# IS JOHNNY FIVE ALIVE OR DID IT SHORT CIRCUIT? CAN AND SHOULD AN ARTIFICIALLY INTELLIGENT MACHINE BE HELD ACCOUNTABLE IN WAR OR IS IT MERELY A WEAPON?

Aaron Gevers*

Imagine the scene: in the not too distant future you, as a U.S. Army Captain, are on a mission in the snowy mountain topography of some far off land. Accompanying you are several soldiers, one of which, Sergeant Johnny Five,[1] sits prone behind a large caliber machine gun as you overlook a village. You have been advised that enemy combatants are the sole residents of the village below and you are instructed to eliminate the targets. You sit behind a spotting scope and direct Johnny to engage five targets dressed in distinct white camouflage and clear the village. Johnny confirms he sees the targets. You give a resounding "Execute" order. Johnny then, being a well programmed, trained, and armored artificially intelligent robot, proceeds down to the village. Johnny opens fire upon reaching a distance of thirty meters from the combatants. Johnny confirms the targets are down. You go to confirm his assessment through the spotting scope but realize that Johnny has not stopped firing and is now clearing the village of all its residents – residents who now lay in the blood-soaked snow – and are, quite clearly, women and children non-combatants.

* Student Note, J.D. 2014, Rutgers School of Law – Camden. Now serving as Assistant General Counsel, Viking Group.

[1] As a Sergeant, Johnny is of the E-5 pay-grade. He is, for our purposes, Johnny Five (E-5) as it were. *See The United States Military Enlisted Rank Insignia*, U.S. DEPARTMENT OF DEF., http://www.defense.gov/about/insignias/enlisted.aspx (last visited Mar. 15, 2015).

You watch speechless in horror, stunned at this effortless slaughter, unable to give an order over the radio.[2]

You begin to wonder how this is possible. How could Johnny, with all his advanced software, his superior intelligence, and understanding of command prompts mistakenly kill innocent civilians? Was it a mistake? Is he at fault? Can he even be at fault? Are you at fault? Weren't you the one utilizing a weapon to clear the village?

These are the questions posed by the inevitable utilization of artificial intelligence (hereinafter "AI") on the battlefield. Like any technology they are sure to malfunction or not work as advertised, regardless of how comprehensive the warranty may be. But perhaps they are no longer machines and have reached a point of becoming sentient beings. If that's the case, then Johnny is at fault for this apparent war crime.[3]

The purpose of this note is to recognize and resolve these issues and come to the invariable conclusion that once a true AI is discovered and deployed, it can and should be held liable as any human actor would. This note goes about this task utilizing several sections. The first section reviews the current state of technology and determines whether the computing power necessary to create and operate Johnny Five is feasible. The

---

[2] For similar scenarios and thought experiments about AIs and robots going awry or coming into their own – not to mention some arguably great cinema – see ROBOCOP (Metro-Goldwyn-Mayer 2014); PROMETHEUS (Twentieth Century Fox 2012); ALIEN (Brandywine Productions 1979); 2001: A SPACE ODYSSEY (Metro-Goldwyn-Mayer 1968); BICENTENNIAL MAN (Columbia Pictures 1999); I, ROBOT (Twentieth Century Fox 2004); THE MATRIX (Warner Bros. 1999); BLADE RUNNER (Ladd Company 1982); TRON (Walt Disney 1982); TRON: LEGACY (Walt Disney 2010); THE IRON GIANT (Warner Bros. 1999); SHORT CIRCUIT (TriStar Pictures 1986).

[3] Though hopefully Johnny Five would have been hardwired with some basic principles like those presented by Isaac Asimov so such a war crime would never occur. *See generally* Isaac Asimov, *Runaround*, ASTOUNDING SCI. FICTION, Mar. 1942. The three laws of robotics as expounded by Asimov are: (1) a robot may not injure a human being or, through inaction, allow a human being to come to harm; (2) a robot must obey any orders given to it by human beings, except where such orders would conflict with the First Law; (3) a robot must protect its own existence as long as such protection does not conflict with the First or Second Law. *Id.* Of course, in our example, laws like these would be altered for conflicts to reflect something like "a robot may not injure a human being except for those aiding, abetting, or directly operating for the Taliban, or through inaction allow a human being except for . . . to come to harm."

second section determines whether an AI can truly ever "think" or essentially become so close to human processes as to be an inorganic human. The third section reviews the basic laws of war and determines that AI robots are not per se excluded from utilization in combat scenarios. The fourth section discusses the liability of conventional human soldiers for acts committed in war or in contingency operations and whether those sole actors or those actors and their commanders are responsible. This section also examines some analogous material, which may be employed in determining the liability of Johnny Five. The fifth section ponders whether Johnny Five or his commanders are liable under current legal regimes including both the Uniform Code of Military Justice and international law such as the Rome Statute for the International Criminal Court. The sixth and final part is a short section, which posits a few changes that could be made to accommodate issues presented by Johnny Five in terms of liability for war crimes.

Before we begin, it's important to note that AI is distinct from a mere autonomous machine. What we have today which we call "drones" or "Predator drones" would be defined as semi-autonomous weapon systems because, once activated, they are only intended to engage individual targets or specific target groups that have been selected by a human operator.[4] Fully autonomous weapons on the other hand are systems that, once activated, can select and engage targets without further intervention by a human operator.[5] However, even these fully autonomous weapons are not of the type this note is concerned with; these weapons are merely weak AI. This note concerns itself with Strong AI – a computer that can truly be called a mind and can think, reason, imagine, and do all the things we currently associate with the human brain.[6] Weak AI as we see currently utilized in combat and in the commercial civilian

---

[4] Directive 3000.09, Autonomy in Weapon Systems, 14 (DOD 2012), *available at* http://www.dtic.mil/whs/directives/corres/pdf/300009p.pdf.

[5] *Id.* at 13.

[6] *Philosophy of Mind-Functionalism: Strong and Weak AI*, PHIL. ONLINE, http://www.philosophyonline.co.uk/oldsite/pom/pom_functionalism_AI.htm (last visited Mar. 15, 2015).

realm merely imitates the human brain and cannot truly think or maintain a consciousness.[7]

## I. SCIENCE FICTION OR SCIENCE FACT?

One of the very first questions we must tackle is whether or not Strong AI is even possible. Some would say it is mere science fiction, while others would posit it is simply a matter of time. In reality, Strong AI is becoming more and more likely. We need only look to our current technology and the rate of progress we have achieved to determine that Strong AI is almost inevitable.

The amount of computing power necessary to replicate the human brain – or any brain which can think – is massive. Researchers have just recently tried to simulate the human thought process for a length of one second.[8] It took 82,944 processors forty minutes to do so while utilizing one petabyte of system memory.[9] Clearly our current technology is not up to the task, but the technology on the horizon is. In fact, though some suggest Moore's law – that the number of transistors that can be placed on silicon doubles every two years – will be hard to keep up with,[10] rapid strides are being made towards what is known as quantum computing. This new type of computational method is likely what will be utilized in a Strong AI being. It allows for the use of qubits, which hold two values, as opposed to the one value held by today's transistors.[11] This allows for chip parts to

---

[7] *Id.*

[8] Ryan Whitwam, *Simulating One Second of Human Brain Activity Takes 82,944 Processors*, EXTREMETECH (Aug. 5, 2013, 1:34 PM), http://www.extremetech.com/extreme/163051-simulating-1-second-of-human-brain-activity-takes-82944-processors.

[9] *Id.*

[10] Agam Shah, *Intel: Keeping up with Moore's Law is Becoming a Challenge*, PCWORLD (May 8, 2013, 11:22 AM), http://www.pcworld.com/article/2038207/intel-keeping-up-with-moores-law-becoming-a-challenge.html.

[11] Cade Metz, *Physicists Foretell Quantum Computer with Single-Atom Transistor*, WIRED (Feb. 20, 2012, 3:46 PM), http://www.wired.com/wiredenterprise/2012/02/sa-transistor/.

be 100 times smaller than current parts, and the processing power increases exponentially with each qubit added.[12] This technology was previously believed to be unsustainable but recent advances have proven it can currently be sustained for forty minutes at room temperature, a huge leap from its previous state.[13]

Some scientists have postulated that with the rate at which technology is progressing, the day AI becomes reality is sooner than we might think. Dr. Ray Kurzwiel, a well-known scientist has stated that the year AI *surpasses* human intelligence will be 2045.[14] He believes, after careful calculations regarding exponential computing, Moore's law, and economics, human intelligence level computing will be reached by the mid-2020's.[15]

To lend more credence to Dr. Kurzweil's prediction and the related likelihood that such intelligence will be applied in robotics and on the battlefield we must look to current and in-development applications. For instance, take the weak AI machines: Watson, Siri, and Deep Blue. IBM's Watson is a question-answer machine, which can comprehend questions posed in natural language and answer them.[16] Watson has done so well at this that it won "Jeopardy!" two years ago[17] and is now poised to be deployed into the Cloud computing community.[18]

---

[12] *Id.*

[13] Emily Chung, *Qubit Record Moves Quantum Computing Forward*, CBCNEWS (Nov. 14, 2013, 4:57 PM), http://www.cbc.ca/news/technology/qubit-record-moves-quantum-computing-forward-1.2426986.

[14] Lev Grossman, *2045: The Year Man Becomes Immortal*, TIME MAG. (Feb. 10, 2011), http://www.time.com/time/magazine/article/0,9171,2048299-1,00.html.

[15] *Id.*

[16] John Markoff, *Computer Wins on "Jeopardy!": Trivial, It's Not*, N.Y.TIMES, Feb. 17, 2011, http://www.nytimes.com/2011/02/17/science/17jeopardy-watson.html?pagewanted=all.

[17] *Id.*

[18] Serdar Yegulalp, *Watson as a Service: IBM Preps AI in the Cloud*, INFOWORLD (Nov. 15, 2013), http://www.infoworld.com/t/cloud-computing/388atson-service-ibm-preps-ai-in-the-cloud-230901.

