Cambridge Law Review (2023) Vol 8, Issue 1, 1–24

The Guilty (Silicon) Mind: Blameworthiness and Liability in Human-Machine Teaming

BRENDAN WALKER-MUNRO* AND ZENA ASSAAD**

ABSTRACT

As science pushes the boundaries of the development of artificial intelligence (AI), the progress has caused scholars and policymakers alike to question the legality of utilising AI in various human endeavours. Debate has raged in international scholarship about the legitimacy of applying AI to weapon systems to form lethal autonomous weapon systems (LAWS). Yet the legality of applying or utilising AI is questionable even when AI is applied to a non-weaponised autonomous system: how does one hold a machine accountable for a crime? What about a tort? Can an artificial agent understand the moral and ethical content of its instructions? These are thorny questions, and in many cases, these questions have been answered in the negative, as artificial entities lack any contingent moral agency. What then occurs if the AI is not alone, but linked with or overseen by a human being, with their own moral and ethical understandings and obligations? Who is responsible for any malfeasance that may be committed? Does the human bear the legal risks of unethical or immoral decisions of an AI? These are some of the questions with which this manuscript seeks to engage.

Keywords: human-machine teaming, liability, criminal law, civil law, military

I. INTRODUCTION

Automation has been a key result of mankind's technological development over the last two centuries. Rather than a reliance on manual labour, we have developed mechanised tools which replace our efforts with streamlined and optimised acts.

* Senior Research Fellow, Law and Future of War Research Group, The University of Queensland; JD, PhD.

** Senior Research Fellow, School of Engineering, Australian National University; BAeroEng, PhD.

The research for this paper received funding from the Australian Government through Trusted Autonomous Systems, a Defence Cooperative Research Centre funded through the Next Generation Technologies Fund.

Even in the most sensitive and value-driven theatre of human endeavour—that of decision making—the march of progress has not slowed, such that we now have computer programs capable of making decisions on everything from restaurant orders and hotel bookings to delivery of healthcare and social welfare programs.[1]

Yet that automation is not without its controversy. A discussion has raged in the international community regarding the legitimacy of merging the 'hard' processing capabilities of a computer with the 'soft' processing abilities of a human.[2] Whilst the reality of such a concept might have been previously restricted to the pages of popular fiction,[3] this is no longer the case. Human-machine interfaces—where a system operates to modulate a human's sensory connection with a machine—are already being used in contemporary applications such as piloting drones and other autonomous and semi-autonomous platforms.[4] Scholars are now examining the next step of this inclusion of machines in the human realm of decision-making with an increased research interest in 'human-machine teaming' (HMT).[5] For the context of this paper, HMT is defined as a bi-directional combination of human and machine capabilities which work together with a dynamic directedness towards an aligned goal.[6]

Conceptually, this definition requires an HMT to include the processing capabilities of both a machine and human component. Each component must be able to send and receive messages from the other that enable them to actively (not passively) aim towards achieving the same goal. The concept of 'dynamic directedness' thus imports two requirements to HMT: one, an element of back-and-forth communication between machine and human component; and two, the need for that back-and-forth communication to be directed towards a similar (but not identical) objective purpose.

Yet despite the research interest in HMT, the literature still lacks a cohesive framework which adequately reflects the legal responsibility for an HMT. Imagine a car assembly, where human workers and robotic workers complete their tasks

[1] Igor Bikeev and others, 'Criminological Risks and Legal Aspects of Artificial Intelligence Implementation' (Proceedings of the International Conference on Artificial Intelligence, Information Processing and Cloud Computing, Sanya, December 2019) <https://www.researchgate.net/publication/337883901_Criminological_risks_and_legal_aspects_of_artificial_intelligence_implementation> accessed 22 March 2023.

[2] Linda Skitka, Kathleen L Mosier, and Mark Burdick, 'Does Automation Bias Decision-Making?' (1999) 51 International Journal of Human-Computer Studies 991; Ericka Rovira, Kathleen McGarry, and Raja Parasuraman, 'Effects of Imperfect Automation on Decision Making in a Simulated Command and Control Task' (2007) 49 Human Factors 76; Gustav Markkula and others, 'Models of Human Decision-Making as Tools for Estimating and Optimizing Impacts of Vehicle Automation' (2018) 2672(37) Transportation Research Record 153; Monika Zalnieriute, Lyria Bennett Moses, and George Williams, 'The Rule of Law and Automation of Government Decision-Making' (2019) 82 Modern Law Review 425.

[3] Alan Turing, 'Computing Machinery and Intelligence' (1950) 59 Mind 433.

[4] Jennifer Riley and others, 'Situation Awareness in Human-Robot Interaction: Challenges and User Interface Requirements' in Michael Barnes and Florian Jentsch (eds), *Human-Robot Interactions in Future Military Operations* (CRC Press 2017) 180.

[5] The terms 'human-machine team' and 'human-machine teaming' are functionally the same for present purposes and can be used interchangeably throughout this paper.

[6] Adapted from Memunat A Ibrahim, Zena Assaad, and Elizabeth Williams, 'Trust and Communication in Human-Machine Teaming' (2022) 10 Frontiers in Physics 1.

side-by-side, assembling the components of a vehicle as part of a smoothly operating team. Both humans and machines, however, are also given a particular values framework imposed by the factory owner: vehicles must be completed to a certain standard and within a certain time. What happens when the machines realise that their human counterparts are the ones that are slowing down the process, making mistakes, and costing time and resources? A human worker might seek to disobey the restrictions imposed on them by the factory owner, go on strike, or maybe just go at their own pace and risk dismissal. Robots, programmed by humans, might be programmed to behave by them. Alternatively, they may lack flexibility in the programming and kill their co-workers inadvertently in the pursuit of improvement.

Although this may sound like the plot to a Hollywood blockbuster, some semblance of these facts can be found in reality. Kenji Urada is widely recognised as the first human to die from an injury caused by a robot. In 1981, Urada was performing maintenance on an automated hydraulic arm which, despite written safety protocol, was still powered on. The system misinterpreted Urada's actions as an attempt to damage the arm, which reacted by knocking Urada into an adjacent machine. Urada was crushed and died instantly.[7] A similarly horrifying (though less serious) incident occurred in 2022 when a 7-year-old chess player had his finger broken by a robotic opponent.[8] In both cases, blame was laid squarely on Urada and the 7-year-old (that is, the human) for violating safety protocol, and otherwise no charges were laid.

Nowhere should this development be more concerning than in the field of military and armed forces given the rapid development of research into the 'deployment of AI-infused systems (e.g. drone swarming, command and control decision-making support systems and a broader range of autonomous weapon systems)'.[9] Whilst the idea of an HMT presents obvious benefits to military operations, the controversy arises by inflaming existing risks or generating new challenges of relevance to military commanders and systems designers. There should be a conceptual question about the attribution of responsibility for unlawful actions committed within military HMT operations: do the actions give rise to civil liability (where the remedy is usually compensation or some remedial order of the court) or criminal liability (where the remedy is usually imprisonment for natural persons, both as a form of punishment and to protect innocent members of society)?

[7] Yueh-Hsuan Weng, Chien-Hsun Chen, and Chuen-Tsai Sun, 'Toward the Human-Robot Co-Existence Society: On Safety Intelligence for Next Generation Robots' (2009) 1 International Journal of Social Robotics 267, 273.

[8] Jon Henley, 'Chess Robot Grabs and Breaks Finger of Seven-Year-Old Opponent' *The Guardian* (London, 24 July 2022) <www.theguardian.com/sport/2022/jul/24/chess-robot-grabs-and-breaks-finger-of-seven-year-old-opponent-moscow> accessed 16 September 2022.

[9] James Johnson, 'The AI-Cyber Nexus: Implications for Military Escalation, Deterrence and Strategic Stability' (2019) 4 Journal of Cyber Policy 442.

