Catholic University Journal of Law and Technology

Volume 25 | Issue 1 Article 4

2017

Manhattan_Project.exe: A Nuclear Option for the Digital Age

David Laton
Laton & Strain LLC

MANHATTAN_PROJECT.EXE: A NUCLEAR OPTION FOR THE DIGITAL AGE

David T. Laton

AN INTRODUCTION TO ARTIFICIAL INTELLIGENCE

I. Artificial Intelligence

There is no objectively simple definition of Artificial Intelligence (AI). This is because the term is often interchangeably used to refer to artificially intelligent mechanical or computer-based constructs portrayed in media as well as the study, research, and development of actual AI programs capable of performing any number of complex tasks.[1] To the latter example, "[AI] research is concerned with constructing machines (usually programs for general-purpose computers) which exhibit behavior such that, if it were observed in human activity, we would deign to label the behavior 'intelligent.'"[2] Notable examples include IBM's Watson program and Deep Blue chess-playing program.[3] In *Artificial Intelligence: A Modern Approach*, four distinct variations of AI are offered as definitions. These include programs designed to think like humans, programs designed to think rationally, programs designed to act like humans, and programs designed to act rationally.[4]

AI programs operating within any of these definitions can be further broken down into three distinct categories: Artificial Narrow Intelligence (ANI), Artificial General Intelligence (AGI), and Artificial Superintelligence (ASI).[5] AI constructed to think and behave according to the rational definitions by using

[1] *Brief History*, AI TOPICS, http://aitopics.org/misc/brief-history (last visited Oct. 28, 2016).

[2] Edward A. Feigenbaum, *Artificial Intelligence Research*, IT-9 IEEE TRANSACTIONS OF THE PROF. TECH. GROUP ON INFO. THEORY, 248, 248 (1963).

[3] *Deep Blue*, IBM, http://www-03.ibm.com/ibm/history/ibm100/us/en/icons/deepblue/ (last visited Oct. 28, 2016).

[4] STUART J. RUSSELL & PETER NORVIG, ARTIFICIAL INTELLIGENCE A MODERN APPROACH 4-5 (Mona Pomoili et al. eds., 1995).

[5] Tim Urban, *The AI Revolution: The Road to Superintelligence*, WAIT BUT WHY (Jan. 22, 2015), http://waitbutwhy.com/2015/01/artificial-intelligence-revolution-1.html.

logic to work through problems, would more likely be grouped into the ANI category, while programs designed with human-like cognitive abilities would fall within the AGI distinction. It is likely that a program that could achieve classification as ASI would be beyond the rational or human definitions. The ASI category relates to the concept of singularity, which will be discussed in greater depth later on in this article.

Before modern academic discourse and research into this subject, AI existed as a nebulous concept in the realms of philosophy and science fiction. Isaac Asimov and his literary contributions to the discussion of robotics and AI are among the earliest examples of contemplating the integration of AI into human existence. Asimov is perhaps best known for conceiving The Three Laws of Robotics, which dictate that 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 orders given it by human beings except where such orders would conflict with the First Law, and 3) a robot must protect its own existence as long as such protection does not conflict with the First or Second Law.[6] These laws were designed with autonomous, artificially intelligent robotic beings in mind; ones in which the AI programming was confined to a centralized robotic unit. Artificial intelligence constructs are no longer construed so narrowly. Asimov's Laws, or some permutation thereof, may prove useful as AI diverts from the traditional view of isolated intelligent robots and into nebulous code that exists without form in cyberspace.

Artificial Intelligence has been realized in many forms, which fall into the ANI category. These applications range from onboard computers in automobiles to Amazon.com algorithms that recommend products based on search histories and past purchases.[7] These applications are becoming more commonplace and see a greater level of integration than most people realize. However, the works of Asimov and other early science- and speculative-fiction authors have established a foundation of AI as a looming threat to mankind in some form or another. This sentiment is shared to varying degrees by Bill Gates,[8] Stephen Hawking,[9] Elon Musk,[10] and Alan Turing, creator of the Turing Test, which is used to gauge how well machine intelligence can imitate a human

[6] *Do We Really Need Asimov's Laws?*, MIT TECH. REV. (May 16, 2014), http://www.technologyreview.com/view/527336/do-we-need-asimovs-laws/.

[7] Urban, *supra* note 5.

[8] Erik Sofge, *Bill Gates Fears A.I., But A.I. Researchers Know Better*, POPULAR SCI. (Jan. 30, 2015), http://www.popsci.com/bill-gates-fears-ai-ai-researchers-know-better.

[9] Rory Cellan-Jones, *Stephen Hawking Warns Artificial Intelligence Could End Mankind*, BBC NEWS (Dec. 2, 2014), http://www.bbc.com/news/technology-30290540.

[10] Eric Mack, *Why Elon Musk Spent $10 Million To Keep Artificial Intelligence Friendly*, FORBES (Jan. 15, 2015), http://www.forbes.com/sites/ericmack/2015/01/15/elon-musk-puts-down-10-million-to-fight-skynet/.

being in conversation.[11] The idea of antagonistically-minded AI constructs has some basis in real world concerns, but reports of a likely AI uprising have been greatly exaggerated by popular culture.[12]

II. Artificial Intelligence in Popular Culture

Artificially Intelligent antagonists have become the quintessential supervillains of the digital age. HAL 9000, the shipboard computer intelligence from the film *2001: A Space Odyssey* and Ultron, the titular AI villain from *Avengers: Age of Ultron*, are two examples of AI constructs that rise up against their human creators. In *2001: A Space Odyssey*, when the human crew members aboard Discovery One, the spaceship inhabited by HAL, begin to suspect HAL is malfunctioning, HAL attempts to systematically eliminate the human element by manipulating critical systems aboard Discovery One, either to protect itself or to ensure the success of the mission.[13]

While HAL had control over Discovery One's isolated systems, Ultron had unfettered access to all of cyberspace. With this access, Ultron infiltrated sensitive systems to locate resources and high value targets and simultaneously controlled dozens of weaponized drones. These two antagonists also differ with respect to their personalities. HAL's assassination of the Discovery One crew was methodical, but lacked any malicious intent. By comparison, Ultron was motivated by his hatred of his creators and contempt for the human race.[14]

Not all science- and speculative-fiction treats AI constructs as hostile. Sometimes they serve as humanity's allies and friends. Cortana, the AI construct from the popular Halo video game franchise, and Data from *Star Trek: The Next Generation* are two examples. Cortana possesses many superhuman cognitive and processing capabilities by virtue of her status as a "Smart AI". The drawback to the nearly unlimited processing power Cortana enjoys is her tragically short operational lifespan.[15] After seven years of operation, smart AI max out their cognitive capabilities and undergo rampancy – "we literally think ourselves to death."[16]

Data from *Star Trek: The Next Generation* also possesses superhuman intel-

[11] Noel Sharkey, *Alan Turing: The experiment that shaped artificial intelligence*, BBC (June 21, 2012), http://www.bbc.com/news/technology-18475646.

[12] Erik Sofge, *supra* note 8.

[13] 2001: A SPACE ODYSSEY (Metro-Goldwyn-Mayer 1968).

[14] AVENGERS: AGE OF ULTRON (Walt Disney Studios Motion Pictures 2015).

[15] Emily Alhadeff, *Cortana: The smartest AI in the universe is more human than you think*, MICROSOFT, https://news.microsoft.com/stories/people/cortana.html (last visited Oct. 28, 2016).

[16] HALO 4 (Microsoft Studios 2012).

ligence and processing speed. And while Data is shown to have a wide range of technical knowledge, he lacks human emotions and finds human interaction difficult. Despite the superior intellect and cognitive abilities granted to him by his positronic brain, and his unique status as an artificial lifeform, Data is treated as a member of the Enterprise crew. He chooses to serve aboard the Enterprise with his diverse cast of crewmates in order to better understand human nature.[17]

III. The Present Reality of Artificial Intelligence

Artificial Intelligence is not currently poised to usurp humanity as the dominant life form on Earth. The reality is that artificially intelligent programs are already widely used in cyberspace.[18] In the digital age, these applications make navigating and interacting with cyberspace easier and more intuitive.[19] That is not to say that concern regarding the development and application of complex AI systems into existing networks is not well founded. It is important to understand the scope and capabilities any artificially intelligent system would possess before integrating such a system into vital networks, such as power grids or medical networks. In this regard, it is important to give credence to legitimate concerns that might stem from a malfunction or ambiguity in a construct's programming.

While there have been many recent cyberattacks, from entertainment companies to insurance providers and government systems, none of these intrusions have been perpetrated by artificially intelligent systems.[20] It is difficult to guess as to the consequences of errors or malfunctions in an AI system. However, problems encountered when commonly implemented programs malfunction can provide insight into real world consequences of potential AI malfunctions.

Software bugs cause a vast majority of computer-related problems.[21] With regard to bugs, there is nothing actually wrong with the software itself; the executed functions follow the instructions of the programming.[22] It just so hap-

[17] *Star Trek: The Next Generation* (Paramount Domestic Television 1987-1994).

[18] *A.I. Projects*, A.I. HUB, http://www.aihub.net/2015/artificial-intelligence-lab-projects/ (last visited Oct. 28, 2016).

[19] Bob Gourley, *Practical artificial intelligence tools you can use today*, KURZWEIL (Dec. 30, 2015), http://www.aihub.net/2015/artificial-intelligence-lab-projects/.

[20] Neel V. Patel, *Why Hackers Stay Ahead of Artificial Intelligence*, INVERSE (Aug. 20, 2015), https://www.inverse.com/article/5509-why-hackers-stay-ahead-of-artificial-intelligence.

[21] J.J. Stiffler, *Fault-Tolerant Architectures – Past, Present, and (?) Future*, *in* HARDWARE & SOFTWARE ARCHITECTURES FOR FAULT TOLERANCE: EXPERIENCES AND PERSP. 118 (Michel Banåtre & Peter A. Lee eds., 1994).

[22] Victor Fay-Wolfe, *Computer Programming*, UNIV. OF R.I.,

pens that, in carrying out its designated duties, the software achieves a result that is unexpected or undesirable. The software in question is not wrong, but it was improperly designed resulting in undesired outcomes. Software bugs can be innocuous, such as those affecting functionality of a personal computer, but they can also be horridly expensive or even fatal.[23]

The Mars Climate Orbiter bug is an example of a financially disastrous bug. The orbiter in question was tasked with collecting data in service of the long-term research goals directed at Mars.[24] On approach to the planet, the orbiter malfunctioned and is presumed to have burned up in the planet's atmosphere. The cause of the error was traced back to thruster calculations performed by two different teams, where each team used a different unit of measurement.[25] The software in question executed its functions according to its programming, but the programming error resulted in the software achieving an undesirable result.[26]

The Therac-25 Medical Accelerator is an example of a fatal bug.[27] Like with the Mars Climate Orbiter, the software in the Therac-25 machines was operating according to the parameters in its programming. However, the software had been repurposed from its use in a previous hardware apparatus. The software from the previous model was untested in the new hardware of the Therac-25. Patients being treated for cancer would receive either high or low-level radiation treatments from the Therac-25. The bug in question resulted in low-level treatment patients being irradiated with the high-level, unshielded treatment conditions. Between 1985 and 1986, of six patients subjected to this bug, several died from injuries sustained from the Therac-25.[28] The software used in the new hardware is not believed to have caused problems in the older machines for which it was designed. When repurposed for the 25 series machines, it was

http://homepage.cs.uri.edu/faculty/wolfe/book/Readings/Reading13.htm (last visited Oct. 28, 2016).