It now runs three times faster than it did when it won "Jeopardy!"[19] IBM's Deep Blue machine recently beat the world champion of chess, Gary Kasparov, in a highly publicized battle of wits.[20] Apple's Siri, which draws from the wealth of information on the internet to generate human-like responses to human queries, can hold short conversations and even provide witty retorts. While Siri may be programmed with select responses, she effectively learns how to interpret and respond by collating voluminous amounts of data, analyzing it, and altering her processes so that she may respond more effectively.[21] Each of these machines demonstrates humanity's will and curiosity to create and develop a machine as smart as or smarter than himself. This drive, coupled with the tremendous amount of funding at the disposal of companies like IBM and Apple, indicates that we are en route to creating true Strong AI.

In terms of robotics to make Sergeant Johnny Five operational – let alone inconspicuous and comforting – we needn't look further than robots like ASIMO, the current military exoskeletons in development, cutting edge prosthetics, and the so-called smart weapons we are soon to employ. Honda's ASIMO, or Advanced Step-in Innovative Mobility, is billed as the most advanced humanoid robot in the world, namely because it can walk independently and climb stairs.[22] Honda has so much faith in ASIMO's spatial proficiency that the four foot, three inch robot works as a receptionist in Honda's Wako, Saitama, Japan office frequently greeting guests and leading them around the facilities.[23] ASIMO can see, recognize,

---

[19] *Id.*

[20] Adam Ford & Tim van Gelder, *Into the Deep Blue Yonder – Artificial Intelligence and Intelligence Amplification*, H PLUS MAG. (Oct. 22, 2013), http://hplusmagazine.com/2013/10/22/into-the-deep-blue-yonder-artificial-intelligence-and-intelligence-amplification/.

[21] Bernadette Johnson, *How Siri Works*, HOWSTUFFWORKS (Feb. 6, 2013), http://electronics.howstuffworks.com/gadgets/high-tech-gadgets/siri.htm.

[22] Lee Ann Obringer & Jonathan Strickland, *How ASIMO Works*, HOWSTUFFWORKS (Apr. 11, 2007), http://science.howstuffworks.com/asimo.htm.

[23] *See id.*; Honda, *Greeting Passers-by* (Dec. 2005), http://world.honda.com/ASIMO/video/2005/ic-greeting/index.html.

and avoid running into objects as well as detect multiple objects, determine distance, perceive motion, recognize programmed faces, and even interpret hand motions.[24]

Like ASIMO, Atlas is another bipedal humanoid robot, albeit much larger and arguably more advanced. Atlas is a 6'2", 330 pound humanoid robot produced by Boston Dynamics and is currently being prepped to undergo new code, which will allow it to navigate degraded terrain, drive a utility vehicle, and enter buildings with the hope that it will one day save lives in disaster zones.[25] Atlas has twenty-eight hydraulically actuated joints allowing it to crouch, kneel, or jump down to a lower level, and despite being less visually appealing than ASIMO, it appears to be more proprioceptive and ergo more stable.[26] Similarly, the company that produced Atlas, Boston Dynamics, has also produced at least two other robots for the U.S. military and DARPA.[27] Bigdog is a three feet long, two-and-a-half foot tall robot capable of throwing fifty pound cinder blocks, or potentially grenades, and carrying a significant amount of weight while traversing rough terrain.[28] WildCat, another Boston Dynamics robot built for DARPA, is a smaller, quadruped robot currently capable of running freely up to sixteen miles-per-hour, but will soon be able to run fifty miles-per-hour over any terrain.[29] Taking all the attributes from these various robots into account–speed, strength, spatial recognition,

---

[24] Obringer & Strickland, *supra* note 22.

[25] Jason Kehe, *This 6-foot, 330-Pound Robot May One Day Save Your Life*, WIRED (Nov. 20, 2013, 9:30 AM), http://www.wired.com/dangerroom/2013/11/atlas-robot/.

[26] *Id.*

[27] DARPA refers to the Defense Advanced Research Projects Agency.

[28] Sebastian Anthony, *U.S. Military's BigDog Robot Learns to Throw Cinder Blocks, Grenades…*, EXTREMETECH (Mar. 1, 2013, 9:04 AM), http://www.extremetech.com/extreme/149732-us-militarys-bigdog-robot-learns-to-throw-cinder-blocks-grenades.

[29] Sebastian Anthony, *Meet DARPA's WildCat: A Free-Running Quadruped Robot That Will Soon Reach 50 mph over Rough Terrain*, EXTREMETECH (Oct. 4, 2013, 11:04 AM), http://www.extremetech.com/extreme/168008-meet-darpas-wildcat-a-free-running-quadruped-robot-that-will-soon-reach-50-mph-over-rough-terrain.

and proprioception – it's naïve to believe that a robot like our Sergeant Johnny Five isn't feasible.

Moreover, the U.S. military has already expressed significant interest in utilizing robotic structures on the battlefield in conjunction with human actors.[30] Take both the Raytheon Sarcos Exoskeleton and Lockheed Martin's appropriately named Human Universal Load Carrier (HULC) system. Both systems allow the wearer to lift 200 pounds repeatedly without tiring. This system would enable soldiers to carry large rucksacks downrange for hours on end without fatigue.[31] DARPA recently announced a new cutting-edge program appropriately dubbed the "Avatar" program. DARPA states the program will "develop interfaces and algorithms to enable a soldier to effectively partner with a semi-autonomous bipedal machine and allow it to act as the soldier's surrogate."[32]

Coupling the genuinely exceptional strides being made in the computing fields with the current state of progress to AI and robotics, we can conclude that we, as humans, wish to create in our own image as God created us in his. It simply takes mankind a bit longer than six days to do so. Have no worry though, we should complete our creations within the next fifteen years.

## II. AI AS A BEING

Having established the feasibility and likelihood of AI and humanoid robots like Sergeant Johnny Five we must now

[30] *See* Warren Peace, *GIs Testing 'Smart' Weapons That Leave Nowhere to Hide*, STARS & STRIPES (Nov. 15, 2010), http://www.stripes.com/news/gis-testing-smart-weapons-that-leave-nowhere-to-hide-1.125514. The U.S. Military is no stranger to researching and testing highly lethal technologically enhanced munitions on the battlefield which know when to detonate near an enemy even if he is not in the line of sight.

[31] *See* David Goldstein, *I Am Iron Man: Top 5 Exoskeleton Robots*, DISCOVERY NEWS (Nov. 27, 2012, 3:00 AM), http://news.discovery.com/tech/robotics/exoskeleton-robots-top-5.htm; *see also HULC Overview*, LOCKHEED MARTIN, http://www.lockheedmartin.com/us/products/hulc.html (last visited Mar. 15, 2015).

[32] Katie Drummond, *Pentagon's Project 'Avatar': Same as the Movie, but with Robots Instead of Aliens*, WIRED (Feb. 16, 2012, 4:51 PM), http://www.wired.com/2012/02/darpa-sci-fi/.

turn to whether or not an artificially intelligent machine is, for all intents and purposes, the same as a human actor. Determining whether someone or something can think has been at the heart of philosophers for some time, so several tests have been developed to ascertain what it means to think and have a true mind.

Alan Turing[33] devised a test known as the "Imitation Game" to determine whether a computer can think.[34] Essentially the test involved three participants: a human interrogator, a human respondent, and a computer respondent. The human interrogator was separate from the respondents and could only communicate with them individually. It was the interrogator's object to discern, by asking questions, which respondent was human and which was computer. If the interrogator could not differentiate human from computer, then the computer was of such intelligence that it could effectively think.[35] Yet it is not enough for the computer to merely fool ordinary observers, but rather to fool interrogators who knew that one of the participants was a machine.[36] The computer was also required to pass this test repeatedly – so much so that the interrogator had no more than a seventy percent chance of guessing which participant was human, or alternatively that the computer had a thirty percent success rate in going undetected.[37] While a seventy percent success rate seems awfully high, it's important to note that this merely indicated *some* level of thought on the part of the machine. This computer was known as ELIZA, a natural language processing system, which deconstructed incoming phrases and matched them with

---

33 Alan Turing is credited with the creation of computer science and AI theory. STEVEN HOMER & ALAN L. SELMAN, COMPUTABILITY AND COMPLEXITY THEORY 35-36 (2001), *available at* http://books.google.com/books?id=r5kOgS1IB-8C&printsec=frontcover#v=onepage&q=35&f=false.

34 Graham Oppy & David Dowe, *The Turing Test*, STAN. ENCYCLOPEDIA OF PHIL. (Jan. 26, 2011), http://plato.stanford.edu/archives/spr2011/entries/turing-test/.

35 *Id.*

36 *Id.*

37 *Id.*

stock responses. [38] This, of course, only maintained the appearance of an intelligent conversation for a brief period. More successfully, a program known as Parry, which used more sophisticated natural language processing techniques than ELIZA, was able to simulate a paranoid patient and achieve the thirty percent pass rate, albeit in front of psychotherapist judges.[39] Because psychotherapists are looking for broken, almost robotic speech, it is a stretch to consider Parry a successful participant of the Turing Test.

Consequently, a more arduous test has been proposed, aptly dubbed the Turing Test 2.0. In this test, machines are asked to mimic certain human visual abilities as opposed to merely written abilities.[40] Humans, unlike machines, are good at describing where one object is in relation to another object and picking up on the relevance of certain objects when it involves subjective judgments.[41] AI must be able to pass this test as well to be considered Strong AI.

While there are many objections to Turing's game,[42] it is still effective in weeding out weak AI like ELIZA and Parry. For all intents and purposes Turing's game could likely only be completed by a sufficiently Strong AI similar to the affable Sergeant Johnny Five.

---

[38] *ELIZA: A Real Example of a Turing Test*, OXFORD DICTIONARIES: OXFORDWORDS BLOG (June 22, 2012), http://blog.oxforddictionaries.com/2012/06/turing-test/.

[39] William J. Rapaport, The Turing Test 8 (2005) (unpublished manuscript), *available at* http://www.cse.buffalo.edu/~rapaport/Papers/ell2.pdf.

[40] Daniel Honan, *A Computer Walked into a Bar: Take the Turing Test 2.0*, BIGTHINK (June 23, 2012, 12:00 AM), http://bigthink.com/think-tank/a-computer-walked-into-a-bar-the-turing-test-20.