We, therefore, set out in this article to advance the proposition that, for military HMT operations, the specifics of the dynamic interactions between the human and machine elements will dictate how liability will be attributed. We intend to approach the problem in the following way. Section II will involve an exploration of the issues of defining HMT in the contemporary context. It will identify that the bi-directionality of communication between the human and machine elements serves to blur the perceptions and observations of both parts, and may have legal and regulatory ramifications. Section III will then introduce some key terms in the context of both civil and criminal law around the establishment of liability, with reference to the idea of blameworthiness. In Section IV, we explore some of the problems with applying these concepts of liability to an artificial agent (including a partial agent such as would be present in an HMT). Before concluding the paper, in Section V, we introduce and explain a possible mechanism of regulating an HMT which we argue will appropriately respect the concepts of blameworthiness and liability whilst retaining utility to incorporate artificial agents.

This article will also specifically focus on an HMT in a military context. There are three reasons for such a focus. The first is that an HMT is a significant component of the technological research for many western military forces including the US, the UK, the EU and Australia,[10] but also of west-adversarial nations such as China and Russia.[11] Secondly, like their comparative cousins in the form of autonomous weapon systems, the application of AI to military decision making in HMT is already recognised as a challenge to the rules-based order of international and comparative domestic law.[12] And thirdly, the military is often a testbed for emerging technologies, with the armed forces standing as the entity which most commonly responds to the legal and regulatory challenges that arise from their implementation.[13]

[10] UK Ministry of Defence, 'Joint Concept Note 1/18: Human Machine Teaming' (UK Ministry of Defence May 2018) <www.gov.uk/government/publications/human-machine-teaming-jcn-118> accessed 19 October 2022; Chad C Tossell and others, 'Appropriately Representing Military Tasks for Human-Machine Teaming Research' in Constantine Stephanidis, Jessie YC Chen and Gino Fragomeni (eds), *HCI International 2020 — Late Breaking Papers: Virtual and Augmented Reality* (Springer 2020); Alex Neads, David J Galbreath, and Theo Farell, 'From Tools to Teammates: Human Machine Teaming and the Future of Command and Control in the Australian Army' (Australian Army Occasional Paper No 7, 20 September 2021).

[11] US Department of Defense, 'Military and Security Developments Involving the People's Republic of China 2021' (US Department of Defense 2021) <https://media.defense.gov/2021/Nov/03/2002885874/-1/-1/0/2021-CMPR-FINAL.PDF> accessed 17 October 2022.

[12] Aiden Warren and Alek Hillas, 'Lethal Autonomous Weapons Systems: Adapting to the Future Unmanned Warfare and Unaccountable Robots' (2017) 12 Yale Journal of International Affairs 71; Aiden Warren and Alek Hillas, 'Friend or Frenemy? The Role of Trust in Human-Machine Teaming and Lethal Autonomous Weapons Systems' (2020) 31 Small Wars & Insurgencies 822.

[13] See for example how drone regulation has emerged in military contexts: Ferran Giones and Alexander Brem, 'From Toys to Tools: The Co-Evolution of Technological and Entrepreneurial Developments in the Drone Industry' (2017) 60 Business Horizons 875; Matthieu J Guitton, 'Fighting the Locusts: Implementing Military Countermeasures against Drones and Drone Swarms' (2021) 4 Scandinavian Journal of Military Studies 26.

II. Definitional Issues of Human-Machine Teaming

One of the most significant challenges facing the academic and industrial community is the lack of a shared definition of what exactly comprises an HMT. Definitions are vitally important for legal and regulatory purposes, not just as academic or theoretical constructs. The blurring of responsibility between the human and machine elements in an HMT and indeed the very concept of identifying where a human ends and a machine begins can present significant challenges to the legal and regulatory framework for future HMT operations. If a legal principle cannot apply to the emergence of HMT operations, or applies weakly or ambiguously, the danger of an unregulated system is apparent. Even absent the possibility that an HMT (especially military HMT) might be operating without a proper form of legal control or oversight, the absence of a proper regulatory system can diminish public trust in the operations of the armed forces. Worse, such systems could expose those same armed forces to liability themselves.[14]

One such example defines an HMT as 'a purposeful combination of human and cyber-physical elements that collaboratively pursue goals that are unachievable by either individually'.[15] Some broader literature on HMT have similarities in their proposed definitions, with many expressing notions of sharing authority to pursue common goals.[16] Such a definition clearly articulates the connection and bi-directionality between the human (natural) and the machine (artificial), yet articulates these by reference to a frame in which goals *cannot* be achieved by one or the other in isolation. Applying such a definition to the simple act of driving a vehicle highlights the definitional issues—clearly, both humans and machines can operate, steer, and control a vehicle without necessary recourse to the other.[17]

Another definition of HMT might be of more utility: 'the dynamic arrangement of humans and cyber-physical elements into a team structure that capitalizes on the respective strengths of each while circumventing their respective limitations in pursuit of shared goals'.[18] Is it the existence of collaboration then, of movement towards a shared goal, which hallmarks human-machine teaming? Yet again, the difficulty in the detail surfaces when applied to a contextual application. Imagine a drone equipped with missiles, deployed in a foreign state but monitored in its home state by a human operator. Both the drone and the operator have a shared goal—the identification, pursuit, and engagement of the State's legitimate military

[14] Consider for example the application of article 36 of Additional Protocol I of the Geneva Convention: Damian P Copeland, 'Legal Review of New Technology Weapons' in Hitoshi Nasu and Robert McLaughlin (eds), *New Technologies and the Law of Armed Conflict* (Springer 2014).

[15] Azad M Madni and Carla C Madni, 'Architectural Framework for Exploring Adaptive Human-Machine Teaming Options in Simulated Dynamic Environments' (2018) 6 Systems 44, 49.

[16] Joseph B Lyons and others, 'Human-Autonomy Teaming: Definitions, Debates, and Directions' (2021) 12 Frontiers in Psychology 1932.

[17] J Levinson and others, 'Towards Fully Autonomous Driving: Systems and Algorithms' (2011 IEEE Intelligent Vehicles Symposium IV, Baden-Baden, June 2011).

[18] Lyons and others (n 16).

targets—but the nature of the relationship is perhaps better characterised as supervision than 'circumventing their respective limitations'. The drone is obviously performing a function in replacement of the human operator and takes the risk in doing so, but the operator still is the one who carries the risk associated with commencing or prosecuting any attack.

How then do the various world militaries approach this definitional issue? The Australian Army broadly defines HMT as the 'incorporation of autonomous or robotic systems within military teams to achieve tactical outputs that neither machines nor people could deliver independently',[19] whilst the United Kingdom's joint concept note on HMT defines the 'effective integration of humans, artificial intelligence (AI) and robotics into warfighting systems'.[20] The US Department of Defense does not strictly define HMT, instead referring to it more obliquely via terminology buried in the program definitions. Take for example the Next-Generation Nonsurgical Neurotechnology (N3) project, which:

> …aims to develop high-performance, bi-directional brain-machine interfaces for able-bodied service members. Such interfaces would be enabling technology for diverse national security applications such as control of unmanned aerial vehicles and active cyber defense systems or teaming with computer systems to successfully multitask during complex military missions.[21]

What is common about these military definitions is the incorporation of, or integration between, human and machine components to achieve outcomes for the armed forces in combat and peacetime. These similarities in the military context also betray the same difficulties in the execution of HMT, which is to explain why military forces strive to achieve the ideal of HMT. The Australian army's definition contains perhaps the most succinct policy purpose of HMT, that is, to '…achieve tactical outputs that neither machines nor people could deliver independently',[22] a concept directly reflecting the ideal in the literature that HMT ought to circumvent the respective limitations of human and machine.[23]

A similar lack of consistency affects definitions in other research spheres. For example, the report published by the UN Institute for Disarmament Research (UNIDIR) does not explicitly define HMT. Instead, UNIDIR focuses on defining a spectrum of HMT operations from 'coactive design' to 'immersion' in which the

[19] Neads, Galbreath, and Farell (n 10).
[20] UK Ministry of Defence (n 10) 39.
[21] Gopal Sarma, 'Our Research: Next-Generation Nonsurgical Neurotechnology' (*DARPA*, 2022) <https://www.darpa.mil/program/next-generation-nonsurgical-neurotechnology> accessed 1 December 2022.
[22] Neads, Galbreath, and Farell (n 10).
[23] Madni and Madni (n 15); Lyons and others (n 16).