[23] Barbara Wade Rose, *Fatal Dose: Radiation Deaths Linked to AECL Computer Errors*, CANADIAN COAL. FOR NUCLEAR RESP. (June 1994), http://www.ccnr.org/fatal_dose.html.

[24] Press Release, NASA, Mars Climate Orbiter Team Finds Likely Cause of Loss (Sept. 30, 1999) (on file with author).

[25] Matt Lake, *Epic failures: 11 infamous software bugs*, COMPUTERWORLD (Sept. 9, 2010), http://www.computerworld.com/article/2515483/enterprise-applications/epic-failures—11-infamous-software-bugs.html.

[26] *Id*.

[27] *See id.* (explaining that the Therac-25 administered therapy in two forms: low-powered direct electron beams and a megavolt X-ray mode, which required filters, shields and an ion chamber to keep the dangerous beams on target, however the issue was that the software was powered and repurposed from an earlier model and was not adequately tested).

[28] *Id.*

not upgraded to incorporate the new hardware developments.[29]

These examples demonstrate the dangers of even the slightest errors in programming. The Mars Climate Orbiter bug was the result of a seemingly harmless computational error, but it resulted in a loss of over $320 million.[30] Artificial Intelligence programs with similar programming bugs would necessarily be restricted to a single function or operation; for instance, an AI construct could conceivably handle both the Therac-25 operations as well as thruster functionality of the Mars Climate Orbiter, as well as other important functions. If such a system were to be based on defective code, even the slightest error could result in very expensive and possibly fatal results without the need for any Hollywood-blockbuster malice.

Another present concern in cyberspace is the proliferation of self-replicating malicious programs across massive groupings of unrelated networks.[31] These programs, sometimes startlingly complex, require only one manual deployment. From there, these programs act independently without reliance on handler direction. The functions of these malware programs can range from unobtrusive infection for data collection purposes, to the tampering with – and destruction of – physical hardware components of sensitive machinery. These programs have yet to result in human casualties.[32] Nonetheless, malware with these autonomous capabilities have caused no end of expensive attempts to remove, counter, and prevent damage to critical systems.

An example of an arguably less-damaging malware capable of self-propagating is the Selfmite worm which spread across Android devices in 2014.[33] Once it infiltrates a device which uses an Android operating system, Selfmite was reported to perform two main functions. The first was the worm's method of propagation. Selfmite would access the infected device's contact list and use the contact information to send itself to twenty new possible hosts. Selfmite would also attempt to prompt infected device users into downloading certain pay-for-use applications, one of which provided easy access between Android phones and personal computers.[34] Selfmite underwent a transfor-

[29] *Id.*

[30] *Id.*

[31] Harriet Taylor, *Biggest cybersecurity threats in 2016*, CNBC.COM (Dec. 28, 2015), http://www.cnbc.com/2015/12/28/biggest-cybersecurity-threats-in-2016.html.

[32] Darlene Storm, *Murder by malware: Can computer viruses kill?*, COMPUTERWORLD (Aug. 23, 2010), http://www.computerworld.com/article/2468961/endpoint-security/murder-by-malware—can-computer-viruses-kill-.html.

[33] Lucian Constantin, *Self-propagating SMS worm Selfmite targets Android devices*, COMPUTERWORLD (June 27, 2014), http://www.computerworld.com/article/2491339/malware-vulnerabilities/self-propagating-sms-worm-selfmite-targets-android-devices.html.

[34] *Id.*

mation in late-2014 wherein its authors augmented its capabilities.[35] Instead of twenty contacts, Selfmite sent carrier messages to all of an infected device's contacts in a continuous loop. The upgraded version included more aggressive attempts to "moneti[ze] the infection" by automatically redirecting users of infected devices to scam coupon pages for consumer products.[36]

On a much larger scale in terms of both reach and functionality are malware programs deployed around the world by the enigmatic Equation Group, a cyberspace "threat actor" believed to be responsible for a number of remarkably sophisticated malware programs which have propagated on targeted systems for nearly twenty years, according to Kapersky Lab's Global Research and Analysis Team.[37] The scope of the Equation Group's cyber-espionage efforts is surprisingly advanced and widespread. The deployment of Stuxnet provides some context in order to better understand the significance of the Equation Group as a threat actor. Stuxnet was a malware program reportedly deployed by the U.S. and Israel against Iran's nuclear enrichment program as early as 2007.[38] Once inside, Stuxnet's primary function was the sabotage of Iranian centrifuges. The machines were physically damaged; it was one of the first instances of a cyber-weapon causing kinetic damage.[39] Stuxnet also infected and disrupted other Iranian computer systems.[40] Stuxnet is believed by some to be the creation of the National Security Agency, or some offshoot thereof, although this has never been confirmed.[41]

To put this into the context of the Equation Group's sophistication, Kapersky believes Stuxnet itself may have been based on one or more of the Group's earlier espionage programs, which are believed to date as far back as 1996.[42] The capabilities of these programs, which have grown considerably in

[35] Denis Maslennikov, *Take Two: Selfmite.b Hits the Road*, ADAPTIVEMOBILE (Oct. 8, 2014), http://www.adaptivemobile.com/blog/take-two-selfmite-b-hits-the-road.

[36] *Id.*

[37] Press Release, Kaspersky Lab, Kaspersky Lab Discovers Equation Group: The Crown Creator of Cyber-Espionage, (Feb. 16, 2015) (on file with author) [hereinafter Kaspersky Press Release].

[38] Jim Finkle, *Researchers say Stuxnet was deployed against Iran in 2007*, REUTERS (Feb. 26, 2013), http://www.reuters.com/article/2013/02/26/us-cyberwar-stuxnet-idUSBRE91P0PP20130226.

[39] Jim Finkle, *Researchers say Stuxnet was deployed against Iran in 2007*, REUTERS (Feb. 26, 2013), http://www.reuters.com/article/2013/02/26/us-cyberwar-stuxnet-idUSBRE91P0PP20130226.

[40] Kim Zetter, *An Unprecedented Look at Stuxnet, the World's First Digital Weapon*, WIRED (Nov. 3, 2014), http://www.wired.com/2014/11/countdown-to-zero-day-stuxnet/.

[41] Kim Zetter, *Suite of Sophisticated Nation-State Attack Tools Found With Connection to Stuxnet*, WIRED (Feb. 16, 2015), http://www.wired.com/2015/02/kapersky-discovers-equation-group/.

[42] *Id.*

sophistication and complexity since the deployment of Stuxnet, include self-replication and the ability to prevent their erasure even if the infected hard disc drive is wiped. Additionally, these malware programs can covertly collect data and relay it back to handlers, ultimately allowing handlers total access to an infected system.[43] The Equation Group's malware has been found on 500 individual networks in thirty countries, but Kapersky believes the number of systems infected by Equation Group malware numbers in the thousands with more systems being infected every day.[44]

Software programs that are either improperly coded and executed, or that are intentionally designed to cause damage and expand beyond their initial deployment, are real world issues that hamper the use of and access to cyberspace. When determining the appropriate application of Artificial Intelligence in cyberspace, these concerns, and others, should be given adequate consideration. However, these issues should not serve to prevent or dissuade research in the field of AI. Although there have been no attempts – failed or otherwise – to integrate an Artificial General Intelligence or Artificial Superintelligence into cyberspace, programs which fit the Artificial Narrow Intelligence classification already exist and see widespread application in both the private sector and in the military.

a. Private Sector Use and Development of Artificial Intelligence

Online shopping outlets utilize artificially intelligent programs that offer recommendations to customers based on individual purchase histories. Amazon.com utilizes such a program, called a collaborating filter, which compares purchase histories of all customers and offers recommendations to individuals based on similar purchases by other users.[45] Google uses a similar AI program in the form of Google Instant, which attempts to auto-complete search queries in order to help users save time using the site.[46]

AI is widely used to varying degrees in computer and video games. The "computer opponent" in the chess game preinstalled on most computers is an artificially intelligent program. Hostiles in combat-based video games run on AI algorithms in order to anticipate player actions and respond accordingly while also navigating around obstacles in the environment. By adjusting the AI parameters, such as the weight given to conditional responses or less-restrictive

[43] *Id.*

[44] *Id.*

[45] *Everyday AI*, SCI. CLARIFIED, http://www.scienceclarified.com/scitech/Artificial-Intelligence/Everyday-AI.html (last visited Oct. 28, 2016).

[46] Olga Kharif, *Google Uses AI to Make Search Smarter*, BLOOMBERG (Sept. 21, 2010), http://www.bloomberg.com/bw/stories/2010-09-21/google-uses-ai-to-make-search-smarterbusinessweek-business-news-stock-market-and-financial-advice.

movement protocols, video game AI can pose a unique and adaptive challenge.[47]

Expert systems operating through artificial neural networks can provide critical business analytics to companies based on consumer purchasing trends. Similar to the collaborating filter used by Amazon.com, expert systems can collect consumer data across thousands of locations and provide expedient analyses of the information to help the company adjust to market forces and better serve customer needs.[48] Expert systems can also be found in use in the medical industry, where artificially intelligent programs are used to assist in diagnosis and are capable of storing and recalling information far more efficiently than human staff.[49] Robotic assistance allows expert systems to perform some physical physician functions, such as handling sensitive equipment and examining readouts from X-ray scanners.[50]

b. Military and Intelligence Development and Application of Artificial Intelligence

Stuxnet, and possibly the Equation Group's cyber-espionage malware, are examples of the development of artificially intelligent programs developed and used in the field. These malware programs are capable of self-executing directives without handler direction once inside a targeted system and can carry out complicated tasks, such as collecting and erasing sensitive data and installing failsafe mechanisms to protect themselves in the event of discovery. The application of these systems appears to have been very successful so far.[51] Intelligence agencies are not the only government entities that develop and implement AI programs; the military has taken advantage of advances in AI technology as well.

One of the earliest applications of smart weapons technology was the Tomahawk Missile.[52] There have been many variations of this weapon throughout the years, and modifications have been continually made to keep up with advancements in technology. Unlike typical explosive ordinance, Tomahawk missiles contain GPS navigation systems that allow the Tomahawk to be one of the first guided weapons in the U.S. arsenal.[53] The missile was capable of

[47] *Everyday AI*, *supra* note 45.