[41] *Id.*

[42] There is the Theological Objection, the Heads in the Sand Objection, the Mathematical Objection, the Argument for Consciousness, the Argument's from Various Disabilities, Lady Lovelace's Objection, Argument from Continuity of the Nervous System, Argument from Informality of Behavior, and Argument from Extra-Sensory Perception. Oppy & Dowe, *supra* note 34. Because these objections and arguments are easily dealt with, in turn, we need only examine the issues raised by the Chinese Room Argument, which poses the biggest question to the Turing Test. *See id.*

Philosopher John Searle felt that imitation of a human's awareness and consciousness does not a human make. Consequently, Searle proposed the Chinese Room Experiment in opposition to the Turing Test.[43] The experiment involves a man situated in a room with an extensive collection of books.[44] The man is slipped a message under a door, which contains numerous Chinese symbols forming, unknown to the man, a coherent sentence.[45] The man then goes through the books and scribbles down the appropriate response, in Chinese symbols, as directed.[46] The man slips the response under the door and the process repeats, effectively forming a conversation in Chinese.[47] The key is that the man does not understand Chinese and is merely responding by the syntax which he receives.[48] Likewise, a computer response is posited to merely identify syntax as it cannot understand language.[49] It, therefore, cannot be aware or conscious of what it is actually stating and, by association, doing. Searle asserts that genuine thought and understanding require something *more* than mere computation.[50]

There are numerous rebuttals to Searle's experiment, chief among them the Systems Reply. It posits that the man is simply a part of a larger system consisting of the books, instructions, and any intermediate phases. [51] While the man

---

[43] David Cole, *The Chinese Room Argument*, STAN. ENCYCLOPEDIA OF PHIL. (Sept. 22, 2009), http://plato.stanford.edu/archives/sum2013/entries/chinese-room/.

[44] David L. Anderson, *Searle and the Chinese Room Argument*, CONSORTIUM ON COGNITIVE SCI. INSTRUCTION (2006), http://www.mind.ilstu.edu/curriculum/searle_chinese_room/searle_chinese_room.php.

[45] *Id.*

[46] *Id.*

[47] *Id.*

[48] *Id.*

[49] *Id.*

[50] Anderson, *supra* note 44. It's important to note that Searle doesn't convey what more is required for genuine thought and understanding. *Id.*

[51] Cole, *supra* note 43.

may not understand Chinese, the system as a whole does. Similarly, while a central processing unit of an AI may not understand the language it projects and receives or the actions it commits, the system as a whole does.[52] Dr. Ray Kurzweil is an advocate of this reply and states that Searle contradicts himself when he says the machine speaks Chinese but doesn't understand Chinese.[53] It's Kurzweil's assertion that if the system displays the apparent capacity to understand Chinese, "[i]t would have to indeed, understand Chinese."[54] Likewise, Jack Copeland equated the calculations the mind makes in catching a ball with understanding Chinese; though we do not understand the specific calculations our mind makes to compensate for speed, gravity, and other variables in catching a ball, it does not mean we do not comprehend what it is we are doing and why.[55]

Equally as strong as the Systems Reply, and especially salient for our purposes, is the Robot Reply. This reply concedes that the man or computer trapped in the Chinese Room does not understand Chinese or know what the words mean.[56] Yet, once the robot is freed from the room, and provided it has all the senses a person would, it could attach meanings to the symbols and actually begin to understand the language through practical application, association, and memory.[57]

Perhaps more abstract, the Other Minds Reply tests the point behind Searle's hypothesis. In Searle's Chinese Room experiment, he assumes that humans truly understand Chinese or any other language as opposed to merely engaging in systematic processes and replies.[58] The problem of "Other Minds" has been a central problem in philosophy and is taught

---

[52] *Id.*

[53] *Id.*

[54] *Id.*

[55] *Id.*

[56] *Id.*

[57] Cole, *supra* note 43.

[58] *Id.*

in most introductory Philosophy classes. The premise – an epistemic one – is that we simply cannot know if anyone or everyone around us experiences the world as we do or has a consciousness to the same extent we do.[59] Given this premise, Searle cannot say with any certainty that we as humans are not simply engaging in the same conduct as a man would in the Chinese Room. Therefore, since we attribute cognition to other people, we must in principle attribute cognition to computers.[60]

Searle's Chinese Room presents the biggest problem in determining whether AI can ever truly be attributed human characteristics to be seen as a fellow sentient being. However, given the many strong replies to Searle's argument, it is reasonable to conclude that we simply will not be able to ascertain whether a computer is merely processing language, images, and the like, or whether it is truly thinking, imagining, and creating, utilizing the same processes we do. In the end, and in the eyes of the law, if it acts like a duck and quacks like a duck, we should treat it like a duck. Because we may never know how a Strong AI reaches its conclusion we should err on the side of caution and, barring evidence to the contrary, conclude that the AI acted as a human would and therefore treat it as a human. Under this conclusion, Sergeant Johnny Five would not be a weapon, but rather a person. Instead of equating Johnny Five to a missile, he is equated to his fellow combatants. Nonetheless, we must consider both interpretations of AI sentience in determining how to hold a machine liable for war crimes.[61] Issues still exist when a weak AI machine commits violations of the rules of war.

---

[59] Alec Hyslop, *Other Minds*, STAN. ENCYCLOPEDIA OF PHIL. (Jan. 14, 2014), http://plato.stanford.edu/entries/other-minds/.

[60] Cole, *supra* note 43.

[61] We need not concern ourselves with the emotional element, or lack thereof, in regard to AI. Very few criminal statutes, let alone International Humanitarian or International Criminal Offenses, require any sort of emotional element for a specific offense. *See* Gabriel Hallevy, '*I, Robot – I, Criminal*' - *When Science Fiction Becomes Reality: Legal Liability of AI Robots Committing Criminal Offenses*, 22 SYRACUSE SCI. & TECH. L. REP. 1, 7 (2010) (citing generally JEROME HALL, GENERAL PRINCIPLES OF CRIMINAL LAW 70-211 (2d ed. 2005) (1960)).

## III. A QUICK PRIMER ON THE RULES OF WAR AND WHY STRONG AI ROBOTS ARE NOT EXCLUDED FROM UTILIZATION IN COMBAT

*Jus in bello,* as opposed to *jus ad bellum,*[62] is the area of law which deals with how the parties to an international conflict conduct an armed conflict once engagement has occurred.[63] *Jus in bello* has four key principles: Military Necessity, Distinction, Proportionality, and Humanity.[64] Arguments have been made that on these bases alone, artificially intelligent machines, or at least autonomous weapons, should never be utilized, but this argument misses on many key elements.[65] These undeveloped arguments fail to truly comprehend the similarities – or rather the identical nature – between AI and human intelligence. For true Strong AI to be banned under the principles of *jus in bello*, it would follow that human beings would also be in violation of the laws of armed conflict.[66]

---

[62] Karma Nabulsi, *Jus ad Bellum/Jus in Bello*, CRIMES OF WAR, http://www.crimesofwar.org/a-z-guide/jus-ad-bellum-jus-in-bello/ (last visited Mar. 15, 2015). *Jus ad bellum* is typically the branch of law that deals with the legitimate reasons a state may engage in armed conflict. *Id.* It focuses on what criteria, if any, render a war just. For instance, the Charter of the United Nations, Article 2 declares: "[a]ll members shall refrain in their international relations from the threat or use of force against the territorial integrity or political independence of any state, or in any other manner inconsistent with the purposes of the United Nations." *Id.* In comparison, Article 51 of the same charter states: "[n]othing in the present Charter shall impair the inherent right of individual or collective self-defense if an armed attack occurs against a Member of the United Nations." *Id.*

[63] *Id.*

[64] DEREK GRIMES ET AL., INT'L & OPERATIONAL LAW DEP'T, LAW OF WAR HANDBOOK 164 (Keith E. Puls ed., 2005), *available at* http://www.loc.gov/rr/frd/Military_Law/pdf/law-war-handbook-2005.pdf.

[65] *See generally* HUMAN RIGHTS WATCH - INT'L. HUMAN RIGHTS CLINIC, LOSING HUMANITY: THE CASE AGAINST KILLER ROBOTS (2012), *available at* http://www.hrw.org/sites/default/files/reports/arms1112ForUpload_0_0.pdf.

[66] This is, of course, an absurd notion, but perhaps one that should be employed given how often the human race decides to demonstrate zero semblance of Military Necessity, Distinction, Proportionality, or Humanity in armed conflict. Perhaps it is something better left to the robots after all.

Military Necessity is essentially composed of two elements: (1) a military requirement to undertake an action and (2) such action must not be forbidden by the law of war.[67] It is important to note that military necessity is typically not a defense to law of war violations because the laws of war were crafted to include consideration of military necessity.[68] So, regarding the legality of employed AI in combat, a state could not simply respond that it was a military necessity whenever an issue arose.

Human Rights Watch (HRW) has argued that it is a matter of military necessity to use the AI robots because they will be far superior to any other type of weapon, but that an armed conflict dominated by machines could have disastrous consequences.[69] The two clauses of this argument seem diametrically opposed. Undoubtedly, the utilization of AI is a military necessity; the prevention of the loss of human life is always an objective and a necessity of armed conflict. Utilizing AI soldiers instead of human soldiers prevents loss of life to those who would've served in combat. This very goal alone – that we could minimize human casualties by employing robotic counterparts – should be enough to garner necessity. Still, HRW's argument that an AI might unnecessarily fire upon a wounded or surrendering soldier[70] assumes the AI has simply been given a gun and told to "shoot anything that moves," which is hard to swallow. More likely is that an AI of the intelligence considered here and in Johnny Five would be at least as capable as a human being, if not more so, in determining whether to kill or capture an enemy combatant or execute a particular objective. In fact, an AI may be more likely to accurately weigh the circumstances and determine with greater ease whether an action is a military necessity.

Distinction is the quintessence of the law of war. The principle requires that military attacks should be directed at combatants and military targets, not civilians or civilian

---

[67] GRIMES ET AL., *supra* note 64, at 164-65.

[68] *Id.* at 165.

[69] HUMAN RIGHTS WATCH - INT'L. HUMAN RIGHTS CLINIC, *supra* note 65, at 35.