machine and human operate in 'virtual worlds that are simulated [and] dynamic'.[24] NATO approaches to HMT also focus not on the definition of the term, but on the supposed benefits to military decision-making, noting that teaming is the ultimate expression of collaboration, trustworthiness and adaptation between the human and machine components.[25]

A critical thread, however, can be observed across both the military and non-military works seeking to define HMT. Contemporary military capabilities already involve collaboration between humans and machines—whether the machine is a sensor, interface, weapon, or system capable of communicating in a shared language, to achieve or move towards some shared goal. The sharing of these capabilities between humans and machines is necessary for achieving outputs that neither entity could complete independently. The critical thread observed, however, in these definitions (and the one that lies at the core of this article) is the bi-directionality of that communication. A machine may communicate with its human capability by displaying sensor information, or the projected results of a weapon detonation,[26] whilst a human may provide commands to select, track or engage targets presented.[27]

Moreover, the bi-directionality of communication also presents a unique challenge to attributing responsibility to the agent in the team who made the particular decision. Autonomous weapons and military robotics have long been suggested to suffer from a 'responsibility gap'[28], that is, the idea that mankind cedes control to machines when they are invested with the capability to learn and

[24] Ioana Puscas, 'Human-Machine Interfaces in Autonomous Weapon Systems: Considerations for Human Control' (*UNIDIR*, 2022) 15–16 <https://www.unidir.org/publication/human-machine-interfaces-autonomous-weapon-systems> accessed 25 February 2023.

[25] Karel van den Bosch and Adelbert Bronkhorst, 'Human-AI Cooperation to Benefit Military Decision Making' (NATO Report STO-MP-IST-160 2018) 8 <https://karelvandenbosch.nl/documents/2018_Bosch_etal_NATO-IST160_Human-AI_Cooperation_in_Military_Decision_Making.pdf> accessed 25 February 2023.

[26] See for example the Athena AI which can differentiate between objects protected under international humanitarian law and legitimate targets: Jonathan Bradley, 'Athena AI Helps Soldiers on the Battlefield Identify Protected Targets' (*Create Digital*, 26 April 2021) <https://createdigital.org.au/athena-ai-helps-soldiers-identify-protected-targets/> accessed 13 July 2022.

[27] Vasja Badalič, 'Automating the Target Selection Process: Humans, Semiautonomous Weapons Systems, and the Assault on International Humanitarian Law' in Aleš Završnik and Vasja Badalič (eds), *Automating Crime Prevention, Surveillance, and Military Operations* (Springer 2021).

[28] Thomas Hellström, 'On the Moral Responsibility of Military Robots' (2013) 15 Ethics and Information Technology 99; Merel Noorman and Deborah G Johnson, 'Negotiating Autonomy and Responsibility in Military Robots' (2014) 16 Ethics and Information Technology 51; Lambèr Royakkers and Peter Olsthoorn, 'Lethal Military Robots: Who is Responsible When Things Go Wrong?' in Mehdi Khosrow-Pour (ed) *Unmanned Aerial Vehicles: Breakthroughs in Research and Practice* (IGI Global 2019); Bernd W Wirtz, Jan C Weyerer, and Carolin Geyer, 'Artificial Intelligence and the Public Sector—Applications and Challenges' (2019) 42 International Journal of Public Administration 596; Isaac Taylor, 'Who Is Responsible for Killer Robots? Autonomous Weapons, Group Agency, and the Military-Industrial Complex' (2021) 38 Journal of Applied Philosophy 320.

evolve, which Matthias postulated would lead to 'injustice of holding men responsible for actions of machines over which they could not have sufficient control'.[29] Ryan writes that 'military organizations must…examine whether it is desirable to have robots able to kill humans based on automated processes and without a human in the decision cycle'.[30] Nevertheless, the idea of a conjoined or collaborative HMT sidesteps Ryan's understanding of the issue: The existence of a human 'in the loop' of decision-making is no safeguard against failure.

Consider the following scenarios involving hypothetical military HMTs, but which have been based on existing automated or autonomous technologies:

1. The pilot of an attack aircraft, assisted by uncrewed sensor drones,[31] attacks a convoy based on the drones' assessment of those vehicles as being legitimate military targets. Following an investigation, it is revealed that the convoy contained fleeing refugees and the sensor data was incorrectly interpreted by the drones;
2. The captain of a Naval destroyer is linked to the automated defences of their ship. The radar detects an aircraft approaching and assesses its behaviour as benign; the captain, however, believes the aircraft is adopting an attack profile and opens fire. The aircraft was in fact an allied fighter in an adjacent battlegroup;[32] and
3. A platoon of soldiers is conducting a patrol in a foreign country, assisted by an armed robotic companion that is teamed with one of the platoon soldiers.[33] Unbeknownst to the platoon, the software underpinning the robot has been hacked by enemy forces and suddenly presents false threat warnings. The teamed soldier opens fire, killing one of his platoon members.

As shown above, each of these scenarios highlights a specific concern with the attribution of responsibility for a military HMT. In the first scenario, there was no malicious or adverse action by any person, merely the occurrence of what might

[29] Andreas Matthias, 'The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata' (2004) 6 Ethics and Information Technology 175, 183.

[30] Mick Ryan, 'Human-Machine Teaming for Future Ground Forces' (Center for Strategic and Budgetary Assessments 2018) 36 <https://csbaonline.org/uploads/documents/Human_Machine_Teaming_FinalFormat.pdf> accessed 25 February 2023.

[31] Based on the Loyal Wingman project developed by the Royal Australian Air Force: RAAF, 'Loyal Wingman' (*Royal Australian Airforce*, 2020) <https://www.airforce.gov.au/our-mission/loyal-wingman> accessed 9 July 2022.

[32] Adapted from the downing of a US intruder aircraft by a Japanese destroyer in 1996: Thomas Newdick, 'The Last Time a Japanese Warship Shot Down a U.S. Navy Plane was Actually not so Long Ago' (*The Drive*, 5 June 2021) <https://www.thedrive.com/the-war-zone/40937/the-last-time-a-japanese-warship-shot-down-a-u-s-navy-plane-was-actually-not-so-long-ago> accessed 14 January 2022.

[33] Based in part on the arming of a Boston Dynamics 'dog' robot: Joshua Rhett Miller, 'Robot Dog Equipped with Submachine Gun is 'Dystopian' Nightmare Fodder' *New York Post* (21 July 2022) <https://nypost.com/2022/07/21/robot-dog-with-submachine-gun-is-dystopian-nightmare-fodder/> accessed 17 August 2022.

be called 'human error'—yet it was an error that resulted in the preventable deaths of civilians.[34] In the second scenario, the naval captain imparted his human bias (contradicting the automated assessment of the aircraft's behaviour) into the decision-making cycle, arguably undermining the purpose of an HMT in the first place. In the last scenario, the addition of a cyber-physical element into human warfighting opens new avenues for misdirection and attack by enemy forces.

In all three scenarios, the issue of bi-directionality is front and centre in the difficulty of attributing responsibility. The pilot in the first scenario may well have been able to avert disaster had they not relied on the drones' sensor information and been able to visually observe the target, something required by pilots in previous conflicts.[35] In the second scenario, the communication with the human is what hampered the machine in (correctly) identifying that the aircraft was not a threat. Inversely, the third scenario demonstrates that the bi-directionality of communication in HMT introduces a vulnerability which can be exploited by adversarial forces. How might this bi-directionality then affect the legal treatment of HMT under civil or criminal law?

III. Liability in Civil and Criminal Law

At this point, it is apposite to examine the concept of liability in both civil and criminal law, so that the requisite characteristics of that concept can be identified which are vulnerable to displacement by HMT. This displacement is likely to occur because of the bi-directionality of communication between the human and machine elements of an HMT, and subsequent reliance on that communication as a basis for taking action: a decision made by a machine or human, where one influences the other, has the potential to affect resulting liability.