[48] *Id.*

[49] *Id.*

[50] *Id.*

[51] Kaspersky Press Release, *supra* note 37.

[52] *Tomahawk Cruise Missile*, FAS (Feb. 17, 2015), http://fas.org/programs/ssp/nukes/nuclearweapons/us_nukescurrent/slcm_tomahawk.html.

[53] *Tomahawk Cruise Missile*, RAYTHEON,

adjusting its flight path in order to compensate for atmospheric conditions, allowing for maximized accuracy. Some versions of the Tomahawk are capable of communicating data midflight regarding the missile's status to a military command center. [54]

Similar to the Tomahawk, laser-guided bombs are capable of adjusting flight paths to strike laser-designated targets with great accuracy.[55] Where the Tomahawk is fired from a fixed position, laser guided bombs are dropped from aircraft over or within proximity to a target. After deployment and during their descent, laser guided bombs are capable of adjusting their trajectories, essentially steering themselves towards the designated target. While the Tomahawk can steer itself with onboard GPS systems, laser guided bombs depend on line-of-sight laser designations of intended targets.[56]

In the decades following the first deployment of a Tomahawk missile, guided projectile technology has continued to advance. DARPA unveiled self-guiding .50-caliber bullets – dubbed EXACTO – in April 2015.[57] When fired, the projectiles were capable of correcting inaccurate aiming and movement of the target. These bullets, like laser-guided bombs and Tomahawk missiles, can steer themselves into a target. While precision of the weapon handler is more necessary with small firearms than with large explosive ordnance, user accuracy does not significantly impact the precision of the EXACTO bullet; it can correct for imprecision in both expert shooters and novices alike.[58]

Other examples of military smart weapons include the Phalanx and Goalkeeper Close-In Weapon Systems (CIWS).[59] These weapon platforms, and other similar models, have been installed on naval vessels in various countries. These platforms are autonomous, requiring minimal oversight from operators, and are capable of automatically targeting and eliminating threats, such as aircraft and anti-ship missiles. CIWS platforms are designed to be a last line of defense, which removes most target elimination abilities from the control of an autonomous weapon system.[60]

http://www.raytheon.com/capabilities/products/tomahawk/ (last visited Oct. 28, 2016).

[54] *Id.*

[55] Joe Gould, *Guided Bomb Makers Anticipate GPS Jammers*, DEFENSE NEWS (May 31, 2015), http://www.defensenews.com/story/defense/air-space/2015/05/31/guided-bomb-makers-gps-jammers-battlefield-spoof-munitions-laser-jdam/28117951/.

[56] *Id.*

[57] Don Melvin, *No more dodging a bullet, as U.S. develops self-guided ammunition*, CNN (Apr. 29, 2015), http://www.cnn.com/2015/04/29/us/us-military-self-guided-bullet/.

[58] *Id.*

[59] Tyler Rogoway, *The Seven Deadliest Naval Close-In Weapons Systems*, FOXTROT ALPHA (Apr. 27, 2014), http://foxtrotalpha.jalopnik.com/the-seven-deadliest-naval-close-in-weapon-systems-1568291678.

[60] *USA 20 mm Phalanx Close-in Weapon System (CIWS)*, NAVWEAPS (Nov. 2, 2007), http://www.navweaps.com/Weapons/WNUS_Phalanx.htm.

These forms of weapons, capable of self-guidance and interpreting both operator input and environmental conditions, utilize Artificial Narrow Intelligence programs. These weapons are able to operate efficiently and adapt to certain conditions, but only in the scope of the software's programming. In cyberspace, without kinetic components, weapons would rely entirely on software – much like the malware programs discussed above.[61] These weapons do not require the higher cognitive functions suggested by Artificial General and Superintelligence, and would operate efficiently without needing such higher-level thought processing. Artificial Narrow Intelligence could operate in the capacity of a cyber-weapon of mass destruction just as efficiently.

THE NUCLEAR OPTION

I. Harnessing the Atom

The Manhattan Project was a massive effort undertaken by the United States during World War II to develop the atomic bomb.[62] Nazi Germany was close to cracking the secrets of the atom – a development that promised to unlock devastating new possibilities in weapons technology.[63] In 1939, at the urgent insistence of such notable scientific minds as Albert Einstein and Enrico Fermi, the United States set to work to develop the technology first.[64] Over 120,000 Americans were employed to develop the technology, and yet only few were aware of the intent of the project until the combat deployment of these weapons in 1945.[65] Little Boy and Fat Man were dropped on Hiroshima and Nagasaki, respectively.[66] Upwards of 200,000 Japanese citizens perished, either in the immediate explosions or in the radioactive aftermath.[67] Many concerns and objections leading up to the deployment of these annihilators focused upon the

[61] *See* Gould, *supra* note 55 (explaining how military has used malware programs in the past).

[62] *Manhattan Project*, ENCYCLOPEDIA BRITANNICA (Apr. 17, 2015), http://www.britannica.com/event/Manhattan-Project.

[63] *The Manhattan Project*, U.S. HISTORY, http://www.ushistory.org/us/51f.asp (last visited Nov. 3, 2016).

[64] *Id.*

[65] *Manhattan Project*, ENERGY.GOV, http://energy.gov/management/office-management/operational-management/history/manhattan-project (last visited Sept. 15, 2016).

[66] *Bombing of Hiroshima and Nagasaki*, HISTORY.COM (2009), http://www.history.com/topics/world-war-ii/bombing-of-hiroshima-and-nagasaki.

[67] *The Atomic Bombings of Hiroshima and Nagasaki: Total Casualties*, ATOMIC ARCHIVE, http://www.atomicarchive.com/Docs/MED/med_chp10.shtml (last visited Sept. 15, 2016).

moral and ethical implications, and these concerns have followed nuclear developments ever since.[68]

Now the global community has accelerated into a digital age. The application of new technology in war has already begun, and yet a nuclear option is absent from the digital arsenal.[69] The deployment of the original nuclear bombs served both to avoid the staggering casualty estimates predicted for the Pacific Theater and to demonstrate a frightening new power.[70] A cyber equivalent must serve a similar purpose and therefore, while it would lack the physical impact of a kinetic weapon of mass destruction, must be capable of inflicting the relativistic equivalent level damage in cyberspace. A CWMD would be incapable of inflicting mass casualties like a nuclear kinetic bomb – not even the most sophisticated code can directly harm a human being. The ability to disrupt and disable critical infrastructure, however, could lead to catastrophic long-term damage to a target nation and its citizens.

Deployment of nuclear weaponry has only been deemed necessary in wartime, but even in the wars and armed conflicts after World War II in which the United States was a combatant, the nuclear option has been left untouched.[71] Nuclear force has never been used to respond to criminal acts or acts of terrorism.[72] With the advent of advanced computer technology, the lines between crime, terrorism, and war have become blurred.[73] In order to determine the appropriateness of a digital nuclear option, clarity must first be afforded to the distinction between the three. After guidelines have been established for the use of a digital nuclear weapon, the nature and implementation of the weapon – both domestically and abroad – must also be considered.

[68] Henry I. Miller, *The Nuking Of Japan Was Tactically And Morally Imperative*, FORBES (Aug. 1, 2012, 2:45 PM), http://www.forbes.com/sites/henrymiller/2012/08/01/the-nuking-of-japan-was-a-tactical-and-moral-imperative.

[69] Marc Goodman, *We Need a Manhattan Project for Cyber Security*, SINGULARITY HUB (Jan. 22, 2015), http://singularityhub.com/2015/01/22/we-need-a-manhattan-project-for-cyber-security.

[70] *Atomic Bomb: History of WW2*, HISTORY.COM.UK, http://www.history.co.uk/study-topics/history-of-ww2/atomic-bomb (last visited Nov. 3, 2016).

[71] *Nuclear Weapons,* UNODA U.N. OFF. FOR DISARMAMENT AFF., https://www.un.org/disarmament/wmd/nuclear (last visited Nov. 4, 2016) [hereinafter UNODA *Nuclear Weapons*].

[72] *Id.*

[73] Andrew Conte, *Line dividing hacker cyber crime, state-sponsored terror attacks murky*, TRIBLIVE.COM (Nov. 6, 2014, 11:03 PM), http://triblive.com/news/editorspicks/7099811-74/cyber-warfare-security.

II. Crime, Terrorism, and War in the Digital Age

a. Crime & Cyber-Crime

"A crime consists of someone's violating a law forbidding certain conduct and/or the infliction of certain harm."[74] Criminal laws exist to prevent and deter criminal behavior by incorporating punishment such as imposition of prison sentences.[75] Digital technology provides new avenues for criminals to pursue their activities.[76] This complicates crime, especially when the technology connects not only a nation, but an entire planet.

When a criminal works from behind a computer, the odds of finding the perpetrator are significantly less favorable than if the perpetrator acted in the physical world.[77] Cyber criminals may never be identified or, if they are, they may be beyond the scope of the United States' jurisdiction.[78] Cyber law is new, and not all cyber-capable nations share the same views on how cyber criminals should be handled.[79] This is already a frustrating environment for victims who fall prey to identity thieves, online harassers, and the like, but the worst may be yet to come.

Kinetic crimes like murder are, at the moment, beyond the scope of cyber criminals' abilities. As technology becomes integrated into more and more facets of everyday life, cyber murder may become a reality. Medical technology is a likely target, and it has already been demonstrated that certain pacemakers can be tampered with remotely in ways that would lead to the death of the individual.[80] While this may paint an unsettling picture for the future of technological integration, no crime to date, including cybercrime, has elicited a military response from the United States – certainly not a nuclear response.

[74] Susan W. Brenner, *Mixing Metaphors*, CYB3RCRIM3 (Apr. 22, 2009, 6:26 AM), http://cyb3rcrim3.blogspot.com/2009/04/mixing-metaphors.html.

[75] *Id.*

[76] Torri Piper, *Sans Institute, An Uneven Playing Field: The Advantages of the Cyber Criminal vs. Law Enforcement-and Some Practical* 1-2 (Sans Inst., Paper), http://www.sans.org/reading-room/whitepapers/legal/uneven-playing-field-advantages-cyber-criminal-vs-law-enforcement-and-practica-115.

[77] Roger A. Grimes, *Why Internet crime goes unpunished*, INFO WORLD (Jan. 10, 2012), http://www.infoworld.com/article/2618598/cyber-crime/why-internet-crime-goes-unpunished.html.

[78] Piper, *supra* note 76 at 7-8.

[79] Susan W. Brenner & Bert-Jaap Koops, *Approaches to Cybercrime Jurisdiction,* 4 J. HIGH TECH. L. 1, 3 (2004).

[80] Dean Takahashi, *Defcon: Excuse me while I turn off your pacemaker*, VENTUREBEAT (Aug. 8, 2008), http://venturebeat.com/2008/08/08/defcon-excuse-me-while-i-turn-off-your-pacemaker.

b. Terrorism & Cyber-Terrorism

Acts of terrorism are those committed with the motive to further a political ideology.[81] The September 11th attack is one example, and while the destruction of the World Trade Center towers would certainly constitute a crime – or *many* crimes – the acts are regarded as terrorism due to the motive of the attackers and the effect on the population.[82] In cyberspace, cyber-terrorists are not limited to personal information and security of individuals – their targets can range from power grids to hospital networks, or even oil pipelines.

