a field of intense scrutiny, see *inter alia* OA Osoba and W Welser IV, *An Intelligence in Our Image: The Risks of Bias and Errors in Artificial Intelligence* (2017).

[148] The ability of disparate, seemingly innocuous information to reveal striking and strikingly-accurate predictions has been described in a seminal newspaper article about early commercial algorithmic prediction, the principles of which have direct national security implications, see C Duhigg, 'How Companies Learn Your Secrets' *New York Times* (16 February 2012) www.nytimes.com/2012/02/19/magazine/shopping-habits.html.

[149] E Smith, 'The Internet on Dead Trees' (*Tedium*, 29 June 2017) https://tedium.co/2017/06/29/90s-internet-books-history/ (hereafter Smith, 'The Internet on Dead Trees').

https://doi.org/10.1017/9781009207898.035 Published online by Cambridge University Press

thinly-disguised dismissal by giving a fairly enduring definition of that 'internet thing': 'Well, it's becoming a place where people are publishing information. . . . It is the big new thing.'[150] Obviously, *Gates* was more clairvoyant than me. Indeed, the Internet would be the new big thing, but he understood that it would take some time until normal people like me could see its value.[151]

Even after search-engines made the increasingly graphical web far more user-friendly, by 2000 the internet was still not mainstream and some journalists wondered whether it was 'just a passing fad'.[152] Like many new cultural phenomena driven by technological innovation, those 'in the know' enjoyed their avant-garde status, as the editor of one of the early magazines serving this new demographic stressed: 'Internet Underground was this celebration of this relatively lawless, boundless network of ideas we call the Internet. It assumed two things about its audience: 1) You were a fan [and] 2) you knew how to use it. Otherwise, the magazine wouldn't have made much sense to you.'[153] The removal of physical, temporal, and pecuniary barriers to the sharing of information indeed created a 'network of ideas', opening new vistas to collective action, new interpretations of established civil liberties, and new conceptions of geography.[154] Early generations of technophiles 'in the know' conjured this non-corporeal geography as a utopia of unfettered information-sharing, non-hierarchical self-regulation, and self-realisation through knowledge. Then-prevailing conceptions of 'cyberspace' were characterised by scepticism of both government power and commercial interests, often espousing anarchist or libertarian attitudes towards community, seeing information as a commodity for self-realisation, not profit.[155]

Early utopians stressed the opportunities created by this new, non-hierarchical 'network of ideas', which many perceived to be some kind of 'samizdat on steroids', subversive to authoritarian power and its attempts to control truth:[156] 'The design of the original Internet was biased in

[150] 'What Is Internet? Explained by Bill Gates 1995, David Letterman Show' (17 November 2019) https://youtu.be/gipL_CEw-fk, emphasis added.

[151] For the purposes of this chapter, we can ignore that he himself turned out to have misjudged how much ordinary people would see value in that Internet thing.

[152] J Chapman, 'Internet "May Be Just a Passing Fad as Millions Give Up On It"' (5 December 2000) *Daily Mail*.

[153] Rob Bernstein quoted in Smith, 'The Internet on Dead Trees' (n 149).

[154] The work of the *Electronic Frontier Foundation* illustrated the spatial metaphor and combines all three aspects that is the perceived need to defend old and necessary new rights through joint political advocacy on the frontier between traditional physical political communities and the non-corporeal space created through electronic communication, https://www.eff.org/de.

[155] S Binkley, 'The Seers of Menlo Park: The Discourse of Heroic Consumption in the 'Whole Earth Catalog'' (2003) *Journal of Consumer Culture* 283; L Dembart, '"Whole Earth Catalog" Recycled as "*Epilog*"' *New York Times* (8 November 1974) https://www.nytimes.com/1974/11/08/archives/-whole-earth-cataog-recycled-as-epilog-new-group-to-serve.html.

[156] Samizdat describes the analog distribution of unauthorised, critical literature throughout the former Communist countries using mimeographs, photocopiers, often simply re-typed carbon-copies or audio-cassettes for music or poetry readings. The effect of such underground criticism on the stability and legitimacy of the Soviet system has been devastating. Islamists used similar methods during the Iranian revolution. The advent of hard-to-monitor electronic communication portended highly destabilising times for local autocrats, but these hopes did not materialise. On the former aspect, see T Glanc, *Samizdat Past & Present* (2019); L Aron, 'Samizdat in the 21st Century' (2009) *Foreign Pol'y* 131; on the role of audio-cassettes and radio in the Iranian revolution, see BBC Persian Service, 'The History of the Revolution [انقلاب داستان]' (n.d.), www.bbc.com/persian/revolution; E Abrahamian, 'The Crowd in the Iranian Revolution' (2009) *Radical History Review* 13–38; on the role of the Internet in post-Communist politics, see S Kulikova and DD Perlmutter, 'Blogging Down the Dictator? The Kyrgyz Revolution and Samizdat Websites' (2007) *International Communication Gazette* 29–50; L Tsui, 'The Panopticon as the Antithesis of a Space of Freedom: Control and Regulation of the Internet in China' (2003) *China Information* 65; on the political space created by electronic communication generally, see JM Balkin, 'Digital Speech and Democratic Culture: A Theory of Freedom of Expression for the Information Society' (2004) *NYU Law Review* 1; O Tkacheva and others, *Internet Freedom and Political Space* (2013); D Joyce, 'Internet Freedom and Human Rights' (2015) 26 *EJIL* 493.