Traditional western legal systems attribute liability on a basis of 'the individual human person as the central unit of action and the appropriate object of blame'.[36] This idea of liability as blameworthiness, both factual and moral, informs how the civil and criminal look to achieve co-regulatory purposes by enforcing breaches of duties in ways that are generally complementary.[37] The criminal law attributes liability at a higher standard and burden of proof, acting in a more

[34] For example, consider the human error that led to the US strike on a Medecins Sans Frontiers hospital in Kunduz, Afghanistan in 2015: John F Campbell, 'Investigation Report of the Airstrike on the Medecins Sans Frontieres / Doctors Without Borders Trauma Center in Kunduz, Afghanistan on 3 October 2015' (United States Forces Command 21 November 2015) <http://fpp.cc/wp-content/uploads/01.-AR-15-6-Inv-Rpt-Doctors-Without-Borders-3-Oct-15_CLEAR.pdf> accessed 22 September 2022.

[35] As an example, the British Royal Air Force bombers colloquially known as 'Dambusters' in the Second World War were not permitted to drop their bombs until the dams were in sight: Paul Brickhill, *The Dam Busters* (Macmillan 2017).

[36] Neha Jain, 'Autonomous Weapon Systems: New Frameworks for Individual Responsibility' in Nehal Bhuta and others (eds), *Autonomous Weapons Systems: Law, Ethics, Policy* (Cambridge University Press 2016) 303.

[37] Kenneth W Simons, 'The Crime/Tort Distinction: Legal Doctrine and Normative Perspectives' (2008) 17 Widener Law Journal 719.

'agent-focused' manner than the civil law, where mere negligence or breach of duty will suffice.[38] Though criminal liability might involve a similar assessment of compliance or non-compliance as civil law,[39] the criminal law is also a tool of social control designed to both punish those who have committed the offences as well as to virtue signal potential future offenders that such behaviour is anathematic to good policy.[40] The stigma of criminal convictions and incarceration also achieves a broader social effect than the necessity of remediating or repairing harm in the context of civil litigation.[41]

In this way, crimes outlaw particular activity and make it impermissible under every circumstance, whilst civil law prevents the breaches of rights and provides reparation of breaches. In economic terms, 'criminal law exclusively imposes *sanctions*, while [civil] law prices an *activity*'.[42] The two are not mutually exclusive—sometimes civil law can be used to punish, whilst criminal law can be used to remediate.[43] Nevertheless, the importance of criminal and civil systems remaining complementary and co-regulatory, but existing as separate strands of law cannot be underestimated:

> …it is a mistake to compare crime and tort. If three persons are incited by a fourth to break into a house and cause damage each will be guilty of a crime and will receive separate punishment. The inciter will be guilty of the criminal offence of inciting others to commit crime. The other three will be guilty of the crime of breaking in. If the damage [is] caused…then in a civil action the three who caused the damage will be jointly and severally liable… The inciter will also be jointly and severally liable for the damage if he procures the commission of the tort and is a joint tortfeasor.[44]

Liability as blameworthiness is thus a common cornerstone to both civil and criminal law, even if they are crafted and applied in different contexts.[45] In civil law, blameworthiness is usually established by applying common law principles such as taking reasonable care not to harm one's 'neighbour' or person proximate to their conduct,[46] whereas for criminal law, it is the written Acts of some governing

[38] Peter Cane, 'Mens Rea in Tort Law' (2000) 20 Oxford Journal of Legal Studies 533, 555.
[39] Andrew von Hirsch and Martin Wasik, 'Civil Disqualifications Attending Conviction: A Suggested Conceptual Framework' (1997) 56 Cambridge Law Journal 599.
[40] Edoardo Greppi, 'The Evolution of Individual Criminal Responsibility Under International Law' (1999) 81 International Review of the Red Cross 531, 536–537. See also Rebecca Crootof, 'War Torts: Accountability for Autonomous Weapons' (2016) 164 University of Pennsylvania Law Review 1347.
[41] Crootof (n 40) 1361.
[42] Robert Cooter, 'Prices and Sanctions' (1984) 84 Columbia Law Review 1523. See also Cane (n 38).
[43] Cane (n 38).
[44] *CBS Songs Ltd v Amstrad Consumer Electronics plc* [1988] AC 1013 (HL).
[45] William J Stuntz, 'Substance, Process, and the Civil-Criminal Line' (1996) 7 Journal of Contemporary Legal Issues 1, 19.
[46] *Donoghue v Stevenson* [1932] AC 562 (HL).

body such as Parliament or Congress that set out the rules to be complied with.[47] In respect of making determinations of liability, the arbiter of law (the judge) and the arbiter of fact (often a judge but occasionally a jury) are called to offer an assessment of whether one party has broken a particular rule or breached a given duty.[48]

Given the further social significance of a criminal finding of guilt (potentially involving the loss of an individual's liability through a custodial sentence) versus the pecuniary imposition of damages through establishing negligence, the standard of proof for criminal liability is objectively higher than in civil law. This concept is expressed in most legal systems as beyond reasonable doubt as opposed to on the balance of probabilities,[49] and is expressed in somewhat equivocal terms in *Currie v Dempsey*:

> In my opinion [the legal burden of proof] lies on a plaintiff, if the fact alleged (whether affirmative or negative in form) is an essential element in his cause of action, eg if its existence is a condition precedent to his right to maintain the action. The onus is on the defendant, if the allegation is not a denial of an essential ingredient in the cause of action, but is one which, if established, will constitute a good defence, that is, an 'avoidance' of the claim which, prima facie, the plaintiff has.[50]

Moral and physical blameworthiness is also imported into other terms used in the determination of liability. Upon assessment of a particular factual situation, questions may be asked around the intent to engage in a particular act, which in turn invoke determinations of whether an action involves 'strict' liability or whether liability is contingent upon finding a person holding a particular state of mind—legally, the *mens rea* or 'guilty mind'.[51] It is only after exploring the complete factual situation that a person can be held responsible for some kind of illegal or wrongful act.[52]

This determination involves the importation of concepts of knowledge and intention to constitute moral blameworthiness, responsibility, and punishment.[53] Put differently, the concept of intent provides for the ascription of blameworthiness, a reflection of the aphorism that 'an agent is responsible for all and only his intentional actions'.[54] Collectively, lawyers commonly talk of intent as both a mental state of intending some action, and intentionality of the action as motivated by

[47] Douglas Husak, *Overcriminalization: The Limits of the Criminal Law* (Oxford University Press 2008) 9–10.
[48] Mike Redmayne, 'Standards of Proof in Civil Litigation' (1999) 62 Modern Law Review 167.
[49] In Australia, this is discussed in the seminal case of *Briginshaw v Briginshaw* (1938) 60 CLR 336.
[50] *Currie v Dempsey* (1967) 69 SR (NSW) 116, 539.
[51] Matthew R Ginther and others, 'The Language of Mens Rea' (2014) 67 Vanderbilt Law Review 1327.
[52] *Vines v Djordjevitch* (1955) 91 CLR 512, 519.
[53] Bertram F Malle and Sarah E Nelson, 'Judging Mens Rea: The Tension between Folk Concepts and Legal Concepts of Intentionality' (2003) 21 Behavioral Sciences and the Law 563, 564.
[54] John L Mackie, *Ethics: Inventing Right and Wrong* (Penguin UK 1990) 208.

that mental state.[55] Intentionality in criminal law has a defined and precise meaning and purpose, consisting of both the intention to engage in certain conduct and an intention to bring about a result because of that conduct (or knowledge that it will occur).[56] This is a deliberate choice: though 'strict' liability exists in crime where no proof of intention is needed, it is usually reserved for minor or regulatory offences where the removal of proving intent is not considered procedurally unfair to the accused.[57] Equally, punishing only those offences that a person actually plans and then carries out severely constrains the legal system in regulating unlawful conduct.[58]