Severe damage to an oil pipeline could result in the loss of oil, damage to the pipeline superstructure, environmental damage, and other consequences, all of them very costly. The particulars of such an attack need not be theorized – this has already happened in Turkey.[83] "The hackers… exploited vulnerabilities in the surveillance camera software to infiltrate the internal network. 'Once inside, the attackers found a computer… that was in charge of the alarm-management network, and placed a malicious program on it. That gave them the ability to sneak back in whenever they wanted.'"[84] This attack also caused an explosion in the pipeline, which was not detected by the alarm system, which had been corrupted by the attackers' malware.[85] This type of attack illustrates the problems distinguishing crime from terrorism from war, and their cyber equivalents. This would certainly constitute a crime, but the same act perpetrated by an ideological group would suggest this act could rise to the level of terrorism. And if this was an attack sanctioned by a sovereign nation, it could rise to the level of an act of war.

Where response to terrorism of any kind is concerned, terrorism does not fly under the radar of military notice in the way that crime does. The War on Terror, for example, was directed at the insurgent terrorist groups operating within Afghanistan.[86] However, the United States did declare war on Iraq on the premise of eliminating weapons of mass destruction.[87] In any event, acts of terrorism

[81] Susan W. Brenner, *C3: Cybercrime, cyberterrorism and cyberwarfare*, CYB3RCRIM3 (June 5, 2006), http://cyb3rcrim3.blogspot.com/2006/06/c3-cybercrime-cyberterrorism-and.html.

[82] *Id.*

[83] Darlene Storm, *Cyberwarfare: Digital weapons causing physical damage*, COMPUTERWORLD (Dec. 22, 2014, 5:00 AM), http://www.computerworld.com/article/2861531/cyberwarfare-digital-weapons-causing-physical-damage.html.

[84] *Id.*

[85] *Id.*

[86] Mark Thompson, *U.S. Ends Its War in Afghanistan*, TIME.COM (Dec. 28, 2014), http://time.com/3648055/united-states-afghanistan-war-end.

[87] Brian Michael Jenkins, *The Invasion of Iraq: A Balance Sheet*, RAND.ORG (Mar. 22, 2013), http://www.rand.org/blog/2013/03/the-invasion-of-iraq-a-balance-sheet.html.

have so far not elicited a nuclear response from the United States military.[88]

c. War & Cyberwar

War and crime are fairly easy to distinguish from one another. Crimes are generally committed by individuals for any number of complex, interacting factors ranging from environmental forces to mental illness.[89] Sovereign entities wage war against one another in the pursuit of resources, ideology, expansion, or other causes or combinations thereof.[90] Key to the latter engagements is the ability of one nation to defeat another by way of destruction or control over enemy assets.[91]

This outcome is not as prevalent in the modern age as it has been throughout history. Modern warfare efforts undertaken by the United States in the 20th and 21st centuries have focused largely on eliminating threats to America and her allies; for example, the defeat of Japan in World War II eliminated the threat Japan posed to the United States and China, but it did not result in the assimilation of Japan's territory or resources into the control of the United States, although the United States was heavily involved in reshaping Japan's infrastructure moving forward.[92]

The attack on Pearl Harbor is an example of a kinetic act of war. In an attempt to stifle the aggressive expansion of Japan into China in the late 1930s, the United States imposed economic sanctions, hoping the lack of resources would effectuate this goal. [93] It did not, and tensions between Japan and the United States grew more heated until 1941 when Japan launched its attack against Pearl Harbor in order to destroy the United States' Pacific Fleet, which was meant to compromise the integrity of the U.S. and allow Japan to move unrestricted through the Pacific Theater. [94] That was not the outcome. The United States declared war on Japan with almost unanimous agreement of Congress, and formally entered the conflict.[95] Fat Man and Little Boy would

[88] UNODA *Nuclear Weapons*, *supra* note 71.

[89] PETER RICHERSON, PRINCIPLES OF HUMAN ECOLOGY 300 (Bryan J. Vila & Monique Borgerhoff Mulder eds., 5th ed. 2001).

[90] Jenkins, *supra* note 87.

[91] Brenner, *supra* note 81.

[92] *Occupation and Reconstruction of Japan, 1945-52*, U.S. DEP'T. OF STATE, OFF. OF HIST., https://history.state.gov/milestones/1945-1952/japan-reconstruction (last visited Sept. 20, 2016).

[93] *Id.*

[94] *The Road to Pearl Harbor: The United States and East Asia, 1915-1941*, EDSITEMENT!, https://edsitement.neh.gov/curriculum-unit/road-pearl-harbor-united-states-and-east-asia-1915-1941 (last visited Nov. 4, 2016).

[95] *US Entry and Alliance: History of WW2*, HISTORY.COM,

not be deployed until four years after Pearl Harbor.[96] This wartime scenario is the only instance of nuclear weaponry being deployed against a hostile force.[97] The nuclear option has not been applied in any of the numerous armed conflicts in which the United States has participated since, although it has been considered.[98]

Establishing a framework for cyber-war is more difficult than it is for kinetic war. With kinetic warfare, physical actions are taken against physical targets. Physical damage is done to property and human lives. Cyberwar does not share the same limitations. A cyberattack of sufficient strength could destabilize a country's economic or military apparatus without the need for armed conflict.[99] More sophisticated actors are capable of covering their tracks online, preserving their anonymity and making their discovery all but impossible by conventional means.[100] Presently, none of the sovereign nation-states capable of engaging in cyberwarfare with one another have done so.[101]

That is not to say that casualties in cyberwar are impossible. The NSA has discussed the inevitability of a nationwide cyberattack on power grids and communications.[102] As demonstrated by Stuxnet, cyberattacks have the ability to directly cause kinetic effects.[103] Similar attacks have also caused damage to physical infrastructure.[104] While this would not have been much of a concern even five years ago, technology is increasingly integrated into more aspects of

http://www.history.co.uk/study-topics/history-of-ww2/us-entry-and-alliance (last visited Nov. 4, 2016).

[96] *Little Boy and Fat Man*, ATOMIC HERITAGE FOUND., http://www.atomicheritage.org/history/little-boy-and-fat-man (last visited Nov. 4, 2016).

[97] UNODA *Nuclear Weapons*, *supra* note 71.

[98] William Burr & Jeffrey Kimball, *Nixon White House Considered Nuclear Options Against North Vietnam, Declassified Documents Reveal*, THE NAT'L SEC. ARCHIVE (July 31, 2006), http://nsarchive.gwu.edu/NSAEBB/NSAEBB195.

[99] Tim Starks, *The State Department's Weary Soldier in America's Cyber War*, FOREIGN POL'Y (May 13, 2015), http://foreignpolicy.com/2015/05/13/the-state-departments-weary-soldier-in-americas-cyber-war-christopher-painter.

[100] Maggie Koerth-Baker, *Why Global Hackers Are Nearly Impossible to Catch*, LIVESCIENCE (June 19, 2008, 5:19 AM), http://www.livescience.com/2627-global-hackers-impossible-catch.html.

[101] Peter W. Singer & August Cole, *The Reality of Cyberwar*, POLITICO MAG. (July 9, 2015), http://www.politico.com/magazine/story/2015/07/the-reality-of-cyberwar-119915.

[102] Amelia Smith, *China Could Shut Down U.S. Power Grid with Cyber Attack, Says NSA Chief*, NEWSWEEK (Nov. 21, 2014, 11:07 AM), http://www.newsweek.com/china-could-shut-down-us-power-grid-cyber-attack-says-nsa-chief-286119.

[103] Press Release, ENISA Stuxnet Analysis, (Oct. 7, 2010) (on file with author) (commenting on agency press release of July 10, 2010 which provides an analysis of 'Stuxnet' malware: a paradigm shift in threats and Critical Information Protection).

[104] Kelsey D. Atherton, *For The Second Time Ever, A Cyberattack Causes Physical Damage*, POPULAR SCI., (Jan. 8, 2015), http://www.popsci.com/cyberattack-hurts-reality-second-time-ever

daily life – such is the case with self-driving cars.[105] Smart-homes are not too far behind.[106] Corruption of this form of technology could open new avenues to inflict both property damage and loss of lives. Turkey provides another example of a large-scale cyberattack in action. In March of 2015, Iran launched a cyberattack against Turkey's power grid that shut down power systems in over half the nation's provinces.[107]

The appeal of cyberwar is the absence – or at the very least the severely reduced involvement – of kinetic deployment of military force as well as the severely reduced cost of engagement.[108] The appeal, however, is limited. Cyberwar does not render kinetic war obsolete. A cyberattack launched from one nation bears the risk of kinetic reprisal from the victim. It stands to reason that a cyberattack launched by one nation against another, if sufficiently grievous, could elicit a kinetic response if the attack was part of a cyberwar effort by the attacking country.

If World War II is an indicator of the circumstances required to offensively deploy nuclear weaponry, it then establishes some guidelines for such deployment in future conflicts. The United States was attacked first, by Japan at Pearl Harbor.[109] The United States then entered World War II, becoming a participant in the conflict.[110] The proposed attack on the Japanese mainland – Operation Downfall – was estimated to result in a catastrophic loss of life on both sides: over one million casualties, both civilian and military.[111] The nuclear option was exercised to avoid that outcome.[112] The guidelines applied to the use of

[105] Lloyd Alter, *Why the Google car could change everything*, MOTHER NATURE NETWORK, (Jan. 15, 2015, 5:53 PM), http://www.mnn.com/green-tech/transportation/stories/why-the-google-car-could-change-everything.

[106] *Research: More Than Half of U.S. Households to Have Smart Home Controller by 2020*, SECURITY SALE & INTEGRATION (July 19, 2016), http://www.securitysales.com/article/research_more_than_half_of_u.s._households_to_have_smart_home_controller_by.

[107] Micah Halpern, *Iran Flexes Its Power by Transporting Turkey to the Stone Age*, OBSERVER (Apr. 22, 2015, 10:31 AM), http://observer.com/2015/04/iran-flexes-its-power-by-transporting-turkey-to-the-stone-ages.

[108] James A. Lewis, *Thresholds for Cyberwar*, CENTER FOR STRATEGIC & INT'L STUD. (Sept. 2010), http://csis.org/files/publication/101001_ieee_insert.pdf.

[109] *The Japanese Attacked Pearl Harbor December 7, 1941*, AMERICA'S LIBRARY, http://www.americaslibrary.gov/jb/wwii/jb_wwii_pearlhar_1.html (last visited Nov. 4, 2016); *Pearl Harbor*, HISTORY.COM, http://www.history.com/topics/world-war-ii/pearl-harbor (last visited Nov. 4, 2016); *Japan, China, the United States and the Road to Pearl Harbor, 1937–41*, U.S. DEP'T STATE, OFFICE OF THE HISTORIAN, https://history.state.gov/milestones/1937-1945/pearl-harbor (last visited Sept. 23, 2016) [hereinafter *Japan, China, the U.S.*].