https://doi.org/10.1017/9781009207898.035 Published online by Cambridge University Press

favor of decentralization of power and freedom to act. As a result, we benefited from an explosion of decentralized entrepreneurial activity and expressive individual work, as well as extensive participatory activity. But the design characteristics that underwrote these gains also supported cybercrime, spam, and malice.'[157] Civilian internet pioneers extrapolated from these core characteristics of decentralisation and unsupervised individual agency a libertarian utopia in the true meaning of the word, a non-place or 'virtual reality' consisting of and existing entirely within a 'network of ideas'. Here, humans could express themselves freely, assume new identities and interests. Unfettered by traditional territorial regimes, new norms and social mores would govern their activities towards personal growth and non-hierarchical self-organisation. Early mainstream descriptions of the Internet compared the novelty to foreign travel, highlighting emotional, cultural, and linguistic barriers to understanding:

> The Internet is the virtual equivalent of New York and Paris. It is a wondrous place full of great art and artists, stimulating coffee houses and salons, towers of commerce, screams and whispers, romantic hideaways, dangerous alleys, great libraries, chaotic traffic, rioting students and a population that is rarely characterized as warm and friendly. . . . First-time visitors may discover that finding the way around is an ordeal, especially if they do not speak the language.[158]

As the Internet became mainstream and eventually ubiquitous, many did, in fact, learn to 'speak its language', however imperfectly.[159] The advent of AI can be expected to bring changes of similar magnitude, requiring individuals and our governing institutions to again 'learn its language'. AI is altering established notions of verification and perceptions of truth. The ability to obtain actionable intelligence despite formidable cultural and organisational obstacles,[160] is accompanied by the ability to automatically generate realistic photographs, video, and text, enabling information warfare of hitherto unprecedented scale, sophistication, and deniability.[161] Interference in the electoral and other domestic processes of competing nations are not new, but the advent of increasingly sophisticated AI is permitting 'social engineering' in novel ways.

First, it has become possible to attack large numbers of individuals with highly tailored misinformation through automated 'chatbots' and similar approaches. Secondly, the quality of 'deep fakes' generated by sophisticated AI are increasingly able to deceive even aware and skilled individuals and professional gatekeepers.[162] Thirdly, the well-known 'Eliza-effect' of human beings endowing inanimate objects like computer interfaces with human emotions, that is imbuing machines with 'social' characteristics permits the deployment of apparently responsive agents at scale, offering unprecedented opportunities and corresponding risks not only for

[157] Benkler, 'Degrees of Freedom' (n 42) 18, 19.

[158] PH Lewis, 'Personal Computers: First-Time Tourists Need a Pocket Guide to Downtown Internet' *New York Times* (5 April 1994) www.nytimes.com/1994/04/05/science/personal-computers-first-time-tourists-need-a-pocket-guide-to-downtown-internet.html; Lewis' reference to Paris and New York was probably not a coincidence, given the somewhat fearsome reputation the inhabitants of these two cities have earned, because he goes on to warn: 'Newcomers to the Internet are warned repeatedly to avoid annoying the general population with their questions.'

[159] Y Benkler, R Faris, and H Roberts, *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics* (2018) (hereafter Benkler, Faris, and Roberts, *Network Propaganda*).

[160] B Hubbard, F Farnaz, and R Bergman, 'Iran Rattled as Israel Repeatedly Strikes Key Targets' *New York Times* (20 April 2021) www.nytimes.com/2021/04/20/world/middleeast/iran-israeli-attacks.html.

[161] Allen and Chan, 'Artificial Intelligence and National Security' (n 135) 29–34.