[70] *Id.*

property.[71] Indiscriminate attacks are strictly prohibited per the Protocols Additional to the Geneva Conventions.[72] Specifically, attacks, which employ a method or means of combat where the effects cannot be directed at a specific military objective are not limited and are of a nature to strike both military objectives and civilians or just civilian objects without distinction.[73]

HRW has argued that robots employed in armed conflict will have difficulty distinguishing between civilians and armed combatants – a difficult assessment even for human standards. This argument is based on the assumption of potentially inadequate sensors and asserts that these inadequacies present a significant problem and limitation for the utilization of Strong AI robots.[74] This assessment is misguided given our lack of knowledge of the ability of the robot's sensors,[75] but also due to the fact that human soldiers are without adequate "sensors" to consistently and accurately detect a combatant versus a civilian. HRW misses that these judgments are not always those of a "snap" nature and that factors other than mere physical appearance are required in the assessment of a combatant. These ancillary factors may be evaluated quicker and more accurately by Johnny Five. Systems which incorporate databases of terror suspects or specific enemy combatants may be readily accessible to the AI and easily matched with targets in sight. Conversely, the AI may just as easily be able to determine that the suspected targets are not those sought and ultimately avoid an unnecessary firefight and deaths.

---

[71] GRIMES ET AL., *supra* note 64, at 166.

[72] Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts (Protocol I), June 8, 1977, 1125 U.N.T.S. 3, art. 51(4) [hereinafter Protocol Additional to the Geneva Conventions of 12 August 1949].

[73] *Id.*

[74] HUMAN RIGHTS WATCH - INT'L. HUMAN RIGHTS CLINIC, *supra* note 65, at 31.

[75] As a reminder, in our scenario of Johnny Five, we operate on the notion that Johnny Five has capabilities, or senses, at least on par with those of his human counterparts; Johnny Five is for all intents and purposes a human with solely mechanical parts.

Furthermore, HRW's assertion that Johnny Five would be incapable of distinguishing an individual's intentions due to a lack of emotion is also inaccurate. The argument presupposes one needs to experience emotions to understand intention, to read body language, or to pick up on other similar cues in the environment which may provide context and require a cease fire order.[76] Rather, one does not need to experience emotion to be able to identify it. For instance, a vacuum cleaner can currently sense whether a person is relaxed or stressed and adjust its movements so as to either comfort or keep its distance from a person.[77] It is not beyond belief to say, given the aforementioned advances made in computing technology and those still yet to be made, that a robot would be unable to be even more adept at this in the near future. Moreover, HRW reaches this conclusion as if the AI, who in their example is analyzing a situation of a mother running to her children playing with toy guns near a soldier, was examining the situation in a vacuum. Instead, it is much more likely the AI would analyze every aspect of the situation – in a manner likely quicker than that of a human – and determine from various bioelectric, body language, verbal language, and general situational elements that the event is non-threatening. HRW assesses the situation as if the AI was rudimentary and incapable of differentiating between threatening behavior and toy or real guns. Our scenario of Johnny Five equips him with the requisite capability to assess these variables. Accordingly, our AI would be able to satisfy the principle of distinction.

Proportionality is a principle which requires the effect on military targets to be in proper proportion to the effect on civilian targets or objects.[78] Indeed, per Article 51(5)(b) of the Protocols Additional to the Geneva Convention, "An attack which may be expected to cause incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination

---

[76] HUMAN RIGHTS WATCH - INT'L. HUMAN RIGHTS CLINIC, *supra* note 65, at 31.

[77] Bill Christensen, *Vacuum Cleaner Senses Human Emotions*, LIVESCIENCE (Mar. 28, 2009, 6:55 AM), http://www.livescience.com/5356-vacuum-cleaner-senses-human-emotions.html.

[78] Protocol Additional to the Geneva Conventions of 12 August 1949, *supra* note 72, art. 51(5)(b).

thereof, which would be excessive in relation to the concrete and direct military advantage anticipated" would violate this principle.[79] However this principle is only applicable when the attack has a likelihood of affecting civilians; a purely military target does not require a proportionality analysis.[80] This principle frequently comes into discussion regarding collateral damage, but it is sure to account for incidental loss of civilian life or property so long as such loss is minimal. The law of war recognizes there may be some death, injury, and destruction during military operations – a more unfortunate paragon of armed conflict now more than ever.[81] The question a military commander must ask himself, in order to avoid a *grave breach* of the Geneva Conventions is whether such death, injury, and destruction are particularly excessive in relation to the military advantage – not whether any will occur whatsoever.

Much like their arguments against military necessity and distinction, HRW asserts an AI would not have the requisite capability to determine whether, essentially, the 'juice was worth the squeeze.' HRW contends that an AI would not have the ability to infer from various situational elements and make a subjective judgment about whether expected civilian harm outweighs the military objective or advantage.[82] This argument hinges on the notion that "a robot could not be programmed to duplicate the psychological processes in human judgment that are necessary to assess proportionality."[83] We can infer from the technology currently being developed, from advances in logic, psychology, and computing that processes at least imitative of human psychological processes can be developed in AI, particularly Strong AI. That isn't to concede that merely logic based determinations of proportionality aren't sufficient. It is very likely that determinations based on simply the number of civilians affected versus combatants rendered ineffective

---

[79] *Id.*

[80] *See id.*

[81] GRIMES ET AL., *supra* note 64, at 167.

[82] HUMAN RIGHTS WATCH - INT'L. HUMAN RIGHTS CLINIC, *supra* note 65, at 32.

[83] *Id.* at 33.

could be sufficient. Indeed, such an ability to program into soldiers precisely what is deemed proportional by a state government could provide immense consistency and prevent incorrect assessments by untrained soldiers in the field. An AI will likely have more accurate diagnostic capabilities than a human soldier when it comes to calculating a precise blast radius, the likelihood of a missed shot, and potential collateral damage as aggregated from numerous previous and test scenarios than a human soldier who may simply "go with his gut." The benefits to utilizing an AI soldier weigh in favor of his ability to employ a fair and accurate proportionality assessment – likely one more favorable to proportionality than any human soldier.

The final principle of *jus in bello* is Unnecessary Suffering. Otherwise known as Humanity, this principle is most pertinent to our discussion. The right of parties in an armed conflict to adopt means of injuring the enemy is limited.[84] This principle is focused primarily on weapons and, as set out in the Hague Convention of 1907, "it is especially forbidden . . . to employ arms, projectiles, or material calculated to cause unnecessary suffering."[85] The two primary elements of this principle are (1) a prohibition on arms that are per se calculated to cause unnecessary suffering, and (2) a prohibition on the use of otherwise lawful arms that result in unnecessary suffering.[86] Both proscriptions require a mens rea element.[87]

It is important to note that the language states "calculated to cause unnecessary suffering" as opposed simply to causing unnecessary suffering. It is a fact of life that unnecessary suffering does unfortunately occur, regardless of how much we try to prevent it. Nonetheless, it is unlikely that AI combatants will be deployed with the intent, in any manner, to create more suffering than is necessary to render an enemy combatant or objective ineffective. Rather, it is likely the case that AI combatants will be more effective at preventing

---

[84] GRIMES ET AL., *supra* note 64, at 168.

[85] *Id.*

[86] *Id.*

[87] *Id.*

unnecessary human suffering given their likely advanced diagnostic techniques, senses, tools, and other capabilities. An AI may also be a more efficient and humane killer. It may be able to do with one bullet what a human would do with several, thus preventing unnecessary suffering, even though the ultimate goal of putting the combatant out of action is still the same. It is especially important to weigh the opposite end of the spectrum as well – that the utilization of AI combatants could prevent the unnecessary suffering of human combatants who would have fought in their stead.

HRW argues the principle of Humanity via the Martens Clause, which essentially states that warfare should be evaluated according to the public conscience.[88] According to HRW, because surveys demonstrate more than fifty percent of people found the "taking [of] life by an autonomous robot in open warfare and covert operations" objectionable, such conduct should be unacceptable per the Martens Clause and the Humanity principle.[89] This conclusion is marred by the information, or lack thereof, likely provided to the participants – evidence apparent in the "autonomous robot" nomenclature utilized. Perhaps had the participants been informed of the capabilities of a Strong AI robot, or more importantly, the harm prevented to friends and families by the employment of such technology, their opinions might have been altered.

Because HRW provides no information as to the parameters of the survey, we can only surmise that the information provided to the participants was that of their own minds – minds which are influenced by today's media and movies. Surely allowing an individual's assessment to be based off what they know from films like *The Terminator, 2001: A Space Odyssey*, *The Matrix,* and other works which portray AI in a negative, hostile, and narcissistic light would result in an objection. Throughout their paper, HRW oversimplifies the process and time it would take to actually implement AI in war. Senate Hearings, House Committees, national debates, election campaigns, news articles, numerous rounds of testing, and informative videos and pamphlets would all surely come to pass

[88] HUMAN RIGHTS WATCH - INT'L. HUMAN RIGHTS CLINIC, *supra* note 65, at 35.

[89] *Id.* at 36.

before a single AI, like Johnny Five, was put to use in the field. During that time, the public would have significantly more information than the survey group had in HRW's survey and would not be basing their opinion off flights of fancy and dystopian fictional futures that are but one of numerous possibilities of AI utilization.

Nonetheless, HRW's argument still misses the mark. Per the Hague Convention of 1907, the utilization of AI robots would not be calculated to cause unnecessary suffering, but rather to ameliorate it. On this basis, these combatants should not be excluded per se. As discussed above, AI would not cause unnecessary suffering and would likely cause less suffering than human actors, thus they cannot be said to be "calculated to cause" or "to result in" unnecessary suffering. Even if the legality of the utilization of AI in war was based off public sentiment, it is likely that a public educated on the capabilities, potential, and ramifications of AI would not be nearly as opposed.

For these aforementioned reasons, an AI robot, specifically of the Strong AI type, should not be per se excluded from utilization in armed conflict under any of the four key principles of *jus in bello*: Military Necessity, Distinction, Proportionality, or Humanity. Indeed, for the same reasons it appears almost a necessity to employ Johnny Five given his capabilities, rather than to prohibit his use.