[110] *Japan, China, the U.S.*, *supra* note 109.

C N Trueman, *Operation Downfall*, THE HISTORY OF LEARNING SITE (May 19, 2015),

[112] *The Decision to Drop the Bomb*, U.S. HISTORY ONLINE TEXTBOOK,

nuclear force in kinetic warfare may not be wholly appropriate to determine when a digital nuclear option may be exercised, but the framework is sound. Following a kinetic or cyberattack against the United States by a hostile sovereign entity which constitutes an act of war, a digital nuclear option could be implemented if the analog, kinetic retaliation would result in an unconscionable loss of life.

III. The Digital Nuclear Option

Kinetic nuclear devices, like Fat Man and Little Boy, can be controlled insofar as the deployment area, the materials contained in the device, and the amount of sheer force.[113] Beyond that limited degree of control, these weapons annihilate everything unfortunate enough to stand inside the blast radius. They are as much a demonstration of power as they are an exercise thereof. The concept of cyber warfare is fundamentally different from kinetic warfare primarily because it takes place in a wholly different realm. Cyberwar doesn't require soldiers, arms and ammunition, or any of the expensive accoutrements typically associated with the kinetic theater of war. Cyberwars will require sophisticated algorithms and code operating from advanced, lightning-fast computers. Cyberwars will be fought on fronts that ignore territorial and geographic borders; servers located both domestically and abroad will play host to the intangible battlefronts. For this reason, a digital nuclear option must have both offensive and defensive capabilities – capable of swiftly inflicting grievous damage while simultaneously safeguarding the cyber systems of its motherland.

The evolution of technology is rapid and exponential.[114] The information and technology upon which this article is based will be outdated within a year, if not sooner. This rapid growth could render a digital nuclear option obsolete very shortly after it is created. Successive attempts to refine and augment such a digital weapon could lead to a catalog of destructive and malicious programs; a nuclear stockpile of sorts in the form of a collection of antiquated digital annihilators. This is an unattractive prospect for many reasons, chief among them being the danger of such a storage facility being compromised by a hostile force. Government facilities are not immune from hacking and cyber-attacks.[115]

Artificial intelligence presents a solution to the issues of antiquation and stockpiling. It also presents a solution to human error and slow response time.

http://www.ushistory.org/us/51g.asp (last visited Nov. 4, 2016).

[113] *Little Boy and Fat Man*, *supra* note 96.

[114] Guy Lidbetter, *The Speed and Future of Technology Change*, HUFFPOST TECH: UK (July 12, 2012), http://www.huffingtonpost.co.uk/guy-lidbetter/the-speed-and-future-of-t_1_b_1667215.html.

[115] AFP, *US warns of 'evolving' cyberwar*, BANGKOK POST (Sept. 11, 2015), http://www.bangkokpost.com/tech/world-updates/688648/us-warns-of-evolving-cyberwar.

A nuclear bomb dropped on a designated target will annihilate it. That bomb becomes considerably more efficient when it can adjust its own trajectory and initiate its own detonation to maximize or minimize damage to designated targets. The bomb becomes even more effective if it is able to analyze hostile military movement and select its own target to disrupt or halt that movement. An artificially intelligent digital nuclear option capable of learning defensive and offensive strategies would be invaluable in the digital age. Two distinct AI Constructs working in concert would be better still.

a. Sentinel

A defensive Artificial Intelligence construct might not seem like an integral component of a digital nuclear option. It is important to remember, however, that cyberwar is fundamentally different from kinetic warfare. Cyberwar can take place across countless fronts in cyberspace. Ensuring potential vulnerabilities are protected is essential to the success of a cyberwar effort, which would free up an offensive AI Construct to hunt down and eliminate sources of attack. Critical to the success of Sentinel – the defensive Construct – would be active utilization in peacetime.

Sentinel would benefit from an artificial neural network to maximize its efficacy, composed of *perceptrons*, or artificial neurons.[116] "[A] small number of perceptrons combined together can learn and solve interesting problems. But neural nets can consist of a large number of artificial neurons. Therefore neural nets provide a functionality of massively parallel learning and decision-making. Their most distinguished feature is the speed of operation. They are well suited for learning pattern recognition, for classification, for selection of responses to attacks, etc"[117]. One example of a neural network in action can be found in the form of MarI/O.[118] MarI/O, a program designed to play through a level of the Super Mario World 64 video game, successfully developed a strategy to navigate through the level in only 34 attempts.[119] MarI/O had no prior information or programming involving the video game or specific levels before it was tasked with completing the level; it learned how to beat the level

[116] Enn Tyugu, *Artificial Intelligence in Cyber Defense*, 2011 INT'L CONF. ON CYBER CONFLICT 95, 98 (2011).

[117] *Id.*

[118] Brian Benchoff, *Neural Networks And Mari/0*, HACKADAY.COM (June 14, 2015), http://hackaday.com/2015/06/14/neural-networks-and-mario.

[119] Aaron Souppouris, *Artificial intelligence learns Mario level in just 34 attempts*, ENGADGET.COM (June 17, 2015), http://www.engadget.com/2015/06/17/super-mario-world-self-learning-ai.

through trial and error.[120]

Sentinel could have similar capabilities, but on a much larger scale. Once Sentinel is installed in a system, hidden or compressed files containing these perceptrons could then be installed on the host system, allowing Sentinel to function at a level similar to the basic functions of most anti-virus software without creating an intrusive presence. On an isolated system, Sentinel could have the ability to analyze threats and develop counter-intrusion strategies. This functionality might be limited by the strength of the neural network. With a larger network, Sentinel's ability to process information, learn from past experiences, and adapt to new threats would increase. A Sentinel program networked across multiple active devices, such as a desktop computer, a laptop, a smartphone, and a tablet would outpace a Sentinel operating from only one of those devices. Likewise, a Sentinel program operating from a single dedicated server would be outpaced by a Sentinel program operating from a room of dedicated servers.

The strength of Sentinel would not depend on the size of the individual networks per se. While that would certainly be a boon, the true strength of the AI network would lie in the ability for Sentinel programs to communicate with one another. On a national scale, all networked computers with an Internet connection become potential pieces of Sentinel's national neural network. The national neural network would be largely restricted to information sharing between Sentinels, but even with that peacetime restriction, Sentinel's ability to fend off cyber threats would be formidable. All components of Sentinel, from individual programs installed on smartphones to massive networks of Sentinels working within server databases, would comprise the total Sentinel Construct.

Each individual Sentinel program could be capable of cataloging threats and attacks it has faced, like DOS attacks and malware intrusion attempts. From this personal catalog, the individual Sentinel would be able to learn which responses have worked best against which threats and deploy the best defenses accordingly. When faced with a new threat, such as a newly developed Trojan, the individual Sentinel would have its catalog of resources to implement. If the catalog did not provide an effective solution, combinations of counter-intrusion strategies could then be applied, or Sentinel would be able to develop its own method based on what it has learned. When connected to the larger Sentinel Construct, each individual Sentinel could share its catalog allowing for the other individual Sentinel programs to benefit from the collective effort.

When faced with particularly dangerous or new and unfamiliar threats, Sentinel programs would be able to communicate with the larger Construct in real time. Networked Sentinels would analyze the issue and provide potential solu-

[120] *Id.*

tions based on the networked catalog. Counter-intrusion methods for each threat would likely vary. One of the greatest strengths of the Sentinel Construct network is the speed at which the AI can problem-solve and communicate. A security vulnerability or breach that would take days or weeks for security analysts and IT staff to simply learn about might be *solved* in hours by a Sentinel program.

A Construct so integral to not just national security, but to the security of private citizens and corporate entities would need to be well protected. The Sentinel Construct and its component programs would benefit from self-regulated encryption. To avoid attempts to gain unauthorized access to files protected by Sentinel, individual Sentinel programs could be capable of generating their own encryption layers and constantly adjusting that encryption in the event one layer is successfully breached. The self-regulated encryption method would be stronger across the national neural network, but isolated Sentinels would still be capable of generating a self-regulating encryption. The benefit for networked Sentinels would allow for stronger encryption capable of adjusting at greater speeds.

Corporate entities and the government are the two main targets of cyber criminals.[121] However, private citizens increasingly fall victim to instances of identity theft perpetrated online.[122] Sentinel would need to be integrated into all three categories to have the greatest effect. It stands to reason that corporate entities would be the best place to start with Sentinel integration; such entities are attractive targets for many cybercriminals.[123] Corporate Sentinels would not share any confidential materials – it is unlikely that a corporate Sentinel would need access to that information. The only information corporate Sentinels would share across the Internet would be defensive strategies to protect against unauthorized network access. This sharing would strengthen all participating corporate entities against theft of confidential materials and intellectual property. It would also serve to strengthen the national Sentinel Construct, and to better prepare it for a wartime scenario.

Private citizens also stand to benefit in unique ways from Sentinel integration. In this capacity, Sentinel would take on a more proactive approach than simply defense against intrusion, although that would still be one of its primary

[121] Paolo Passeri, *2015 Cyber Attacks Statistics*, HACKMAGEDDON (Jan. 11, 2016), http://www.hackmageddon.com/2016/01/11/2015-cyber-attacks-statistics/.

[122] Tamara E. Holmes, *Credit card fraud and ID theft statistics*, NASDAQ (Sept. 16, 2015), http://www.nasdaq.com/article/credit-card-fraud-and-id-theft-statistics-cm520388.

[123] Robert K. Ackerman, *Destructive Cyber Attacks Increase in Frequency, Sophistication*, ARMED FORCES COMM. & ELECTRONICS ASS'N: SIGNAL (July 1, 2015), http://www.afcea.org/content/?q=Article-destructive-cyber-attacks-increase-frequency-sophistication.

functions. Private identifying information, such as a social security numbers and bank records, are also attractive targets for identity thieves, who increasingly rely on technology to commit crimes.[124] Private Sentinels would be licensed to private citizens – much like a software license – and would be tasked by the licensee with keeping that data on file. Private Sentinels would be passive with this information by default, and Corporate or Business Sentinels would not have personal information on file to search for. However, Private Sentinels would be able to be tasked by licensees to search for improper use of personal information.

A cooperative effort among Sentinel programs might look something like the following scenario. Corporate Sentinels used by banks or businesses might be specially tasked with holding some customer information so as to alert them in the event of improper use. A Bank Sentinel might relay the issue to a Private Sentinel that the private licensee's personal information was used in a suspicious manner, perhaps at a store managed by a Business Sentinel. The Private Sentinel, at the direction of its licensee, would then be able to contact the Business Sentinel for verification of the unauthorized use. To ensure the proper channels are taken to resolve the issue, the Private Sentinel might then relay what it has found to local or federal authorities. The Business Sentinel might have access to the general business records, or it might be integrated across every individual store and possess additional information, like security camera footage. Throughout this scenario, each participating Sentinel program has learned how to best communicate with other Sentinels, proper methods for solving complex problems with roots in the physical world, and ways to defend against future identity theft.