Hence, although intentionality and intention may appear similar in both civil and criminal law, they are treated differently and can achieve different outcomes. Good motives cannot rescue or defend wrongful conduct, either in tort or crime. In *Caldwell*[59] an individual erected a wharf on public property and was charged with public nuisance. His defence—that the wharf was at the request of, and benefitted, the local community—was dismissed by the court because he had infringed a common right. On the other hand, a malign motive will taint any form of conduct, even if the conduct itself is morally acceptable. For example, a contract is a lawful arrangement between two parties and may be undertaken by any persons in society at large to regulate their dealings. However, a contract that is objectionable on public policy or legal grounds—such as a contract to commit murder—is void and unenforceable.[60]

Thus, criminal law departs from civil law because the bare formulation of mental state and conduct grounds liability, and there is no need to prove a particular effect or outcome. This explains the criminalisation of conduct even where both parties may consent (such as drug dealing or prostitution[61]), where the offence never actually took place (such as attempting to commit a crime[62]) or where the offence was committed by someone else (inchoate crimes such as aiding or abetting, which are treated differently to contributory negligence[63]). Further, it is almost always the State—and not the infringed party—who brings proceedings for the commission of crimes.[64] Conduct might also be criminalised without reference

[55] Malle and Nelson (n 53).
[56] Issues of automatism and involuntariness are beyond the scope of this paper.
[57] David Prendergast, 'The Constitutionality of Strict Liability in Criminal Law' (2011) 33 Dublin University Law Journal 285. Cf Federico Picinali, 'The Denial of Procedural Safeguards in Trials for Regulatory Offences: A Justification' (2017) 11 Criminal Law and Philosophy 681.
[58] Cane (n 38) 553.
[59] *Respublica v Caldwell* 1 US (1 Dall) 150 (Pa Ct of Oyer & Terminer 1785).
[60] *Commonwealth Bank of Australia Ltd v Amadio* (1983) 151 CLR 447.
[61] Barbara Sullivan, 'Rape, Prostitution and Consent' (2007) 40 Australian & New Zealand Journal of Criminology 127.
[62] Ian D Leader-Elliott, 'Framing Preparatory Inchoate Offences in the Criminal Code: The Identity Crime Debacle' (2011) 35 Criminal Law Journal 80.
[63] Joachim Dietrich, 'The Liability of Accessories under Statute, in Equity, and in Criminal Law: Some Common Problems and (Perhaps) Common Solutions' (2010) 34 Melbourne University Law Review 106.
[64] Ric Simmons, 'Private Criminal Justice' (2007) 42 Wake Forest Law Review 911.

to culpability if there was a serious social cost. Referencing Blackstone's *Commentaries*, Binder observed that the formulation of early Crown offences such as treason, carnal knowledge of the queen, piracy, serving a foreign monarch, or harbouring a Catholic priest were punishable without any proof of intent.[65]

Conversely, the purpose of proving intention in civil law (especially torts)—as opposed to in criminal law, where intention may be a fundamental proof of the charge—may be unnecessary. Torts are almost always actioned by the aggrieved parties, and not the State, to receive remedies that place the aggrieved parties as near to their original position before the infringement.[66] Because the focus of tort liability is generally on the existence of a duty of care, a breach of that duty, and in most cases, the suffering of harm, one cannot attempt a tort, plan one, or conspire to cause one.[67] Intention is usually relevant to penalty, not liability; again, this is a deliberate choice. For the victim whose rights have been infringed, they might not necessarily care if an infringement was actuated by malice, recklessness, or negligence. A search for intentionality may well be meaningless to compensate the victim for the harm suffered.

That is not to say that intention in civil law is a useless concept. Exemplary damages may be issued by the court in cases where the conduct was deliberately engaged in and 'of a sufficiently reprehensible kind'.[68] In this way, torts can punish intentional conduct in circumstances where an 'assertion of one's autonomy… produces harmful consequences [it] may justify more onerous liability than negligence'.[69] Intention is also more relevant where torts regulate activity with high social value high risk, such as transporting dangerous goods or manufacturing poisonous chemicals. In these contexts, it is apparent that the differences between negligence and malice are far more relevant to tortious conduct. In the words of Cane, 'when a harm-causing activity has high social value, a requirement of intention for tort liability helps to protect society's interest in the continuance of that activity'.[70]

IV. Problems Applying Liablity to HMT

The appropriate and proportionate imposition of liability in the context of artificial agents is not a novel problem to confront the law. The European Union (EU)

[65] Guyora Binder, 'The Rhetoric of Motive and Intent' (2002) 6 Buffalo Criminal Law Review 16.
[66] Scott Hershovitz, 'The Search for a Grand Unified Theory of Tort Law' (2017) 130 Harvard Law Review 942. Cf Seth Davis and Christopher A Whytock, 'State Remedies for Human Rights' (2018) 98 Boston University Law Review 397.
[67] Though a party may face contributory negligence for playing some part in its commission: Paul S Davies, 'Accessory Liability for Assisting Torts' (2011) 70 Cambridge Law Journal 353; Paul S Davies and Philip Sales, 'Intentional Harm, Accessories and Conspiracies' (2018) 134 Law Quarterly Review 69.
[68] *Lamb v Cotogno* (1987) 164 CLR 1.
[69] Cane (n 38) 553.
[70] ibid.

for example has proposed an Artificial Intelligence Act, the first one like it anywhere in the world.[71] Such regulation would provide a broader contextual and developmental framework for the design and implementation of AI systems, with the European Commission already having adopted a directive regarding the liability of AI systems.[72] Whilst the EU Directive only deals with civil and not criminal liability, it does emplace liability markers on designers, manufacturers, testers and end-users of AI systems where those systems do not comply with the principles in the proposed AI Act.

This places the EU at the forefront of regulating AI, but other countries are aware of the regulatory impacts of AI. The US has passed several legislative instruments which impose obligations upon government agencies to develop rules and guidelines for the design and testing of AI systems, especially for making decisions in the government.[73] By contrast, the UK intends to make no global rules governing AI, but to leave regulation down to the existing regulators such as the Competition and Markets Authority, the Information Commissioner's Office, and the Financial Conduct Authority.[74]

One of the key threads linking all these proposals is that military use of AI—whether meeting the terms of our proposed definition of HMT or otherwise—is excluded, either explicitly or by implication. Further, all of these legislative proposals do little to oust existing rules of liability, where States are largely able to invoke the protection of the 'act of State' doctrine to prevent courts from imposing liability on defence decisions.[75] The EU AI Act explicitly carves out military use whilst the EU Directive provides that national courts must limit disclosure and preserve secrecy in cases involving potentially confidential evidence (which one presumes would cover the technical specifications of military technology). In the US, it is the Office of the Under Secretary of State for Arms Control and International Security that provides for military AI regulation, yet the agency

[71] Noting that the proposed regulation 'shall not apply to AI systems developed or used exclusively for military purposes': Commission, 'Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts' COM (2021) 206 final (EU AI Act).

[72] Commission, 'Proposal for a Directive of the European Parliament and of the Council on Adapting Non-Contractual Civil Liability Rules to Artificial Intelligence' COM (2022) 496 final (EU Directive).

[73] Namely, the National Artificial Intelligence Initiative Act of 2020 (Division E, § 5001) and the AI in Government Act of 2020 (Division U, Title I). See also Executive Order 13960, 'Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government' (Federal Register 12 August 2020) <https://www.federalregister.gov/do-cuments/2020/12/08/2020-27065/promoting-the-use-of-trustworthy-artificial-intelligence-in-the-federal-government> accessed 25 February 2023.

[74] Department for Digital Culture, Media, and Sport, 'Establishing a Pro-Innovation Approach to Regulating AI: An Overview of the UK's Emerging Approach' (Department for Digital Culture, Media, and Sport 18 July 2022) <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1092630/_CP_728__-_Establishing_a_pro-innovation_approach_to_regulating_AI.pdf> accessed 20 January 2023.

[75] Brendan Walker-Munro, 'Exploring Manufacturer Strict Liability as Regulation for Autonomous Military Systems' (2022) 27 Torts Law Journal 182. Compare the scope of the defence in the US courts in *Boyle v United Technologies Corp* 487 US 500 (1988) to the English courts in *Smith v Ministry of Defence* [2013] UKSC 41, [2014] AC 52.

is completely omitted from US AI legislation. The UK carves out military applications under its proposed policy by explicitly mentioning the Ministry of Defence as being a 'domain that [has] existing and distinct approaches to AI regulation' and therefore permitted to write its own ruleset. At the global level, the Group of Governmental Experts under the Convention on Certain Conventional Weapons have argued for seven years about the definitional scope of AI in the military.[76] It is appropriate then that we consider some of the challenges to these existing models of legal regulation that are being applied in the sphere of AI and how they may not sit easily in the context of a military HMT.