## IV. LIABILITY OF CONVENTIONAL HUMAN ACTORS DURING TIMES OF WAR

Before I turn to liability of an actor for his commission of a war crime, I must first do as the machines do – set out precisely what the parameters are under which liability shall be assessed. Foremost, I consider what precisely is a war crime and how and who is held liable for their commission.

War crimes have been defined as "such hostile or other acts of soldiers or other individuals as may be punished by the enemy on capture of the offenders."[90] There are, however, numerous definitions because neither the words "war" nor

[90] GRIMES ET AL., *supra* note 64, at 206.

"crime" has a single definition.[91] The term "war crime" has become the technical expression for any violation of the law of war, by any person or persons, be they military or civilian; every violation of the law of war is a war crime.[92]

Crimes against humanity on the other hand are those inhumane acts which are essentially *in flagrante delicto* to the entire international community and humanity at large.[93] They are committed as part of a widespread or systematic attack on civilian population.[94] Crimes against humanity typically involve a civilian population, which is targeted due to some distinguishing characteristic; civilians are attacked due to national, ethnic, racial, political, or religious discrimination.[95] The requirement that the attack must be systematic or widespread addresses the number of victims or organized nature of the attacks.[96] Moreover, the accused must know of the attack and that his acts are part of such an attack or may further that attack.[97] For this reason, I shall not be considering crimes against humanity in my analysis, because it is unlikely Johnny Five was part of a massive robot conspiracy to overtake the world or commit genocide.[98] Nonetheless, it is important to note the differences between war crimes and crimes against humanity. "First, war crimes require an armed conflict, whereas crimes against humanity do not. Second, war crimes do not

---

[91] *Id.*

[92] *Id.*

[93] *Id.* at 209-10.

[94] *Id.*

[95] *Id.* at 216. The requirement of general persecution was unique to one of the ad hoc tribunals – it is not in the Nuremberg Charter or the Rome Statute, although persecution is one kind of proscribed "act" in the Rome Statute. *See* E-mail from Roger Clark, Bd. of Governors Professor, Rutgers University School of Law – Camden, to author (Apr. 23, 2014, 15:22 EST) (on file with author).

[96] GRIMES ET AL., *supra* note 64, at 216.

[97] *Id.*

[98] *But see* THE TERMINATOR (Orion Pictures 1984); TERMINATOR 2: JUDGMENT DAY (TriStar Pictures 1991); TERMINATOR 3: RISE OF THE MACHINES (Warner Bros. 2003); TERMINATOR: SALVATION (The Halcyon Company, 2009).

require a connection to a widespread or systematic attack. Finally, war crimes are a broader category of offenses, some of which could be the underlying foundational offense for a crime against humanity."[99]

With war crimes, there are two types: "grave" and "simple" breaches.[100] Grave breaches are the most serious felonies and include: (1) willful killing; (2) torture or inhumane treatment; (3) biological experiments; (4) willfully causing great suffering or serious injury to body or health; (5) taking of hostages; or (6) extensive destruction of property not justified by military necessity.[101] Simple breaches,[102] include things such as: (1) treacherous request for quarter; (2) firing on localities which are undefended and without military significance; (3) killing without trial; and (4) spies or other persons who have committed hostile acts. Specific to our discussion, are the simple breaches of the proportionality principle; attacking or bombarding towns or villages, which are undefended and are not military objectives; and violations of the principle of humanity, including the "treacherous killing or wounding of individuals" belonging to the enemy nation or army.[103]

These are not crimes that are borne simply on the person pulling the trigger, but also on the commanders responsible for their criminal subordinates. Commanders may be held liable for the criminal acts of their subordinates even if the commander did not personally participate in the underlying offenses, provided certain criteria are met.[104] Primarily, the commander's

---

[99] GRIMES ET AL., *supra* note 64, at 216.

[100] *Id.* at 208.

[101] *Id.*; *see also* Rome Statute of the International Criminal Court, art. 8 para. 2(a), U.N. Doc A/CONF.183/9 (July 1, 2002), *available at* http://www.icc-cpi.int/nr/rdonlyres/ea9aeff7-5752-4f84-be94-0a655eb30e16/0/rome_statute_english.pdf.

[102] GRIMES ET AL., *supra* note 64, at 208-09. Take note that "Simple Breach" is a bit of misnomer. These breaches are still considered serious offenses per the Rome Statute of the ICC. *See* Rome Statute of the International Criminal Court, *supra* note 101, para. 2(b).

[103] *Id.*

[104] GRIMES ET AL., *supra* note 64, at 218.

personal negligence must have contributed to, or failed to prevent the offense; thus a commander is not strictly liable for crimes committed by subordinates.[105]

One of the seminal cases of command responsibility is that of General Tomoyuki Yamashita who was convicted and sentenced to hang for war crimes committed by his soldiers in the Philippines.[106] There was no evidence of General Yamashita's direct participation in the crimes; however, the tribunal determined that the violations of the rules of war were so widespread that he had to have a hand in ordering their commission or otherwise failed to discover and control them.[107] The Court came to this conclusion even though Yamashita had been cut off from various regiments of his forces due to American military attacks and disruptions.[108] Most scholars have concluded *Yamashita* stands for the proposition that "where a commander knew or should have known that his subordinates were involved in war crimes, the commander may be liable if he did not take reasonable or necessary action to prevent the crimes."[109]

The International Criminal Court has expanded upon this notion of command responsibility. Per the Rome Statute, a superior can be held criminally liable when:

> The superior either knew, or consciously disregarded information which clearly indicated, that the subordinates were committing or about to commit such crimes, the crimes concerned activities that were within the effective responsibility and control of the superior, and the superior failed to take all necessary and reasonable measures within his or her power to prevent or repress their commission or to submit the matter

---

[105] *Id.*

[106] *Id.*

[107] *Id.*

[108] *Id.*; *see generally In re* Yamashita, 327 U.S. 1 (1946) (Murphy, J., dissenting).

[109] GRIMES ET AL., *supra* note 64, at 218.

> to the competent authorities for investigation and prosecution.[110]

More specifically, Article 28 of the Rome Statute has two particularly pertinent paragraphs. One paragraph deals with the responsibility of military superiors – paragraph (a) – and the other applies to other superiors, such as cabinet ministers – paragraph (b). Paragraph (a) has a negligence standard where "should have known" is sufficient to find culpability.[111] Paragraph (b) however, has what would be defined under the Model Penal Code as a recklessness standard, where one knew or consciously disregarded the occurrence of the commission of crimes.[112] Nonetheless, the Tokyo Tribunal in the Rape of Nanking tragedy applied a negligence standard to both military and civilian leaders.[113] So it would appear that the lower negligence standard reigns supreme in most command responsibility cases, especially those with truly egregious transgressions at play, like the Rape of Nanking.[114]

However, the doctrine of command responsibility has it limits. In a case known as the High Command Trial, which was prosecuted in Germany after World War II, the court stated:

> A high commander cannot keep completely informed of the details of military operations of subordinates . . . . He has the right to assume that

---

[110] Rome Statute of the International Criminal Court, *supra* note 101, art. 28 para. (b).

[111] *See* E-mail from Roger Clark, *supra* note 95.

[112] *Id.*

[113] *Id.*

[114] There is, however, another formulation of superior or command responsibility. In Protocol 1, Article 86, the statute talks about not "absolving" superiors, which seems to assume some sort of strict liability, but then goes on to speak about responsibility, "if they knew, or had information which should have enabled them to conclude in the circumstances . . . ." *See id.*; *see also* Protocol Additional to the Geneva Conventions of 12 August 1949, *supra* note 72, art. 86. This is not quite recklessness or negligence, and so the decision was ultimately made in the Rome negotiations to avoid it as many thought it was confusing. *See* E-mail from Roger Clark, *supra* note 95.

> details entrusted to responsible subordinates will be legally executed . . . . There must be personal dereliction. That can only occur where the act is traceable to him or where his failure to properly supervise his subordinates constitutes criminal negligence on his part. In the latter case, it must be a personal neglect amounting to a wanton, immoral disregard of the action of his subordinates amounting to acquiescence. Any other interpretation of international law would go far beyond the basic principles of criminal law as known to civilized nations.[115]

At first glance, this ruling seems inapposite to the ruling in *Yamashita*, yet they came out during the same time period. However, the key similarity between the two cases is that they still require the commander to meet the mens rea element of knowledge. This latter case merely seems to stand for the notion that liability can only travel up so far the chain of command before it is no longer applicable.

This mens rea element is exemplified in the Department of the Army Field Manual on The Law of Land Warfare ("MLLW"). A commander is only responsible, in instances, when: (1) he has ordered the commission of the crime; (2) has actual knowledge, or should have knowledge that persons subject to his control are about to commit or have committed a war crime; and (3) he fails to take reasonable steps to insure the law of war is upheld.[116] The Army's knowledge requirement can be demonstrated through reports received by the commander or through other unspecified means.[117] Of course, knowledge through reports made by others is secondary to instances where

---

[115] GRIMES ET AL., *supra* note 64, at 219; *see also* 7 UNITED NATIONS WAR CRIMES COMM'N, LAW REPORTS OF TRIALS OF WAR CRIMINALS-THE GERMAN HIGH COMMAND TRIAL 76 (1948), *available at* http://www.loc.gov/rr/frd/Military_Law/pdf/Law-Reports_Vol-12.pdf.

[116] GRIMES ET AL., *supra* note 64, at 222 (citing DEP'T OF THE ARMY, FIELD MANUAL 27-10: THE LAW OF LAND WARFARE 178-79 (1956)).

[117] *Id.*

the commander in question gives an order in direct violation of the laws of war.[118]

An illustrative example of the doctrine of command responsibility is the infamous My Lai incident. The incident, more commonly referred to as the My Lai Massacre, involved the ruthless slaughter of the village of My Lai by American soldiers in Vietnam.[119] A central question concerning the incident was whether the soldiers were acting on orders of their commander, Captain Medina.[120] At that time, much consideration was given to Article 77 of the Uniform Code of Military Justice (UCMJ), which required that the non-participant share in the perpetrator's purpose of design and "assist, encourage, advise, instigate, counsel, command, or procure another to commit, or assist . . . ."[121] The court instructed that the panel would have to find that Captain Medina had actual knowledge, as opposed to constructive knowledge, of the actions of his subordinates in order to hold him criminally liable.[122]

The only instance of command responsibility and a possible malfunction of software causing the loss of life is the Iran Air Flight 655 incident. On July 3, 1988, the USS

---

[118] DEP'T OF THE ARMY, *supra* note 116, at 178.