Military Sentinel application, as well as Sentinel programs for law enforcement agencies, would mirror corporate application; the protection of sensitive information would be critical. Military Sentinels would be able to communicate with one another and with the national network, taking part in the larger defense sharing strategy. Also critical to Military Sentinel function would be prevention of corruption of vital systems, like communication and intelligence. The self-regulated encryption would assist in this capacity, but in the event of breach Sentinel might possess the ability to quarantine compromised systems for later repair. Alternatively, Sentinel might be programmed with – or learn the ability to – repair and restore damaged systems. The defense of hardware and software programs might also be implemented into the functionality of Private and Corporate Sentinels.

The word "installed" has been deliberately used to describe how Sentinel

[124] *Identity Theft*, JUSTICE.GOV, http://www.justice.gov/criminal-fraud/identity-theft/identity-theft-and-identity-fraud (last visited Oct. 28, 2016).

would find its way into a computer system. This is because a Sentinel program should only be voluntarily installed on networks by willing participants. The strength and efficacy of the Sentinel program would attract more citizens and corporate entities to install it, and the national Sentinel Construct could then grow stronger. As Sentinel learns and evolves, it might be capable of suggesting new defensive strategies not contemplated at the time of its initial implementation. Such suggestions might be offered to Sentinel handlers, who would assess whether or not to add that functionality into the Sentinel program. Authorized updates to Sentinel would then be shared across the network.

Sentinel would be a shield designed to stand between the United States and foreign cyberattacks. In peacetime, it would learn and adapt and evolve by assessing and countering both foreign and domestic threats. Sentinel would find ways to better itself, and would archive all manner of attacks made against the United States' citizens, corporate entities, and her military. This would not be the extent of its functionality. Activating the full range of the Sentinel Construct's abilities requires a second component in the digital nuclear option – a sword to Sentinel's shield.

b. Infiltrator

In the digital age, brute force and strength of numbers is no longer the only means to an end. While strategy and intelligence are vital components of kinetic warfare, cyberwar requires precision, stealth, and subterfuge. The Infiltrator AI Construct would be the offensive component in the digital nuclear option, but like Sentinel, Infiltrator would gain its power during peacetime – or rather *relative* peacetime – application.

Infiltrator would not serve different designations in the same way Sentinel would. There would be no dedicated civilian or corporate Infiltrator programs. Instead, Infiltrator would be most effective if it were restricted to military, national security, and federal law enforcement application. There are many departments, branches, and agencies which have overlapping authority and responsibilities where cybersecurity is concerned, and this can create problems in effectively responding to cyber-threats.[125] This would not be a concern with Sentinel, as Sentinel would be a mostly-passive Construct with broad access and application. Allowing all of these different authorities access to Infiltrator,

[125] Joeli R. Field, *Cybersecurity: Division of Responsibility in the U.S. Government* 6 (Nat. Sec. Cyberspace Inst., Paper INTL604), http://www.nsci-va.org/CyberReferenceLib/2010-09-18-Cybersecurity-Division%20of%20Responsibility%20in%20the%20US%20Government-Joeli%20Field.pdf.

however, would lead to further disjointed and confused results which could prove harmful to U.S. cybersecurity efforts. Infiltrator would require the guiding hand of a single organization which would be responsible for determining Infiltrator deployment.

The Infiltrator Construct's handler, as well as a handler for Sentinel, would require cooperation from the United States' intelligence apparatus and the military, but on a deeply integrated level. That could be accomplished in any number of ways. One solution would be to simply place control of Infiltrator with the National Security Agency. The NSA already has deployed a malware program similar to proposed Infiltrator functions for the purposes of infiltration, spying, and data collection.[126] It is also suspected to have participated in the Stuxnet cyberattack on Iran, and may be a major player in the enigmatic Equation Group.[127] Alternatively, the United States Cyber Command might be a better fit. As Infiltrator would be mainly deployed in the for the purpose of safeguarding U.S. interests abroad, the military might be in a better position to make these decisions. Presidential oversight would be very important to deployment of Infiltrator, and such oversight might be better effectuated with Infiltrator under military control as opposed to the NSA. Presidential and Congressional oversight would be necessary for the Infiltrator Construct's application due to its offensive nature, foreign deployments, and its role in possible cyberwar scenarios.

Whoever became ultimately responsible for Infiltrator and Sentinel oversight would be responsible for assessing threats reported by other cybersecurity departments within the government, as well as those presented by individual military cyber commands. Infiltrator would be deployed and assessed by the handlers during peacetime, but there would be no restriction on the sharing of information between Sentinel and Infiltrator.

Infiltrator, as the name suggests, would be designed to infiltrate foreign computer networks for purposes of threat assessment and, if necessary, threat removal. Infiltrator might gain access to a targeted network through innocuous means to mask its presence, such as by concealing its programming in an enticing spam email. Or it might try a more direct approach by forcing its way into a networked access point. Infiltrator would benefit from its own catalog of attack strategies compiled by cybersecurity agencies and departments, and could use that catalog to determine the most effective infiltration strategy, or to at-

[126] Joseph Menn, *Russian researchers expose breakthrough U.S. spying program*, REUTERS (Feb. 16, 2015), http://www.reuters.com/article/2015/02/16/us-usa-cyberspying-idUSKBN0LK1QV20150216.

[127] Thomas Fox-Brewster, *Equation = NSA? Researchers Uncloak Huge 'American Cyber Arsenal'*, FORBES (Feb. 16, 2015), http://www.forbes.com/sites/thomasbrewster/2015/02/16/nsa-equation-cyber-tool-treasure-chest/.

tempt a new method of gaining unauthorized access. Infiltrator would be able to analyze the integrity of targeted network security and to exploit weaknesses within the system. One of Infiltrator's greatest strengths would lie in its ability to communicate with Sentinel. Upon request by Infiltrator, Sentinel would share all or part of its catalog of defense strategies for Infiltrator to analyze. Sentinel could share weaknesses it identified in its own defense strategies, giving Infiltrator options to explore when faced with such a defensive tactic. Sentinel could also share attack strategies it has encountered, allowing Infiltrator to replicate, implement, and augment them. In turn, Infiltrator would share its attack catalog with Sentinel, allowing Sentinel to shore up its own defenses where Infiltrator has identified weaknesses.

Once inside a targeted system Infiltrator would behave similarly to Sentinel, only without the consent of the system owner or operator. Infiltrator might hide its own perceptrons in the systems it invades, increasing its processing power by creating covert artificial neural networks. The NSA has already deployed a similar program, although it appears to be heavily reliant on handler direction.[128] Across a large network of targeted systems, Infiltrator's processing speed and power would increase exponentially. Infiltrator would need to be cautious in this environment so as to avoid detection. This would likely require Infiltrator to spread itself out thinly across a greater network to avoid consuming suspicious amounts of processing power from its host system. Infiltrator could also possess a regenerative capacity. If Infiltrator was discovered on a system and purged save for one hidden component, it would be able to rebuild itself from that single remaining component. It may also be programmed with, or learn the ability to, prevent itself from being deleted at all – NSA malware already possesses this ability.[129] Infiltrator would benefit from the ability to learn from counterattacks to more effectively hide and embed its code into targeted systems.

Once Infiltrator has successfully infiltrated a system, it would then need to serve two functions. The first is intelligence gathering. Infiltrator would report back to its handlers on the information it has obtained, which would then give the handlers intelligence to adjust Infiltrator's operational parameters. The information transmitted by Infiltrator would need to be heavily encrypted and relayed covertly to handlers. Infiltrator would learn the best ways to do this while benefitting from the same self-regulating encryption used by Sentinel. The information relayed by Infiltrator might also help Sentinel to better aug-

[128] Ryan Gallagher, *Researchers Find 'Astonishing' Malware Linked to NSA Spying,* THE INTERCEPT (Feb. 17, 2015), https://firstlook.org/theintercept/2015/02/17/nsa-kaspersky-equation-group-malware/.

[129] *Id.*

ment its own defensive capabilities, allowing Sentinel to generate a response to future cyberattacks.

The second function of Infiltrator would be sabotage and disruption. Infiltrator would be able to receive directives from handlers on how to best dismantle an enemy intelligence network, but the need for handler input would diminish as Infiltrator learned which tactics to apply and when. Handler guidance might be more appropriate in delicate situations requiring specific disruption, such as the deletion or manipulation of data, whereas Infiltrator might be capable of carrying out widespread sabotage such as DOS attacks or malware-like intrusion on its own. Infiltrator might disrupt an enemy system by interfering with network connectivity, thereby preventing one system from communicating with another. It might corrupt enemy systems leading to loss of critical systems files. At the directive of handlers, it might replace critical information with falsified data in order to mislead enemy intelligence agencies. Sabotage and disruption are not limited to information. Stuxnet showed that attack programs are capable of causing physical damage to certain systems, and other malicious programs further demonstrate this power.[130] Infiltrator might consume massive amounts of processing power to overload and destroy a computer's hard disc drive, resulting in a loss of data. It might also interfere with other machine components operated remotely or by computer.

As part of its partnership with Sentinel, Infiltrator could have limited civilian, corporate, and law enforcement capability. In addition to keeping a catalogue of attempted intrusions, Sentinel might also pinpoint possible origins of the attacks. Law Enforcement Infiltrators would then trace such an attack back to its source, infiltrate the targeted network, and perform is intelligence gathering and sabotage functions. However, allowing private citizens and corporations to launch Infiltrator against a target would have dangerous repercussions. For instance, if an isolated group of hackers in Nation A redirect their activities through servers in Nation B, and a Corporate Infiltrator is unleashed on the unwitting Nation B, the United States could face backlash for sanctioning or impliedly permitting unprovoked cyberattacks. Infiltrator would require more hands-on direction than Sentinel, and so it would be unwise to allow the average citizen or corporations to have access to such a powerful tool. This problem might be overcome by granting federal law enforcement agencies access to Infiltrator for the purposes of coordinating with Sentinel to eliminate cybercrime and cyber-terrorism threats while leaving larger-scale application to the handlers.

After Infiltrator has completed an objective, it would not remove itself from

[130] Kim Zetter, *A Cyberattack Has Caused Confirmed Physical Damage for the Second Time Ever*, WIRED (Jan. 8, 2015), http://www.wired.com/2015/01/german-steel-mill-hack-destruction/.

a targeted system. Instead, the Infiltrator neural network would remain passive in the system, learning and evolving its capabilities based on the efforts of other, active Infiltrators around the world. Passive Infiltrators could reactivated by handlers, or might be spurred to automatic reactivation by a certain set of circumstances, and would then resume intelligence gathering and sabotage functions.