A. Theoretical challenges to HMT liability

It is this focus on blameworthiness that will likely be disrupted by the appearance or adoption of HMT and its bi-directional communication between man and machine. In a legal system where the focus is on the punishment of unlawful conduct or the remediation of breaches of rights, any circumstance influencing the blameworthiness of an agent will have serious ramifications for attribution of liability:

> [A]n agent can only be held responsible if they know the particular facts that surround their action, they are able to freely form a decision to act, and are able to select one of the suitable available alternative actions based on the facts of the given situation.[77]

Breaking apart this statement, we can consider three consecutive notions of attribution of liability and blameworthiness that are worth further exploration in the context of HMTs: *knowledge* of the facts, the existence of *suitable alternatives*, and the *freedom* to decide on one of them:[78]

1. Knowledge: Consider for a moment an HMT where the machine component of the team merely provides information or feedback to the human component, but the machine's programming suffers a catastrophic error and the feedback the human receives is completely nonsensical. A decision to engage in some form of conduct based on that erroneous information might be actionable if the conduct is subsequently proven to be unlawful;
2. Suitable alternatives: Such an assessment is in part subjective and in part objective, considering what the individual thought suitable as

[76] Klaudia Klonowska, 'Article 36: Review of AI Decision-Support Systems and Other Emerging Technologies of Warfare' (2022) 23 Yearbook of International Humanitarian Law 123.
[77] Matthias (n 29) 175.
[78] Giovanni Sartor and Andrea Omicini, 'The Autonomy of Technological Systems and Responsibilities for their Use' in Nehal Bhuta and others (eds), *Autonomous Weapons Systems: Law, Ethics, Policy* (Cambridge University Press 2016) 62.

well as the broader social context in which the act occurred.[79] Of course a lack of blameworthiness because of no suitable alternatives does not always exculpate the actor, as acts which are 'the lesser of two evils' can still be an infringement on another's rights;[80] and

3. Freedom of choice: Whether the human component of the HMT experienced a removal of freedom of choice will depend on the circumstances of the conduct and the context of the teaming operation. In circumstances of extreme emergency, or where the human and machine components are inextricably combined, there may be no way to divorce the human in any way that would render a valid freedom of choice. In others, the factual circumstances are highly relevant to blameworthiness: a military HMT operation in a combat zone may result in far less freedom of choice than an administrative HMT operation within an office.

By examining and weighing all three concepts we suggest it is possible to assess the degree to which an HMT might be liable for acts undertaken, and appropriately adjust for the artificial component's effect on human decision-making. The actions within an HMT, assessed partly on actions by a machine and partly on actions by a human, will become enmeshed to varying extents and may lead to overlapping spheres of liability for blameworthiness. Given the nature of HMTs, it is perhaps easiest to 'conceive of their actions as creating a web of overlapping chains of responsibility, both criminal and civil in nature'.[81] As a concept, this idea already appears in the literature in the context of attributing liability to autonomous systems more generally:

> The *mens rea* of the direct perpetrator therefore must be judged in terms of the secondary party's mental state, and will require intent or knowledge. This can also apply to AWS, as their code gives them the ability to perform some decision-making capabilities, and therefore be able to comprehend certain elements of their actions. However, ultimately, their actions are limited by a human agent, who sets parameters for how they are able to act. Therefore, responsibility can be shared by both the AWS and another human counterpart who is involved in its behaviours and actions.[82]

[79] Randolphe Clarke, 'Moral Responsibility, Guilt, and Retributivism' (2016) 20 The Journal of Ethics 121, 124.
[80] James Goudkamp, 'The Spurious Relationship Between Moral Blameworthiness and Liability for Negligence' (2004) 28 Melbourne University Law Review 343.
[81] Jain (n 36) 304.
[82] ibid 310.

B. Pragmatic challenges to HMT liability

There are also some unresolved difficulties in the application of liability and blameworthiness to HMT more generally. The first is identifying which actor within an HMT, whether the machine or human actor, is the one 'making' a decision when the tasks being completed are not repetitive or deterministic.[83] Consider the theoretical effects explored above: what if a human is presented with a tactical scenario in which there are no alternative options which the human considers acceptable? If the human takes what is the only 'reasonable' option, are they really making a decision? The decision has already been made by the machine—perhaps inadvertently—by presenting the information in a way that only one option was possible.

The second challenge, particularly in the military and armed forces context, is the effect of HMTs on the inquisitorial process (such as criminal investigation or civil discovery). These processes often involve determining both a factual substrate of the conduct, but also an assessment of liability. Unfortunately, HMT presents two distinct barriers to these processes. Firstly, much of the technology, automation, or software underpinning HMTs is likely to be protected by trade secrets or military secrecy;[84] and secondly, the opacity of AI-automation programs in HMT means that even where such the code of such programs can be exposed, the apparent nature of decision-making by that code is not readily discernible in a manner understandable by jurors or judges.[85]

The third is the differing legal treatment of various mental defences within and across jurisdictions. It is not within the scope of this article to consider the various natures of impairment, automatism or insanity defences (however they might be labelled); instead, it is to note that the varying degrees, scope, and application of these defences will lead to entirely varied treatments of HMTs in circumstances where judges are called to assess the 'voluntariness' of actions to assign blameworthiness.[86] This is especially the case where many of the mental defences often involve some level of 'impairment' to functioning. Does the human in an HMT really become 'impaired' because of the inclusion of a machine component?[87] Again, the machine may have removed the scope of involuntariness merely by presenting the information to the human in a particular format or fashion.

[83] Shagun Jhaver and others, 'Human-Machine Collaboration for Content Regulation: The Case of Reddit Automoderator' (2019) 26(5) ACM Transactions on Computer-Human Interaction 1, 8; Vincent Boulanin and others, *Limits on Autonomy in Weapon Systems: Identifying Practical Elements of Human Control* (Stockholm International Peace Research Institute 2020).

[84] Gabriele Spina Ali and Ronald Yu, 'Artificial Intelligence Between Transparency and Secrecy: From the EC Whitepaper to the AIA and Beyond' (2021) 12(3) European Journal of Law and Technology 1, 6–8.

[85] Ashley Deeks, 'The Judicial Demand for Explainable Artificial Intelligence' (2019) 119 Columbia Law Review 1829.

[86] Peter Cane, 'Fleeting Mental States' (2002) 59 Cambridge Law Journal 273.

[87] These defences are generally exculpatory doctrines to apply in only the most extreme cases: Steven Yannoulidis, 'Excusing Fleeting Mental States: Provocation, Involuntariness and Normative Practice' (2005) 12 Psychiatry, Psychology and Law 23, 27.

The last challenge for regulating HMTs in a pragmatic sense is determining a remedy that adequately reflects the blameworthiness of the conduct. Most western legal systems have evolved from the perspective that irrespective of the legal entity a claim is brought against, there is nevertheless a 'human who decides whether or not to comply'.[88] For example, civil and criminal actions are often brought against companies as legal entities, but where the actions of those companies are often a sheeted home to individuals within them.[89] Where a machine can be attributed with blameworthiness, there comes the question of how to achieve a penalty or restitution in a manner that is relevant to the machine. Alternately, there is a question of how to apply a remedy to a human who may have had no conscious control of or over the actions they are now alleged to have engaged in.[90]

In summary, these various principles of theoretical and pragmatic challenges in the context of military HMTs can be accounted for. Whatever the intended scheme of regulation is proposed, we consider that it must be capable of addressing the difficulties of applying blameworthiness in the context of HMT operations generally, but also the military and armed forces context more specifically. We consider the best and most efficient approach to involve modifying an existing regulatory scheme to apply to the future use of military HMT operations.