[162] KM Sayler and LA Harris, 'Deep Fakes and National Security' (26 August 2020) Congressional Research Service https://apps.dtic.mil/sti/pdfs/AD1117081.pdf; DK Citron and R Chesney, 'Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security' (2019) *California Law Review* 1753.

https://doi.org/10.1017/9781009207898.035 Published online by Cambridge University Press

'phishing' and 'honey trap' operations,[163] but especially to circumvent an enemy government by directly targeting its population.[164]

A distinct problem fueled by similar technological advances is the ability to impersonate representatives of governments, thereby undermining trust and creating cover for competing narratives to develop.[165] Just as with any other technology, it is reasonable to expect that eventually corresponding technological advances will make it possible to detect and defuse artificially created fraudulent information.[166] It is furthermore reasonable to expect that social systems will likewise adapt and create more sophisticated consumers of such information better able to resist misinformation. Such measures had been devised during wars and ideological conflicts in the past and it is therefore correct to state that 'deep fakes don't create new problems so much as make existing problems worse'.[167] *Jessica Silbey* and *Woodrow Hartzog* are, of course, correct that the cure to the weaponisation of misinformation lies in strengthening and creating institution tasked with 'gatekeeping' and validation:

> We need to find a vaccine to the deep fake, and that will start with understanding that authentication is a social process sustained by resilient and inclusive social institutions. … it should be our choice and mandate to establish standards and institutions that are resilient to the con. Transforming our education, journalism, and elections to focus on building these standards subject to collective norms of accuracy, dignity, and democracy will be a critical first step to understanding the upside of deep fakes.[168]

The manner in which this is to be achieved goes beyond the scope of this chapter, but it is important to keep in mind that both accurate information itself, as well as misinformation have long been part of violent and ideological conflict.[169] Their transformation by the advent of AI must, therefore, be taken into account for a holistic assessment of its impact on national security and its legal regulation. This is particularly pertinent due to the rise of legal argumentation not only as a corollary of armed conflict but as its, often asymmetric, substitute in the form of 'lawfare',[170] as well as the evident importance of legal standards for such societal 'inocculation' to be successful.[171]

[163] Forsvarsministeriet, 'Center for Cybersikkerhed' (18 September 2020) https://www.fmn.dk/da/arbejdsomraader/cybersikkerhed/center-for-cybersikkerhed/.

[164] On such 'information attacks,' see generally MJ Blitz, 'Lies, Line Drawing, and (Deep) Fake News' (2018) 72 *Okla L Rev* 59; Benkler, Faris, and Roberts, Network Propaganda (n 159).

[165] S Agarwal and others, 'Protecting World Leaders against Deep Fakes' (2019) *IEEE Xplore* 38.

[166] For an account of the technology involved, see for instance S Agarwal and others, 'Detecting Deep-Fake Videos from Appearance and Behavior' (2020) *IEEE International* 1.

[167] J Silbey and W Hartzog, 'The Upside of Deep Fakes' (2019) 78 *Maryland Law Review* 960, 960.

[168] Ibid, 966.

[169] R Darnton, 'The True History of Fake News' *The New York Review* (13 February 2017) www.nybooks.com/daily/2017/02/13/the-true-history-of-fake-news/.

[170] The term has been suggested by General Charles Dunlap who offered the following definition: 'the strategy of using – or misusing – law as a substitute for traditional military means to achieve a warfighting objective.' CJ Dunlap, 'Lawfare Today: A Perspective' (2008) *Yale J Int'l L* 146. 146. See also D Stephens, 'The Age of Lawfare', in RA Pedrozo and DP Wollschlaeger (eds), *International Law and the Changing Character of War* (2011); CJ Dunlap, 'Lawfare Today … and Tomorrow', in RA Pedrozo and DP Wollschlaeger (eds), *International Law and the Changing Character of War* (2011).

[171] See *inter alia* Chapter 13 "What Can Men Do against Such Reckless Hate?" in Benkler, Faris, and Roberts, *Network Propaganda* (n 159) 351–380.

https://doi.org/10.1017/9781009207898.035 Published online by Cambridge University Press

VIII. ECONOMICS

National security is affected by economic competitiveness, which supplies the fiscal and material needs of military defence. The impact of the ongoing revolution in AI on existing labour markets and productive patterns is likely to be transformational.[172] The current debate is reminiscent of earlier debates about the advent of robotics and automation in production. Where that earlier debate focused on the impact on the bargaining power and medium-term earning potential of blue-collar workers, AI is also threatening white-collar workers, who hitherto seemed relatively secure from cross-border wage arbitrage as well as automation.[173] In a competitive arena, whether capitalism for individual firms or anarchy for nations, the spread of innovation is not optional but a logical consequence of the 'socialising effect' of any competitive system:[174] 'Machine learning is a cool new technology, but that's not why businesses embrace it. They embrace it because they have no choice.'[175]