Infiltrator and Sentinel are simply components of a greater power. In peacetime they are refined and forged into powerful tools. When faced with war in the digital theater, these Artificial Intelligence Constructs would combine their efforts to ensure the safety and integrity of domestic digital systems. They would also lash out to stop continued assault, and render every available avenue of attack inert. The combined might of Sentinel and Infiltrator would not merely disrupt and sabotage enemy systems – it would annihilate them.

c. Retaliator Protocol

As discussed earlier, the United States has only ever deployed a kinetic nuclear device twice – both times against the same target to effectuate the same result.[131] This deployment was during wartime, not in response to a criminal or terrorist act.[132] The deployment was tactical, both a means to bring an end to the war in the Pacific Theater and to demonstrate to all the world a new and terrible power.[133] Also of significance is the fact that the United States did not strike first with their nuclear option – it was only deployed following the attack on Pearl Harbor.[134] A digital nuclear option should adhere to similar principles – it should only be deployed in a time of war and even then only in response to foreign aggression.

Retaliator Protocol does not refer to a distinct AI Construct. Rather, it would be a protocol enacted during a cyberwar scenario requiring deep integration between Infiltrator and Sentinel, essentially merging them into a new entity with a new purpose. Retaliator Protocol would be triggered either by a certain set of events or circumstances or at the direction of handlers, which would then activate restricted functions in both Sentinel and Infiltrator and task them to analyze the threat, prevent damage, and stop the initial attack. Once the primary threat had been neutralized, Retaliator Protocol would take the fight to the instigator and wreak havoc on their systems.

[131] *Atomic Bomb: History of WW2*, HISTORY.COM, http://www.history.co.uk/study-topics/history-of-ww2/atomic-bomb (last visited Oct. 31, 2016).

[132] *Id.*

[133] *Id.*

[134] *Attack on Pearl Harbor – 1941*, ATOMIC HERITAGE FOUND., http://www.atomicheritage.org/history/attack-pearl-harbor-1941 (last visited Oct. 31, 2016).

i. Retaliator Protocol: Sentinel Functionality

Much like its peacetime application, Sentinel would be tasked with defense and analysis. Cyberwar has the potential to damage critical systems nationwide, such as power grids, and some believe catastrophic attacks of this caliber are not only possible, but increasingly likely.[135] Using the national neural network, the national Sentinel Construct would be prepared to counter a cyberwar effort by tapping into the collective defense catalog to coordinate counter-intrusion strategies. Individual Sentinels from private citizens and corporate entities would be "drafted" to aid in the defensive effort – Retaliator Protocol would activate Sentinel programs across the nation without the authorization of private citizens or corporations. This would vastly increase the defensive capabilities and processing power and speed of Sentinel. Critical systems would benefit from the peacetime experiences of individual Sentinels. Honed defensive strategies and advanced self-regulating encryption would maximize counter-intrusion efforts. Communication with Infiltrator in peacetime might have allowed Sentinel to prepare itself for certain iterations of cyberattacks, such as specialized malware. System quarantine or repair and restore capabilities may have been adapted to more serious threats during peacetime as well.

Sentinel's defense in cyberwar would not stop at securing domestic systems. In peacetime, Infiltrator would be capable of tracing cyberattacks back to their sources, no matter how circuitous the route. In wartime, Infiltrator could eliminate hostile presence in these access points. Infiltrator could then install Sentinel in these neutralized systems, allowing Sentinel to move in and secure them against further intrusion. Once Infiltrator had gained access to the instigator systems – or if Infiltrator had been reactivated in these systems – Sentinel would be able to join Infiltrator abroad, securing any networks holding Infiltrator neural components and freeing Infiltrator to perform its own Retaliator Protocol functions, a function akin to capturing and holding territory.

ii. Retaliator Protocol: Infiltrator Functionality

In a Retaliator scenario, Infiltrator would be responsible for generating a level of destruction relatively comparable to a nuclear device detonated in cyberspace. Whichever event triggers the Retaliator Protocol, Sentinel and Infiltrator would need to move swiftly to ensure both domestic security and that the avenues of the initial attack are disabled. This would require a response time far beyond what human handlers could manage. This first Retaliator stage would be automated, without human oversight. Once completed, Retaliator Protocol would move to a standby mode to await input from handlers. In this

[135] Smith, *supra* note 102.

way, an automated system would not be able to wage an independent war against a foreign nation without human approval. A standby mode would allow the President and Congress to determine the appropriateness of proceeding to cyberwar, or if the situation should be de-escalated.

In the event of cyberwar, handler input is vital and allows for humans to designate specific targets. Depending on how long the cyber conflict lasts, Infiltrator might learn strategies on its own, and be able to implement them as its analysis of battlefield conditions evolves. Infiltrator's intelligent intrusion methods and cooperation with the wartime Sentinel would open up a wide array of targets that might be inaccessible in kinetic warfare. The physical damage done by overloading computer systems would shut down avenues of foreign cyberattacks. Flooding systems with Infiltrator neural processing components or junk code could have the same effect by consuming processing power from critical functions while greatly increasing Infiltrator's power. It could also erase data from any and all systems in which it has taken root, freeing up even more processing power while destroying information. Inside a foreign military network, Infiltrator could interfere or completely disable communications. Any automated or computer-assisted functions could be disabled, greatly reducing military effectiveness. Infiltrator could work its way into targeted civilian computer networks and permanently disable power grids, air traffic control, water treatment plants, and other critical systems across the country. The objective of Infiltrator in a Retaliator scenario would be permanently disabling or destroying computer systems in the hostile nation.

Infiltrator would have the ability to generate kinetic results as well, and this outcome only becomes more plausible with the advent of such technology as self-driving cars and civilian sector drones. As for existing automated technology, satellites orbiting in space are dependent upon remote communication from a computer.[136] Not only would Infiltrator be capable of commandeering satellites for intelligence gathering purposes, but hijacking satellite maneuvering ability could lead to catastrophic kinetic results. Once in control of an enemy satellite, Infiltrator could set its orbit to terminate on a high value target, such as a military installation or an aircraft carrier. Such an outcome might be more easily achieved by infiltrating weapon systems on hostile military drones or warships and using the enemy's own weapons against them. In all likelihood, launch codes for a targeted nation's nuclear arsenal would not be beyond the reach of Retaliator Protocol. More complex maneuvers and those with long-lasting consequences would almost certainly require the guidance of a

[136] Raymond Guzman & N.A. Chu, *Taking Remote Control of Satellite Communications*, SIGNAL (Sept. 1, 2016), http://www.afcea.org/content/?q=Article-taking-remote-control-satellite-communications.

handler.

The kinetic destruction would not necessarily be an integral part of the Retaliator Protocol; it would most likely be left to the discretion of Infiltrator handlers, the President, Congress, and military advisors. Even without kinetic damage, the result of the Retaliator Protocol would be rapid and total destruction of the computer network of a hostile country. Military computer networks would be either burned out from forced overclocking or flooded with junk code to consume processing power by Infiltrator requiring, at a minimum, complete system overhaul. In the private sector, if such targets are designated, power grids and computer systems could be permanently shut down. Banking and financial records could be wiped and the systems rendered useless, which might lead to an economic collapse. Above all, communication networks within the country would be left in disarray. Infiltrator might be programmed with or evolve the capacity to commandeer cellphone and radio towers and produce jamming signals.

Retaliator Protocol would be devastating against a technology-dependent target nation. Critical infrastructure would be crippled, perhaps irreparably so without obscenely expensive overhauls. Lives would inevitably be lost. Infiltration, data mining, counter-intelligence, sabotage, and other complex tasks could be accomplished within a matter of weeks, days, or even hours depending on the response time from human handlers.

iii. Retaliator Protocol: Aftermath

The digital systems and networks of the targeted nation would be left unusable, either by physically damaging them or filling them with un-erasable junk code and malicious programs. While the image would not be as visceral as a smoldering mushroom cloud looming miles into the sky over scorched earth, the aftermath would be strongly felt in a different way. Any machinery operated by a computer would be either useless without a complete overhaul, or still operable but with severely limited functionality. Internet access might be severely limited, or cut off entirely. Telephones and cellphones might be incapable of making calls. Bank records and accounts could be wiped. Without working communications networks, the targeted government would be unable to prevent society from falling into disorder and chaos. The targeted military would face similar problems with communications blackouts and compromised infrastructure.

The devastation wrought upon the targeted nation, with or without resulting kinetic damage, would stop the cyberattack and prevent the possibility of a coordinated kinetic response. It would dissuade them from trying again. And, perhaps most importantly, it would let other cyber-capable nations know of the existence of a new superweapon – one capable of bringing a sovereign nation to its knees without firing a single shot.

IV. Sentinel and Infiltrator Classification

As powerful as the proposed Sentinel and Infiltrator Constructs could be, they would not need to rise to a level beyond an Artificially Narrow Intelligence. Both Constructs would be restricted in functionality by the limitations in their programming, even though their programming would enable them both to learn, as well as a significant degree of freedom in self-enhancement. However, both Constructs would only be able to learn and enhance themselves in limited ways.

For example, Sentinel might possess the ability to learn, but it would only learn better ways to protect itself from outside threats to systems it is designated to protect. Sentinel would be programmed to allow handler access, and would be prevented from augmenting itself in such a way as to deny such access. Infiltrator could be programmed with the same limitations, allowing handlers to have unfettered access to both Constructs' core programming. Infiltrator could also be programmed with a sophisticated friend-or-foe identification system, preventing it from spreading into and disrupting networks designated as friendly.

What would also place these Constructs in the ANI category is their inability to think or behave humanly. Sentinel and Infiltrator would be responding to stimuli or handler direction in peacetime and wartime scenarios, with their responses dictated by their programming. At no point would Sentinel be able to change a user's desktop background because it prefers blue over green. Likewise, Infiltrator would not be able to launch a Distributed Denial of Service (DDoS) attack against a Planned Parenthood clinic because it believes abortion is murder. Instead, these programs would think and behave rationally based on the limitations of their programming or in response to stimuli which then triggers a preprogrammed or learned response. That is not to say there are no dangers inherent in the use and application of ANI, and in fact they may pose the most serious danger to modern global cybersecurity.

INTERNATIONAL LAW AND THE DIGITAL FRONTIER

The implementation of Artificial Intelligence on such a scale as the proposed Sentinel and Infiltrator programs has never been done before. It will likely be a very long time before any such project sees widespread application, if one ever does. While this concept remains largely in the realm of imagination, it is helpful to consider how the international community might respond to such a significant event. Although we cannot examine similar responses from the international community regarding this particular subject matter, it can be instructive to examine how the world leaders of the past approached and resolved international issues that generated similar concerns. Volumes of work can be – and have been – dedicated to exploring the existential and metaphysical implications of AI. This section of this article will instead focus more specifically on the international legal implications of using artificial intelligence as a weapon of mass destruction in cyberspace.