[119] *See Vietnam Online- The My Lai Massacre*, PBS (Mar. 29, 2005), http://www.pbs.org/wgbh/amex/vietnam/trenches/my_lai.html. On March 16, 1968, Charlie Company, 11th Brigade, commanded by Lieutenant William Calley entered the village of My Lai on a search and destroy mission, ready for a firefight. *Id.* Upon Lt. Calley's orders they entered the village firing even though no enemy fire had been received at that time. *Id.* It quickly devolved into the brutal massacre of 300 men, women, and children and added fuel to the fire that was the debate over the U.S.'s involvement in the Vietnam War. *Id.*

[120] *Id.*

[121] GRIMES ET AL., *supra* note 64, at 221. The current Article 77 of the UCMJ has been reformulated to state that a person is punishable if they commit an offense, aid, abet, counsel, command, or procure its commission or otherwise cause an act to be done which if directly performed by him would be punishable. *See* 10 U.S.C. § 877 (2012).

[122] GRIMES ET AL., *supra* note 64, at 221. Note that many disagree with the way Medina's court martial was handled. *See, e.g.,* Roger S. Clark, *Medina: An Essay on the Principles of Criminal Liability for Homicide*, 5 RUTGERS L.J. 59 (1973).

Vincennes, a Ticonderoga class guided missile cruiser equipped with the Aegis combat system, mistakenly shot down Iran Air Flight 655, killing all 290 people on board.[123] The Vincennes and her crew, after engaging Iranian boats, mistakenly believed Iran Air Flight 655 was an Iranian F-14 Tomcat fighter jet on hostile approach.[124] The Aegis system employed on the Vincennes was – and still is – a centralized, automated, command and control weapon system that is designed to work from detection to kill.[125] An automatic detect and track feature is included in the system, which is ancillary to the command and decision element at the core of the system.[126] This system, along with the crew of the USS Vincennes, failed to distinguish the Airbus commercial airliner from an F-14 Tomcat based upon its profile.[127] Moreover, it also failed to ascertain the Identification Friend or Foe ("IFF") communication, which would have distinguished a commercial jet aircraft from a military aircraft.[128] Among other reasons, but largely because of these failures in the system, the Vincennes fired upon and destroyed Iran Air Flight 655. No formal action within the military justice system of the United States was brought against the captain or crew of the Vincennes, nor was there any mention made about the responsibility of the Aegis combat system. The only action

---

[123] *See* Max Fisher, *The Forgotten Story of Iran Air Flight 655*, WASH. POST (Oct. 16, 2013 7:00 AM), http://www.washingtonpost.com/blogs/worldviews/wp/2013/10/16/the-forgotten-story-of-iran-air-flight-655/; *see also* George C. Wilson, *Navy Missile Downs Iranian Jetliner*, WASH. POST (July 4, 1988), http://www.washingtonpost.com/wp-srv/inatl/longterm/flight801/stories/july88crash.htm; Shapour Ghasemi, *Shooting Down Iran Air Flight 655 [IR655],* IRAN CHAMBER SOC'Y (2004), http://www.iranchamber.com/history/articles/shootingdown_iranair_flight655.php.

[124] Ghasemi, *supra* note 123.

[125] *United States Navy Fact File: Aegis Weapon System*, AM.'S NAVY, http://www.navy.mil/navydata/fact_display.asp?cid=2100&tid=200&ct=2 (last updated Nov. 22, 2013).

[126] *Id.*

[127] *See* Wilson, *supra* note 123; *see also* Ghasemi, *supra* note 123.

[128] *Id.*

came from a civil suit brought against the United States.[129] This speaks to the issues presented here, primarily because the Aegis system contained a decision system, allowing it to make conclusions and demonstrate some rudimentary logic – the very beginnings of AI. The fact that nothing was mentioned with regard to the liability of the Aegis system supports the contention that it was but a mere weapon, even if the crew of the ship relied upon its representations and decisions. The responsibility remained solely with the human commanders.

Likewise, the same can be said for current drone strikes where civilians are mistakenly killed. No mention is made of product liability or command responsibility for these machines as actors, but instead for those piloting or deploying those machines.[130] Although these machines are autonomous and not AI powered, the idea of holding a machine responsible is not in the minds of those looking for justice.[131]

Command responsibility has even come into discussion in civilian Article III courts. In *Chavez v. Carranz*, the Sixth Circuit applied a three-factor test for command responsibility when determining whether an El-Salvadorian military commander was responsible for the crimes of torture, extrajudicial killing, and crimes against humanity in his tenure as El-Salvador's Vice Minister of Defense and Public Security.[132] The Court required: (1) a superior-subordinate relationship between the commander and the persons who committed the human rights abuses; (2) the commander knew or should have known that subordinates had committed, were in the process of committing, or were about to commit human rights abuses; and (3) the commander failed to take all reasonable measures to prevent human rights abuses and punish human rights

[129] *See generally* Koohi v. United States, 976 F.2d 1328 (9th Cir. 1992).

[130] *See, e.g.,* Ewen MacAskill & Owen Bowcott, *UN Report Calls for Independent Investigations of Drone Attacks*, THE GUARDIAN (Mar. 10, 2014 11:16 AM), http://www.theguardian.com/world/2014/mar/10/un-report-independent-investigations-drone-attacks.

[131] Or motherboards, as it were.

[132] Chavez v. Carranza, 559 F.3d 486 (6th Cir. 2009).

abusers.[133] Thus, it is evident that even civilian courts follow the contention presented in *Yamashita* and the MLLW.

The doctrine of command responsibility bears resemblance to another doctrine more common to civil litigation: *respondeat superior*. After all, command responsibility is essentially an issue of agency – there is no better corollary to draw from than that of "let the master answer." The Restatement Third of Agency defines agency as "the fiduciary relationship that arises when one person (a 'principal') manifests assent to another person (an 'agent') that the agent shall act on the principal's control, and the agent manifests assent or otherwise consents so to act."[134] Lest we forget that a soldier is ultimately an employee who is subordinate of his commander, and by virtue of his enlistment or voluntary commission, consents to act as his senior orders him. Thus, it is naturally axiomatic that a soldier is an agent of his commander.

*Respondeat superior* is defined as an employer being subject to liability for torts committed by employers acting within the scope of their employment.[135] Similar to command responsibility, *respondeat superior* "[m]ost often . . . applies to acts that have not been specifically directed by an employer but that are the consequence of inattentiveness and poor judgment on the part of an employee . . . ." [136] Liability is essentially "strict" as far as the employer is concerned.[137]

Though *respondeat superior* is a civil law invention, its concerns and overarching ideas are worth considering in the context of command responsibility. Indeed, one can see the similarity where a supervisor is essentially liable through preventative or retributive measures for not properly supervising or training his subordinate. Much like *respondeat superior*, command responsibility creates an incentive to choose and train subordinates and structure work within the command

---

[133] *Id.* at 499 (citing Ford v. Garcia, 289 F.3d 1283, 1288 (11th Cir. 2002)).

[134] RESTATEMENT (THIRD) OF AGENCY § 1.01 (2006).

[135] *Id.* § 2.04.

[136] *Id.* § 2.04, cmt. b.

[137] E-mail from Roger Clark, *supra* note 95.

hierarchy to reduce the incidence of misconduct.[138] Ergo, *respondeat superior* bears mentioning in our discussion and should be considered in determining whether Johnny Five or our Army Captain is liable for the deaths of those villagers.

What these different cases and theories show is an uneven application of the doctrine of command responsibility by U.S. courts and military tribunals since the Second World War. As previously described: *Medina* required actual knowledge; *Yamashita*, the MLLW, and *Chavez* created a standard of "knew or should have known;" *High Command* goes further, requiring a personal dereliction on the account of the commander; Article 28 paragraph (a) of the Rome Statute requires knowledge; while Article 28 paragraph (b) requires conscious disregard. *Respondeat superior* provides that the mere utilization of a subordinate agent who commits misconduct suffices to establish responsibility on the part of the superior or even the organization as a whole. Given that the "knew or should have known" standard appears to be the most commonly embraced, I shall analyze Johnny Five's and the Captain's responsibility primarily under this standard, but the other tests I shall examine are prudent.

## V. WHO THEN IS LIABLE FOR JOHNNY FIVE?

These legal standards and statutes affect Johnny Five and his situation in two ways: first, in determining whether Johnny Five himself would be subject to the UCMJ or the Rome Statute, and second, in discerning whether a captain, as Johnny Five's commander, would be responsible for his actions.

### A. DO THE LAWS APPLY TO JOHNNY FIVE?

The UCMJ is broad in its jurisdiction over the armed services both in statute and in common law. Article 2 of the UCMJ states, inter alia, persons subject to the UCMJ include "members of a regular component of the armed forces,"[139] "other persons lawfully called or ordered into or to duty in or

[138] *See* RESTATEMENT (THIRD) OF AGENCY § 2.04, cmt. b (2006).

[139] 10 U.S.C. § 802(a)(1) (2009).

training for in, the armed forces,"[140] "[i]n time of declared war or a contingency operation, persons serving with or accompanying an armed force in the field,"[141] or "persons serving with, employed by, or accompanying the armed forces."[142] Furthermore, Article 3 states the military has the jurisdiction to try personnel who are or were at the time of the act in question a status to be subject to the UCMJ. In other words, the UCMJ's jurisdiction extends to members of the armed forces or other persons encompassed by Article 2 at the time the act in question took place.[143]

Essentially, the UCMJ applies to any person within or accompanying the armed forces. Johnny Five might think he is able to get away scot-free since he is not necessarily a person, but that is not the case. While the UCMJ does not expound upon the meaning of "person", the United States Code in its very first provision certainly does. It provides "[i]n determining the meaning of any Act of Congress, unless the context indicates otherwise . . . the words 'person' and 'whoever' include corporations, companies, associations, firms, partnerships, societies, and joint stock companies, as well as individuals."[144] It would be no different to give the same rights to an AI being as those conferred on corporations; both AI persons and corporations would be legal fictions.