This embrace of AI has at least three important national security implications, with corresponding regulatory challenges and opportunities. First, dislocations resulting from the substitution of machines for human labour has destabilising effects for social cohesion and political stability, both domestic and international.[176] These dislocations have to be managed, including through the use of proactive regulation meant to further positive effects while buffering negative consequences.[177] The implications of mass unemployment resulting from this new wave of automation is potentially different from earlier cycles of technological disruption because it could lead to permanent unemployability of large sectors of the population, rendering them uncompetitive at any price. This could spell a form of automation-induced 'resource curse' affecting technologically advanced economies,[178] suddenly suffering from the socio-economic-regulatory failings historically associated with underdeveloped extractive economies.[179]

Second, the mastery of AI has been identified by all major economic powers as central to maintaining their relative competitive posture.[180] Consequently, the protection of intellectual property, the creation of a conducive regulatory, scientific, and investment climate to nurture the sector has itself increasingly become a key area of competition between nations and trading blocs.[181]

[172] European Commission, 'Proposal for a Regulation Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act)' (*European Commission*, 26 April 2021) https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence-artificial-intelligence.

[173] K Roose, 'The Robots Are Coming for Phil in Accounting' *New York Times* (6 March 2021) www.nytimes.com/2021/03/06/business/the-robots-are-coming-for-phil-in-accounting.html.

[174] KN Waltz, *Theory of International Politics* (1979) 129.

[175] P Domingos, *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World* (2015) 13.

[176] Allen and Chan, 'Artificial Intelligence and National Security' (n 134) 36–39.

[177] Denmark and the other Scandinavian economies have a long history of seeking productivity gains in both the public and private sector as a way to keep their costly welfare systems fiscally sustainable and labour markets globally competitive. See *inter alia* Forsvarsminisiteriet, 'National strategi for cyber- og informationssikkerhed. Øget professionalisering og mere viden' (December 2014); C Greve and N Ejersbo, *Moderniseringen af den offentlige sektor* (3rd ed. 2014); J Hoff, *Danmark som Informationssamfund. Muligheder og Barrierer for Politik og Demokrati* (2004); PA Hall, 'Danish Capitalism in Comparative Perspective', in JL Campbell, JA Hall, and OK Pedersen (eds), *National Identity and the Varieties of Capitalism: The Danish Experience* (2006).

[178] Allen and Chan, 'Artificial Intelligence and National Security' (n 135) 37.

[179] See *inter alia* G Luciani, 'Allocation v Production States: A Theoretical Framework' in G Luciani and B Hazem (eds), *The Rentier State* (2015).

[180] Mozur and Myers, 'Xi's Gambit' (n 5); R Doshi and others, 'China as a "Cyber Great Power" – Beijing's Two Voices in Telecommunications' (*Brookings*, April 2021) www.brookings.edu/wp-content/uploads/2021/04/FP_20210405_china_cyber_power.pdf.

[181] Bird and others, 'The Ethics of Artificial Intelligence' (n 59).

https://doi.org/10.1017/9781009207898.035 Published online by Cambridge University Press

Third, given the large overlap between civilian and military sectors, capabilities in AI developed in one are likely to affect the nation's position in the other.[182] Given inherent technological characteristics, especially scalability and the drastic reduction of marginal costs, and the highly disruptive effect AI can have on traditional military capabilities, the technology has the potential to drastically affect the relative military standing of nations quite independent of conventional measures such as size, population, hardware, etc.: 'Small countries that develop a significant edge in AI technology will punch far above their weight.'[183]

IX. CONCLUSION

Like many previous innovations, the transformational potential of AI has long been 'hyped' by members of the epistemic communities directly involved in its technical development. There is a tendency among such early pioneers to overstate potential, minimise risk, and alienate those not 'in the know' by elitist attitudes, incomprehensible jargon, and unrealistic postulations. As the comparison with cyberspace has shown, it is difficult to predict with accuracy what the likely impact of AI will be. Whatever its concrete form, AI is almost certain to transform many aspects of our lives, including national security.

This transformation will affect existing relative balances of power and modes of fighting and thereby call into question the existing normative *acquis*, especially regarding international humanitarian law. Given the enormous potential benefits and the highly dynamic current stage of technological innovation and intense national competition, the prospects for international regulation, let alone outright bans are slim. This might appear to be more consequential than it is, because much of the transformation will occur in operational, tactical, and strategic areas that can be subsumed under an existing normative framework that is sufficiently adaptable and broadly adequate.

The risk of existential danger by the emergence of super-intelligence is real but perhaps overdrawn. It should not detract from the laborious task of applying existing international and constitutional principles to the concrete regulation of more mundane narrow AI in the national security field.

[182] Allen and Chan, 'Artificial Intelligence and National Security' (n 135) 35–41.

[183] Ibid, 3 and 58–59.

https://doi.org/10.1017/9781009207898.035 Published online by Cambridge University Press