V. A Proposed Framework for Liability in HMT

In this final Section, we propose the leveraging of a concept that has already been explored in the literature—chains of responsibility or 'COR'—as a mechanism for attributing liability in HMT. Originating in the logistics and supply chain industry, COR applies a proactive model of compliance to prevent road and freight accidents. COR legislation for heavy vehicles is already a feature of the legal landscape in Australia.[91]

It is with this framework in mind that we present a COR model for the HMT context in Table V.1. Along the vertical axis, Table V.1 charts the lifecycle of an HMT from conception and design, through manufacture and testing, to procurement and deployment (both domestic and foreign). At each stage of that lifecycle, those involved with HMT will carry responsibilities explored horizontally. These responsibilities are non-exhaustive and intended to provide a high-level example of the types of activities at each stage which are relevant in determining potential legal culpability from the use of HMT. Each of them has been derived

[88] Lawrence Lessig, 'The Zones of Cyberspace' (1996) 48 Stanford Law Review 1403, 1408.

[89] Michael Nietsch, 'Corporate Illegal Conduct and Directors' Liability: An Approach to Personal Accountability for Violations of Corporate Legal Compliance' (2018) 18 Journal of Corporate Law Studies 151.

[90] David Watson, 'The Rhetoric and Reality of Anthropomorphism in Artificial Intelligence' (2019) 29 Minds and Machines 417.

[91] Heavy Vehicle National Law (Queensland); as applied by the Heavy Vehicle National Law Act 2012 (Qld), s 4; Heavy Vehicle National Law Act 2013 (ACT), s 7; Heavy Vehicle (Adoption of National Law) 2013 (NSW), s 4; Heavy Vehicle National Law (South Australia) Act 2013 (SA), s 4; Heavy Vehicle National Law (Tasmania) Act 2013 (TAS), s 4; Heavy Vehicle National Law Application Act 2013 (Vic), s 4.

from the theoretical and pragmatic challenges to HMT liability in Section II and is intended to be read from the perspective of 'reasonable foreseeability' For example, legal and ethical advice in the conduct of military operations is best sought well before the first shot is fired, when advising how a conflict may be fought.[92] Thus, legal and ethical advice should be incorporated into the very design of the HMT.[93] To discharge their responsibilities at the manufacturing stage, those producing HMT interfaces and software should have rigorous testing regimes in place which are capable of detecting flaws and errors to a low tolerance (noting that the systems will ultimately be used in a warfighting capability[94]). Those involved in the manufacture of subcomponents will also need to meet these rigorous standards and be informed by the principal manufacturer of the potential risks.[95]

TABLE V.1

Chain of Responsibility for HMT			
Design Phase	Legal and ethical advice	Are system functions being automated appropriately?	Comply with international and domestic law
Manufacturing	Advise subcontractors of potential liability	Testing at or above industry standard	Fully investigate adverse incidents
Testing	Frequent and robust testing	Address any potential safety concerns, that is, with mandatory warnings	Safety must be 'such as persons generally are entitled to expect'
Contract Negotiation	Include, as contractual terms, the intended scope and operating environment	Disclose all possible safety issues	Include maintenance and upkeep cycles in contract
In-Service	Further testing compatible with intended operational environment	Be aware 'handover' to military authorities does not end liability	Operators must not operate systems until deemed competent
Domestic Deployment	Platform must operate consistent with domestic and international laws	Human decision-maker must be capable of intervening at any time	Systems must be operated in accordance with the manufacturer's instructions and training at all times
International Deployment	Be aware that international and domestic law will apply	Human decision-maker must be capable of intervening at any time	Consider IT security in overseas environments

The concept of COR outlined in Table V.1 is relatively simplistic: it imposes a primary, direct, and non-delegable duty on every person interacting with an

[92] Marcus Schulzke, 'Autonomous Weapons and Distributed Responsibility' (2013) 26 Philosophy & Technology 203, 209.
[93] Heather M Roff, 'The Strategic Robot Problem: Lethal Autonomous Weapons in War' (2014) 13 Journal of Military Ethics 211.
[94] Jai Galliott, 'The Soldier's Tolerance for Autonomous Systems' (2018) 9 Paladyn, Journal of Behavioral Robotics 124.
[95] Walker-Munro (n 75).

HMT to ensure the safety of their individual activities so far as is reasonably practicable.[96] At each level, from design, through manufacture and testing, to 'handover' to military authorities and eventual deployment in military operations, an HMT must be rigorously tested in all intended operational environments. Legal and ethical advice should be sought and incorporated into the design, manufacture, and testing stages. Such testing must be performed by both the manufacturers and military authorities, and testing performed at any specific stage should not be regarded as being conclusive. A 'cut-off' or similar system should always be included in any HMT that permits a human operator (or other person acting remotely) to deactivate the machine component in the event of a failure or incident. Any safety defects, issues, and injuries must be rigorously investigated and either remediated or repaired, or a mandatory warning provided in relation to conduct likely to cause that issue again. In both training and operational use, military commanders bear an additional non-delegable duty to ensure their staff are trained on HMTs and deemed competent in their use. In the absence of clear legal guidance to the contrary, principles of both domestic and international law should be deemed to always apply to the use of HMTs in operational military environments.

In the event of an accident or incident, an investigation is conducted that examines the entire logistic chain to determine where the duty was breached, and by which agent. Breaches of that duty of care may result in the commencement either of civil action (involving pecuniary penalties) or criminal offences (involving potential for penal sentences in severe cases). There exists a legitimate question as to how COR might address any of the theoretical or pragmatic challenges that HMT poses to existing civil and criminal liability approaches. It is therefore the focus of this Part of the article to demonstrate how COR could apply in the context of military HMTs.

A. Applying COR to theoretical challenges to HMT liability

It should be recalled that we examined three consecutive notions of attribution of liability and blameworthiness applying to HMTs: *knowledge* of the facts, the existence of *suitable alternatives*, and the *freedom* to decide on one of them.[97] How should COR apply to these three notions of liability?

Firstly, COR examines the nature of actions taken and decisions made up to and inclusive of the decision to engage in the impugned conduct. In such an *ex*

[96] Amanda Beesley, 'Improving Safety and Compliance, and Simplifying Enforcement – Recent Reforms to Australia's Heavy Vehicle Chain of Responsibility Laws' (International Symposium on Heavy Vehicle Transport Technology, Rotorua, 5 October 2016); Geoff Farnsworth and Jarrad McCarthy, 'Heavy Vehicle National Law Reform: New Approach to Chain of Responsibility Liability' (2016) 68 Governance Directions 41; Wonmongo Lacina Soro, 'Towards an Understanding of Financial Influences on Heavy Vehicle Safety Outcomes' (PhD thesis, Queensland University of Technology 2020).

[97] Sartor and Omicini (n 78).

ante examination, it is not just the blameworthiness of the ultimate decision that is determinative, but of each of the 'steps' that led up to its final execution. Consider the earlier example of a military HMT where the machine programming fails and the human is presented with nonsensical information. In this case, the application of COR's 'reasonably practicable' assessment of safety might determine that the nature of the manufacturer's pre-deployment testing was insufficient and that this was the blameworthy failure. Alternately, it might be a repairer who inserted a faulty component who bears the blame for the incident at hand. This concept of extended liability is certainly not unknown to the literature and reinforces the idea that delictual responsibility is not a pie—'[a]ll involved can theoretically take all the responsibility for the harm caused, and if unjustified, punished'.[98]

Secondly, whilst COR applies equally across the nature of military and non-military actors, it has the flexibility to consider the unique challenges of military service. The application of COR is based on precautions that are 'reasonably practicable' by reference to the controls available, the suitability of those controls, and the cost of controls proportional to the risk posed. Consistent with other approaches to applying liability to emerging military technologies, it is easier to control risks from a manufacturer's office or a designer's factory, places which are far removed from operations against the enemy in a foreign war zone.[99]

Thirdly, COR sidesteps many of the theoretical issues to liability attribution that might occur in the context of military HMT. Again, the curial search in COR is for 'reasonable practicability', not necessarily strict blameworthiness. In circumstances where a military force has not properly trained a human for HMT operations but could have easily done so with the resources and time it had available, underlying fault questions do not arise. The military force bears responsibility under COR and may be prosecuted or litigated accordingly. Military forces are already assessed for this level of compliance under most western work health and safety systems, suggesting that the level of adaptation required to adopt COR is unlikely to be onerous or disruptive to military operations.[100]

Fourthly, by adopting a less prescriptive system for attribution of blameworthiness, there is potential to avoid the injustice of liability being applied to persons not having sufficient control of machines or in circumstances where the machine has malfunctioned.[101] At the same time, COR renders irrelevant the need for militaries to scrutinise the role of humans in a decision cycle.[102] The focus of inquiry in situations of failure is on the reasonableness of safeguards enacted to protect against harms, not the actions of the individual HMT.