Johnny Five can't be said to be any less of a person than a corporation. In fact, because he is an individual with cognitive and communicative abilities, he is more so a person than any corporation. At the very least, if a corporation can be considered a person and is defined as such per the United States Code, with

---

[140] *Id.*

[141] *Id.* § 802(a)(10).

[142] *Id.* § 802(a)(11).

[143] *Id.* § 803(a); *see* Solorio v. United States, 483 U.S. 435 (1987) (holding that subject matter jurisdiction of military tribunals is dependent upon the status of the accused as a member of the armed services). *But cf.* O'Callahan v. Parker, 395 U.S. 258 (1969) (holding that the jurisdiction of military tribunals was dependent upon the action in question being "service-connected", that is, the action had to be related to the person's duties in his capacity as a member of the armed services), *overruled by*, *Solorio*, 483 U.S. 435.

[144] 1 U.S.C. § 1 (2012).

nothing else to the contrary in the UCMJ, he should be subject to the articles of the UCMJ and the jurisdiction of military tribunals.[145]

Likewise, Johnny Five should be considered a person in any other criminal proceeding domestically or internationally because the meaning of person has long been understood to include inanimate objects. While "person" is typically understood to mean an actual human being, a legal person is anything that is subject to rights and duties.[146] So long as an inanimate object is the subject of a legal right, the will of a human is attributed to it in order for the right to be exercised.[147] Surprisingly, this is not a new theory. Legal proceedings against inanimate objects have been in existence since ancient Greece and in fact continue until this day, albeit infrequently. In Greece, proceedings against inanimate objects were almost commonplace.[148] Such objects were often referred to as deodands and, in England as late as 1842, these items were forfeited to the church or Crown.[149] In fact, anything that had killed a man, such as an errant locomotive, was liable to be forfeited.[150] For killing those people in our scenario, Johnny Five then would be liable for those deaths and subject to forfeit under those rules – rules which have been around for thousands of years and should go undisturbed or, at the very least, provide example for the discussion of how to treat Johnny Five in a potential war crime scenario.

---

[145] This assumes that the convening authority for the court martial of Johnny Five does in fact convene the court martial per Article 16 or recommend some other non-judicial punishment per Article 15 of an AI instead of letting the issue fall by the wayside and simply deactivating the machine. 10 U.S.C. §§ 815-816.

[146] JOHN CHIPMAN GRAY, THE NATURE AND SOURCES OF THE LAW 27 (1909); *see also* Lawrence B. Solum, *Legal Personhood for Artificial Intelligences*, 70 N.C. L. REV. 1231, 1239-40 (1992).

[147] GRAY, *supra* note 146, at 46.

[148] *Id.* at 47.

[149] *Id.*

[150] *Id.* at 46.

This conception of liability of inanimate objects is not one solely of the old world or other countries, but has been a staple of domestic U.S. law. There were instances in the Plymouth and Massachusetts colonies where a gun or even a boat would be forfeited for causing the death of a man.[151] Indeed, this notion of liability of objects has had the most discussion in maritime and admiralty law.[152] Oliver Wendell Holmes, Jr., in his treatise on the common law, has stated that in maritime collisions, the owner of the ship is not to blame, nor necessarily is the captain, but rather all liability is to be placed upon the vessel – freeing the owner from all personal liability.[153] Chief Justice Marshall even stated outright that proceedings in maritime law are not against the owner, but against the vessel for offenses committed by the vessel.[154] Like a vessel, Johnny Five would often be referred to with a gender – "he" for Johnny Five is no different than "she" for a sea faring ship. This attribution of gender is something unique to maritime ships and is a likely reason for this unique, if not strange, rule of liability against vessels.[155] If something as simple as gender can lead to legal person status, surely something with gender, speech, movement, logic, and appearance similar to human persons should be treated in at least the same respect.

---

[151] *Id.*

[152] *See* The China, 74 U.S. (7 Wall.) 53, 64 (1868) (after reviewing multiple cases of wrongdoing vessels at sea, the Court deduced, inter alia, the colliding vessel is in all cases prima facie responsible).

[153] OLIVER WENDELL HOLMES, JR., THE COMMON LAW 27 (1881).

[154] United States v. The Little Charles, 26 F. Cas. 979, 982 (C.C.D. Va. 1818) (No. 15,612); *see also* United States v. Brig Malek Adhel, 43 U.S. (2 How.) 210, 234 (1844). It should be noted that the purpose of this rule had foundation in the idea that on the high seas, the vessel may be the only way to secure a judgment against the at-fault party since the vessel is often of international origin. Indeed, this was a form of *in rem* jurisdiction and the likely genesis for the rule. *See* HOLMES, *supra* note 153, at 27-28. Nonetheless, the law is still relevant here since the AI, like the vessel, presents an equal conundrum in determining its owner or commander – if it even has one – or who to hold responsible generally, a problem demonstrated by this very paper. Moreover, such an AI could easily be used at sea and could, for all intents and purposes, technically be considered a vessel in which these laws discussed by Holmes, Marshall, and others would apply directly.

[155] HOLMES, *supra* note 153.

It is no stretch to relate Johnny Five to a vessel, just as it is no stretch to relate him to a corporation. Both corporations and vessels display substantially larger differences from the traditional human person than Johnny Five would, yet they are held liable and in the case of corporations, afforded certain constitutional rights.[156] Johnny Five would be able to think, speak, move, listen, make decisions, and take action all on his own. He would be tasked with the legal right to kill, capture, or otherwise deter an enemy combatant. If a legal person is anything that is subject to legal rights and duties, then because Johnny Five is tasked with not only the legal right to kill, but also the duty not to kill in certain situations, it only follows that he is a legal person. He should, like vessels and corporations before him, be considered a person for purposes of the UCMJ, Rome Statute, and any other international law he may meet.[157] Inanimate AI objects such as Johnny Five should most assuredly be legal persons.

## B. OUR SCENARIO: WHO IS RESPONSIBLE?

Accepting that Johnny Five is a person under the UCMJ and other international laws, Johnny Five would be liable for his actions. And in my scenario, because the Captain did nothing to stop Johnny Five and instead paused out of shock, the Captain too would likely be liable, provided the Captain failed to act for a

---

[156] Note though that the concept of corporate criminal responsibility is not a widely accepted custom of international law. Indeed, several countries – like Austria – do not accept corporate criminal responsibility and the default position in international law is not to include it unless there is some express provision. E-mail from Roger Clark, *supra* note 95. Interestingly, France and the Solomon Islands tried unsuccessfully to have legal persons included in the Rome Statute but ultimately failed. *Id.*

[157] Though not mentioned outright in international law, because the liability at sea is attributed to an inanimate object, and because this concept appears to be readily utilized across multiple countries, it would appear safe to conclude that it is customary international law. However, the Rome Statute specifically uses the words "natural person" in Article 25 in asserting over whom the court may have jurisdiction. *See* Rome Statute of the International Criminal Court, *supra* note 101, art. 25. This would directly exclude Johnny Five from the jurisdiction of the ICC. Nonetheless, the theories the ICC would employ under the Rome Statute still give significant fguidance as to how to assess the situation.

sufficient period. Moreover, if the Captain did nothing and this happened again, the Captain would be even more likely to be held responsible, as the Captain had knowledge that Johnny Five had certain proclivities to violate the rules of war.

This result is admittedly hard to swallow. After all, if Johnny Five is held liable, what good is it to actually punish him? Putting an expensive piece of machinery, albeit a thinking, speaking, and moving piece of machinery, behind bars seems ineffective. The deprivation of freedom and time may not mean the same to Johnny Five as it would to a human actor. True, he can think and understand he is being punished, and potentially even feel sorry for his actions, but what does twenty years in prison mean to a machine that may not have a life span? Similarly, if the point of punishment is to be a deterrent to others, does putting Johnny Five behind bars truly deter other AIs in combat from doing the same thing? Granted, these are questions potentially already posed by the retributive criminal justice system as a whole, but perhaps ones that may be more easily quelled in the instance of a machine as opposed to a human.

Perhaps the simple solution is to shut him down and repurpose or reconfigure him. However, does one run into significant hurdles when they do this to something that is, for all intents and purposes, human but for the organic component? Though we may develop bonds or affection towards our AI counterparts as if they were our brothers, the ultimate reality that they are machines will never fade. No matter how similar in appearance they become, no matter how identical they are in mannerisms and speech, or how friendly they may be, the notion that they are naught but metal and plastic will never truly be overcome.[158] Ultimately, punishment will simply have to be crafted to Johnny Five and will likely entail reconfiguration or decommissioning.

Regardless, procedures such as court martials and military commissions or tribunals can and should still be employed. They can be employed for the reasons mentioned above, that is, an AI could qualify as a "person" and therefore be

[158] This however is plagued by scenarios in which we begin integrating inorganic components into other humans, whether they are prostheses or augmentations. The line becomes blurred as we become closer and closer to our Johnny Five counterparts.

subject to the UCMJ and other courts. They should be employed because an AI who can think and feel should be afforded the same rights as their human counterparts, at least in terms of due process. It would be easy for a group of soldiers to commit a war crime, blame an AI, and have the AI simply shut down while they escape scot-free. For the very necessity of ascertaining the facts of any situation, proceedings should be held. That these proceedings ultimately end in different punishments have little effect on their necessity.[159]

Setting aside the rationale behind holding a robot liable and applying punishment, let us look at why both Johnny Five and the Captain may be held responsible. First, Johnny Five's actions are reminiscent of the My Lai incident and are per se in violation of the rules of war. Namely, Johnny Five's actions violate three of the four principles of *jus in bello*: military necessity, proportionality, and distinction. The fourth principle of *jus in bello*, humanity, is not triggered by this because Johnny Five was not calculated to cause unnecessary suffering, nor had his actions prior to this instance resulted in unnecessary suffering. However, if this instance continued or if a ban was suggested on AI after this instance, one could assert a violation of the humanity principle, citing this incident as reason enough.

There was no military necessity, and there is never any military necessity, in willfully killing unarmed noncombatants. There was no concrete and direct military advantage calculated by the murder of these innocent civilians. And of course, there was no distinction made here; attacks directed solely at civilians are in direct conflict with the rule of distinction since the only distinction made was to kill civilians. Thus, we can be sure Johnny Five's actions were in violation of the rules of war, just as the actions of Lieutenant Calley and his men were in My Lai.

But how alike is My Lai to this incident? A stark contrast to My Lai is that the Captain did not order Johnny Five to murder those people as Lt. Calley ordered his men to kill the villagers of My Lai. However, like Lt. Calley, the Captain did nothing to stop the slaughter of a whole village.[160] Looking to

---

[159] This quickly glances over the discussion about whether AI should be afforded civil rights. If they should, then of course military tribunals should occur. *See generally* Solum, *supra* note 146.

[160] To be sure, had Johnny Five acted on his own and our Captain immediately ordered him to stop, the Captain would not be held liable. That is,

the *Yamashita* and *Chavez* standards, so long as the Captain knew or should have known of Johnny Five's actions, he can and will be held liable. Here, he knew by watching through the scope what Johnny Five was doing and, like Yamashita himself, he did not take reasonable or necessary steps to prevent the murders – or rather to prevent them from continuing to occur after he became aware. Similarly, under the Article 77 and the *Medina* standard, the Captain had actual knowledge and would be liable. The same result occurs under the Rome Statute, albeit by a moderately different analysis, as he both had knowledge and, it can be argued by willful inaction, consciously disregarded the massacre that was taking place. Looking next to the *High Command* case, we may run into a bit of a kerfuffle. If a commander is not responsible but for cases where there is a personal dereliction on his part, does the Captain's failure to act create responsibility for Johnny Five? It most certainly does. After all, this exact scenario is almost perfectly fit into the *High Command* court's decision. The Captain's inaction – depending throughout this analysis on precisely how long Johnny Five went on killing, that is minutes as opposed to mere seconds – certainly amounts to a personal dereliction, and is tantamount to criminal negligence. He had actual knowledge of what was occurring and failed to do anything.

If, however, we were to utilize the civil doctrine of *respondeat superior*, not only is our Captain potentially liable, but so is the United States as a whole, barring of course some sovereign immunity. Because the U.S. decided to employ the AI in combat, the deaths were ultimately a result of their negligent utilization of this technology, and so they should be made to pay reparations, much like they ultimately did in the Iran Air Flight 655 incident.[161]

Nonetheless, our Captain is stuck holding the proverbial smoking gun here on one level or another and will be punished for an error a hulking bit of metal committed. This is an

---

unless some strict liability regime was put into place in this future of ours where the commanding officer is liable for any acts of his AI subordinate – an unfair but certainly possible scenario.

[161] INT'L COURT OF JUSTICE, SETTLEMENT AGREEMENT ON THE CASE CONCERNING THE AERIAL INCIDENT OF 3 JULY 1988, para. 1 (1996), *available at* http://www.icj-cij.org/docket/files/79/11131.pdf.

unfortunate, but ultimately correct, result under the current regime of command responsibility.

## VI. A PROPOSAL FOR LIABILITY OF JOHNNY FIVE AND HIS COMMANDERS

True, these outcomes would be the same if Johnny Five was human, but that is entirely the point. An AI with exactly the same faculties and characteristics as a human, but merely inorganic sinews, still acts, decides, and exists as a human does. Still, this seems an awfully strange outcome; it does not sit right. Perhaps instead we should look to other parties to hold liable in addition to, or in place of, the poor Captain.

The manufacturer of Johnny Five is one party to look to for responsibility. Along a sort of products liability theory, comingled with command responsibility, we can ascertain that the manufacturer was apparently negligent in its creation, programming, wiring, or other technique used to make Johnny Five alive. But while this seems an obvious conclusion when you consider Johnny Five as a machine, it becomes a much more difficult conclusion to draw when you realize he can see, think, act, speak, and move just like a person can. In that instance, the situation seems less like returning a faulty washing machine to its manufacturer and more similar to punishing the mother for sins of her son. If the manufacturer is creating something that is essentially an inorganic human, how do we hold them responsible for the acts of this now artificially independent being? It may be that we consider the AI much like an adolescent child and as an agent of the manufacturer. Perhaps in this instance it provides incentive for the creator to construct Johnny Five with the utmost care and, in a limitation on Johnny Five's free will, hardwire him with specific directions. The trouble is when you hardwire those directions in him, there is almost always going to be a situation you cannot account for which circumvents the hardwiring, or perhaps one that allows a maligned soldier to order Johnny Five to slaughter the villagers of My Lai. The question is, can this be corrected? And are we amenable to limiting something's free will?

What about Johnny Five as solely a weapon? If Johnny Five was an MI6, and the Captain was the soldier pulling the trigger, the Captain and not the weapon is the one responsible

for what damage that MI6 does. But Johnny Five is not a mere weapon. If he were a mindless automaton programmed to respond to any order, he, perhaps, could be just a weapon. Instead, he has the ability to think and decide like a human does. And like humans, every so often the wiring is not quite right. It is true that in this scenario, the Captain could have intervened with an order – he is still at fault for not at least trying to stop the atrocity. If this situation were different, though, as where the Captain sends out an entire independent platoon of Johnny Fives and they commit a My Lai sort of atrocity, can we say he pulled the trigger there? Surely not. And he is surely not liable under any regime of command responsibility, barring some previous event that occurred with these particular AI robots. He would not have actual or constructive knowledge of their actions unless he gave the specific order to wipe out the entire village, which for our purposes is not the case. It is the perfect crime. Yes, the robots themselves could be decommissioned, destroyed, confined, or otherwise, but no human actor is responsible. This too seems an unfitting result; the death of 100 civilians and not one human person to blame is unsettling.

Foremost, it is undoubted in my mind that the robots – either the lone Johnny Five or platoon Johnny Five – should all be put out of service, an option which, while potentially harsh when you consider the bonds that may be created with an inorganic human, is the only sure way to prevent them from doing the same thing again. The death penalty for robots is hardly a capital offense, regardless of the humanity of the machine. A machine that cannot "die" just as it cannot be "born," should not be allowed the same trepidation in enacting an end of life sentence as a human would. This seems the only logical outcome for the machine actor.

It seems the only answer here then is *respondeat superior*, or rather a slightly modified version. If *respondeat superior* translates to "let the master answer," then our version for Johnny Five shall be "let the creator answer and be careful." For policy concerns, the best bet to ensure Johnny Five is created with the utmost care and properly constructed is to place the burden for ensuring these machines do not commit war crimes with the manufacturers. While we can punish the derelict commanders and dispose of or repurpose the miscreant machines, the reparations should be paid not by the country

utilizing the machines, but rather the creators. This puts significant pressure on the manufacturers and designers to create the soundest ethical machines for not only financial but also publicity purposes. No manufacturer would want to be known as the one who creates baby killers or women slaughterers. Ergo, in this version of *respondeat superior*, the creator is responsible for the actions of his creations as opposed to the employer responsible for the actions of his employees.

This doctrine, to be accompanied by the command responsibility and normal criminal liability imposed on the actor, provides for complete criminal and civil remedies. Most importantly though, it solves the issue of deterrence. If decommissioning, imprisoning, or repurposing an AI does not deter others from acting, by nipping the problem in the bud we can deter other manufacturers from producing equally malfeasant AI.

Understanding that much of this paper has focused on the humanization of AI and that it will be no different cognitively than a human, the fact that the AI is still *not* human is most salient. As close to human as it can get, it will still never be human, and for that reason we must hold its creators liable. Only they can install specific parameters within the mind of the AI, which should never be broken.[162] True, there will always be a potential scenario that allows the rules to be obeyed, and still a grizzly outcome may ensue, but machines that can conceptualize, deduce, and understand from a single or multiple hardwiring, can also ascertain the correct course of action in those scenarios. This is much unlike an automaton programmed not to learn and adapt but merely to follow predetermined rules – rules that to the automaton are square pegs to fit only in square holes.

This "let the creator answer and be careful" policy provides for constant innovation and the highest level of care in an AI creator. Though incidents like our Captain and Johnny Five may occur, the impetus not only to prevent them, but also to learn from them, is there. The creator cannot merely rest on his laurels and be insulated from the misanthropy of his creations. Rather, the hope and policy behind this regime is that

[162] The answer then, to the question posed at the end of the second paragraph in this section, is that we must limit the free will of the AI. It would seem that our greatest asset is also what we fear most in our creations.

even if these events occur, the drive to improve will be so great that the learning curve for creating foolproof machines will be almost nonexistent. And in a scenario where you are employing inorganic humans with the killing power of an entire platoon, you cannot afford anything else.

While it is easy to write off this entire scenario, indeed perhaps this entire paper, as being implausible, one must remember that humans like to create in their own image and likely will continue to create until we reach the phase where machine and human are essentially carbon copies. The need for a true, strong AI might not seem prima facie evident, but when you look at the human need for companionship, creation, innovation, and to push the envelope, you can understand why Johnny Five will – not may – come to exist. What the people will want is not mindless automatons who will succumb to the will of any person and could potentially fall into the wrong hands, but human analogues who can do our dirty work, yet still have the discretion, cognitive faculties, and logic to know when to do something, how to do it, why they are doing it, and if they should do it. What the people want is humans who cannot "die." When the world realizes it can replace its soldiers, among other professions, with identical AI robots and no longer have to suffer the human casualties, this push will come and it will succeed. What the world wants, and has always wanted, is not peace, but rather a lack of human suffering; people have always thrived on destruction and war. If they can have both war and a lack of human suffering, they will strive for it.

*"I don't know why he saved my life. Maybe in those last moments he loved life more than he ever had before. Not just his life - anybody's life; my life. All he'd wanted were the same answers the rest of us want. Where did I come from? Where am I going? How long have I got? All I could do was sit there and watch him die."*

*-Deckard* [163]

[163] BLADE RUNNER, *supra* note 2.