[98] Ross W Bellaby, 'Can AI Weapons Make Ethical Decisions?' (2021) 40 Criminal Justice Ethics 86, 96.
[99] *Smith* (n 75); Walker-Munro (n 75).
[100] Nick Turner and Sarah J Tennant, '"As Far as is Reasonably Practicable": Socially Constructing Risk, Safety, and Accidents in Military Operations' (2010) 91 Journal of Business Ethics 21; John A Casciotti, 'Fundamentals of Military Health Law: Governance at the Crossroads of Health Care and Military Functions' (2016) 75 Air Force Law Review 201.
[101] Matthias (n 29).
[102] Ryan (n 30).

Finally, COR is equally adaptable to the vast array of field environments in which modern militaries are prepared to operate. The reasonableness of safety precautions to avoid specific harms recognises that 'a safety measure that would be enough in one situation might be completely inadequate in another and excessive in a third set of circumstances'.[103] Therefore, what may be deemed acceptable to limit HMT risk in a training setting might be inadequate for foreign operations but excessive in joint or allied exercises. Acceptability is dependent entirely on the circumstances of the HMT deployment and the likelihood and magnitude of the risk being guarded against.

B. Applying COR to pragmatic challenges to HMT liability

In the same vein, COR has the potential to drastically limit or eliminate the pragmatic risks to the attribution of HMT liability. The application of COR to HMT operations, especially military operations, recognises the unique factual circumstances in each deployment of HMT and seeks to impose a sliding duty of reasonableness to determine whether liability should apply and to what degree.

Firstly, the idea of needing to determine which aspect of an HMT—human or machine—'made' a decision for any liable conduct is irrelevant. The focus of COR is not strictly limited to the liable conduct in question, but on all the antecedent decisions and circumstances along the chain leading to that conduct. Where a programming error which presents inaccurate or misleading data in an HMT is the causative agent, liability is still attributable to the human for not verifying the information using another technique or system. The fighter pilot who bombs a target without visually verifying and satisfying themselves of a target's validity (and instead relying on the automated or autonomous system) stands to carry some of the punishment or rectification for that fault.

Secondly, COR considers but does not rely on the mental element of each of the individual actors along the chain. Each individual actor has the same duty ('to limit risk to the extent reasonably practicable') but different capacities and methods of discharging that duty, in the same way as modern work health and safety laws operate in the military context.[104] Defences of automatism, involuntariness, or diminished capacity are relevant only to the extent that the actor can discharge the duty, not the existence of the duty. It further recognises that the liability of one individual of the chain may be contingent, or rely upon, the liability of others. The failure of a soldier relying on faulty data is influenced by and partly reliant upon the failure of a manufacturer to properly test the machine components.

[103] National Heavy Vehicle Regulator, 'Primary Duty Definitions' (*National Heavy Vehicle Regulator*, 2022) <https://www.nhvr.gov.au/safety-accreditation-compliance/chain-of-responsibility/the-primary-duty/primary-duty-definitions> accessed 20 January 2023.
[104] Turner and Tennant (n 100).

Thirdly, COR has the potential to move with new developments in technology, including the ability for machines to achieve 'artificial general intelligence'.[105] At such a point, machine components in HMTs may well be treated as moral actors with their own level of agency, at which point they become another link in the COR and their potential for blameworthiness becomes examinable. If a machine can achieve general intelligence in a manner that can be attributed to moral agency, there is no reason why its actions could not be examined through the lens of reasonable practicability for preventing harm.

The fourth benefit to the application of COR in military applications of human-machine teaming is the ability to attribute liability to specific human actors interacting with the system over time. As we stated above, liability is attributed to the human based on the scope and scale of their interaction with the AI system, not the level of their involvement in the blameworthy decision. In that way, COR applies a remedy for blameworthiness to a human actor on the basis of what reasonable steps could or should have been taken. Equally, COR also recognises the taking of reasonable steps against the realisation of harm as a partial or full defence to that liability, incentivising actors to behave in protective and alleviative ways.

Of course, COR has limited ability to counter the challenges of trade secrets, secrecy, or explainable AI. Indeed, these potentially introduce a challenge in the form of the broadly conceptualised 'state of the art' defence. This defence obviates responsibility in COR and similar regimes for defects that could not be detected by reasonably practicable testing available at the time of manufacturing or programming.[106] Many western legal systems have, however, grappled with like concepts for decades. In most cases, they involve the interpretation and application of rules of procedure which are dealt with at the level of individual courts or tribunals (including their military equivalents). There are already calls for action across multiple domains in respect of vesting arbiters of fact with appropriate powers of inquiry, coupled with broader education of the legal fraternity in concepts like AI.[107]

VI. Conclusion

There can be little doubt that although the integration of human and machine elements offers significant benefits to the armed forces, insufficient consideration has been given to how to regulate these integrations. Given the significance of decisions made in the context of military operations, which might involve the deaths of hundreds or thousands of people, we cannot leave the regulation of such events to mere chance and ambiguity. Nor does there appear to be much benefit in

[105] Ben Goertzel, 'Artificial General Intelligence: Concept, State of the Art, and Future Prospects' (2014) 5(1) Journal of Artificial General Intelligence 1; Pei Wang, 'On Defining Artificial Intelligence' (2019) 10(2) Journal of Artificial General Intelligence 1.

[106] Mabel Tsui, 'The State of the Art Defence: Defining the Australian Experience in the Context of Pharmaceuticals' (2013) 13(1) QUT Law Review 132.

[107] Deeks (n 85).

merely outlawing the pursuit of HMT applications, driving the research underground, and delegitimising a purposeful line of human research.

Instead, what is required is a nuanced and purposeful regulatory regime which considers the reasoning for attribution of responsibility, whilst also providing appropriate mechanisms for restitution and punishment. This is much for the benefit of our armed forces as for the protection of the rules-based global order: military officers and personnel need to know the legal limits of their conduct, what can be done in war and peacetime, and what consequences might attach when they step outside those boundaries.

There is still more to be done. The exact parameters of technologies designed to constitute HMT and how they are defined in law will need a more comprehensive examination than was possible in this article. The definitions will need to be expansive enough to capture those technologies at the forefront of military and civilian research, but also those yet to be contemplated. Alternately, new legal definitions for those technologies will need to be included in their own regulatory regime to eliminate grey areas and ambiguity. Just like our treatment of AI, we need to ensure that the definition is clear, unambiguous, and is not leading to inaccurate or oversimplified definitions of the technology.[108]

The work on regulating HMT also will not end with the possible introduction of a COR regime. There will no doubt be developments in warfighting technology which escape even the most carefully drafted working definitions of HMTs. Systemic difficulties which cannot be resolved at the procedural level of courts and tribunals will inevitably arise. Future avenues of research might look at how COR regimes could be tailored to specific military operations, or how military COR might be adapted to civilian environments. At the same time, broader calls for 'explainable AI', rules of evidence for AI, and judicial education need to be heeded to ensure that any COR regime enacted by an armed force is capable of being dealt with properly and justifiably.[109]

[108] Wang (n 105).

[109] Katherine Quezada-Tavárez, Plixavra Vogiatzoglou, and Sofie Royer, 'Legal Challenges in Bringing AI Evidence to the Criminal Courtroom' (2021) 12 New Journal of European Criminal Law 531.