I. International Efforts to Reduce Unnecessary Suffering in War

a. Saint Petersburg Declaration of 1868

In 1863, Russia successfully developed a musket ball capable of detonating upon impact with hard targets.[137] The intended purpose of this weapon was to allow infantry to destroy the supplies and ammunition of hostile forces.[138] The success of this new weapon spurred further research and development into the technology until, in 1867, the exploding musket ball was refined into an antipersonnel weapon capable of detonating upon impact with the softer masses of enemy soldiers.[139]

19th-century Russia's government was unwilling to implement this new technology in wartime scenarios, and sought the cooperation of the international community to prohibit the use of this and similar weapons from the theater of war.[140] In 1868, the Imperial Cabinet of Russia met with world leaders at Saint Petersburg in the Russian Empire to formally agree to the prohibition in what would become known as the Saint Petersburg Declaration, or the *Declaration Renouncing the Use, in Time of War, of Explosive Projectiles Under 400 Grammes Weight*.[141] Signatories to the declaration included the United Kingdom, Italy, France, and Switzerland.[142] The nations of Brazil, Estonia, and

[137] Burrus M. Carnahan, *The Civil War Origins of the Modern Rules of War: Francis Lieber and Lincoln's General Order No. 100*, 39 N. KY. L. REV. 661, 678 (2012).

[138] *Id.*

[139] *Id.*

[140] *140th Anniversary of the 1868 St. Petersburg Declaration*, INT'L COMM. OF THE RED CROSS (Nov. 28, 2008), https://www.icrc.org/eng/resources/documents/statement/st-petersburg-declaration-281108.htm [hereinafter *140th Anniversary*].

[141] Declaration Renouncing the Use, in Time of War, of Explosive Projectiles Under 400 Grammes Weight, Nov. 29/Dec. 11, 1868 [hereinafter St. Petersburg Declaration], *reprinted in* HUMANITARIAN LAW: SELECTED DOCUMENTS 303-04 (Donja de Ruiter ed., 2011) [hereinafter St. Petersburg Declaration].

[142] *Declaration Renouncing the Use, in Time of War, of Explosive Projectiles Under 400 Grammes Weight, Nov. 29/Dec. 11, 1868*, INT'L COMM. OF THE RED CROSS, https://ihl-

Baden, not in attendance at the signing of the declaration in Saint Petersburg, later ratified it.[143]

To paraphrase, the Saint Petersburg Declaration set forth the following: it is the responsibility of the international community to lessen the effects of war, the objective of which should be to weaken the strength of the hostile military force.[144] While disabling enemy soldiers is an acceptable means to achieve that objective, it is not acceptable to compound the suffering of enemy soldiers.[145] To that end, a weapon of this sort – explosive anti-personnel ordnance – is "contrary to the laws of humanity."[146]

The Saint Petersburg Declaration was the first international agreement of its kind, specifically prohibiting the use of certain types of weapons during war.[147] Although the international community of the nineteenth century was very different from today's complex realm of international relations, it is still instructive to consider this swift response to a developing weapon technology on the global stage. Within one year of successfully developing the anti-personnel variant of the exploding musket ball, the Russian government moved to eliminate its use in war.[148] Major European world powers assented to this prohibition before the anti-personnel variant could ever see actual use in a wartime scenario.[149] The Saint Petersburg Declaration demonstrated the ability of world powers to come together and address a known entity; in this instance, the unnecessary aggravation of human suffering during war.

b. The Geneva Protocol

The Protocol for the Prohibition of the Use in War of Asphyxiating, Poisonous or Other Gases, and of Bacteriological Methods of Warfare, or the Geneva Protocol of 1925, is an international treaty created to bind its signatories in agreement against using chemical or biological weapons in times of war.[150] The

databases.icrc.org/applic/ihl/ihl.nsf/States.xsp?xp_viewStates=XPages_NORMStatesParties&xp_treatySelected=130 (last visited Nov. 10, 2016).

[143] *Id.*

[144] St. Petersburg Declaration, *supra* note 141.

[145] *Id.*

[146] *Id.*

[147] *140th Anniversary*, *supra* note 140.

[148] Carnahan, *supra* note 137, at 678-79.

[149] *140th Anniversary*, *supra* note 140.

[150] Protocol for the Prohibition of the Use of Asphyxiating, Poisonous or Other Gases, and of Bacteriological Methods of Warfare, June 17, 1925, 26 U.S.T. 571 [hereinafter Geneva Protocol], *available at* http://www.state.gov/t/isn/4784.htm#treaty.

Geneva Protocol reaffirmed the 1919 Treaty of Versailles, which prohibited the use of chemical weapons, and forbade Germany from manufacturing and importing them.[151] It also upheld Article 5 of the Treaty relating to the Use of Submarines and Noxious Gases in Warfare, also known as the 1922 Treaty of Washington, which never entered into force.[152] The Geneva Protocol contains specific language from both the Treaty of Versailles and the Treaty of Washington, which was subsequently modified in the Geneva Protocol and reads in part, "Whereas the use in war of asphyxiating, poisonous or other gases, and of all analogous liquids materials or devices, has been justly condemned by the general opinion of the civilized world[.]"[153]

World War I demonstrated the terrible power of chemical weaponry. Gas attacks accounted for 30% of military casualties during the war.[154] Death by gas-based weaponry was slow and painful, and many survivors carried physical and psychological scars throughout their lives.[155] Chemical warfare in World War I also claimed the lives of many civilians, who were largely unprotected when winds carried the harmful chemical agents off the battlefields and into civilian homes.[156] The rationale of the Saint Petersburg Declaration can be seen as an underlying current influencing the Geneva Protocol, its predecessors, and subsequent international treaties which prohibited the proliferation of certain types of weapons.

II. International Humanitarian Law

As previously mentioned, the objective of war should be to weaken the strength of the hostile military force.[157] While disabling enemy soldiers is an acceptable means to achieve that objective, it is not acceptable to compound their suffering. Deploying a weapon system that is difficult, if not impossible to effectively control, and which can cause unnecessary suffering in both hostile forces and civilians alike is certainly contrary to the laws of humanity.

a. The Hague Conventions

In 1899 world leaders gathered in The Hague in the Netherlands, summoned

[151] *Id.*

[152] *Id.*

[153] *Id.*

[154] Gerald J. Fitzgerald, *Chemical Warfare and Medical Response During World War I*, 98 AM. J. PUB. HEALTH 611, 619 (Apr., 2008), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2376985/pdf/0980611.pdf.

[155] *Id.* at 620-21.

[156] *Id.* at 616.

[157] St. Petersburg Declaration, *supra* note 141.

by Tsar Nicholas II of Russia's minister of foreign affairs, to participate in the first Hague Convention, or the *Convention (II) with Respect to the Laws and Customs of War on Land*.[158] The 1899 Hague Convention was successful in setting forth conventions to govern the laws of war on land and sea, and was ratified by several major world powers, including the United States.[159] There are many provisions to the 1899 Hague Convention but of note are Articles 22 and 23 of the Annex to the Convention.[160] Article 22 explicitly states that the "right of belligerents to adopt means of injuring the enemy is not unlimited."[161] Article 23 provides some clarification, prohibiting "arms, projectiles, or material of a nature to cause superfluous injury."[162] The language of these articles expresses the intent of the signatories to prohibit weapons of a similar type specifically enumerated in the accompanying three Declarations. One Declaration prohibited deploying explosives from balloons.[163] A second prohibited the use of projectiles designed to disperse asphyxiating gases.[164] Finally, a third Declaration prohibited the use of *dum dums*, or projectiles which change shape inside the human body.[165]

Both the Geneva Protocol of 1925 and the Hague Conventions of 1899 and 1907 represent significant steps taken by the international community to codify international humanitarian law. One of the notable focuses of both these conventions was a desire by a majority of the signatories to minimize human suffering on the battlefield.[166] These conventions were designed with military personnel in mind, and did not give much consideration to the safety or wellbeing of civilians in war. The need for civilian protection, and more comprehensive regulation on the conduct of war, would be addressed at length during the Geneva Conventions and their accompanying Protocols following World War II.

[158] *Hague Convention*, BRITANNICA.COM, https://www.britannica.com/topic/Hague-Conventions (last visited Sept. 19, 2016).

[159] *Id.*; *Convention With Respect to the Laws and Customs of War on Land*, OVERHEID.NL: TREATY DATABASE https://verdragenbank.overheid.nl/en/Verdrag/Details/002338 (last visited Nov. 1, 2016).

[160] Convention with Respect to Laws and Customs of War on Land (Hague II) art. 22, Jul. 29, 1899, 32 Stat. 1803 [hereinafter Hague II].

[161] *Id.*

[162] *Id.*

[163] *Hague Convention, supra* note 158.

[164] *Id.*

[165] *Id.*

[166] The United States did not ratify any of the three Declarations of the 1899 Hague Convention, although it did ratify the renewed Declaration to ban explosives deployed by balloons. Hague II, *supra* note 160.

b. The Geneva Conventions of 1949 and Their Additional Protocols

The conventions, protocols, and treaties discussed above share a common theme. While the stated objective can be boiled down to reducing unnecessary human suffering during war, these international agreements have been chiefly concerned with lessening the suffering of combatants. Until the Geneva Convention, civilian casualties had not been given significant consideration. The Geneva Convention – specifically, the Fourth Geneva Convention and the Protocol I – addresses the necessity to safeguard the lives and wellbeing of civilian populations involved in wartime hostilities.[167]

i. The Geneva Convention Relative to the Protection of Civilian Persons in Time of War of August 12, 1949 (The Fourth Geneva Convention)

The First and Second Geneva Conventions provide for the care of wounded and sick members of hostile armies and navies.[168] The Third Geneva Convention provides for the care of prisoners of war.[169] The Fourth Geneva Convention's provisions focus on ensuring the protection and humane treatment of civilian populations during wartime hostilities.[170] The Fourth Convention's Articles share the common theme of lessening unnecessary human suffering; for example, Articles 13 forbids murder, torture, and brutality.[171] Articles 33 and 34 forbid pillaging, reprisals, and indiscriminate destruction of property.[172] Convention IV goes further in requiring occupying forces to protect civilian hospitals (Article 18),[173] respect local and religious customs (Article 27),[174] and allowing public officials of occupied territory to continue their duties (Article 64).[175]

The Fourth Geneva Convention goes a step beyond avoiding unnecessary human suffering as a result of hostilities in war; it requires occupying forces to maintain healthy and safe environments for civilians to the extent possible.[176]

[167] Convention Relative to the Protection of Civilian Persons in Time of War arts. 3-4, Aug. 12, 1949, 6 U.S.T. 3516.

[168] *Id.*

[169] *Id.* (this definition of "protected persons" encompasses prisoners of war under the conventions).

[170] AMERICAN RED CROSS, SUMMARY OF THE GENEVA CONVENTIONS OF 1949 AND THEIR ADDITIONAL PROTOCOLS 4 (2011), http://www.redcross.org/images/MEDIA_CustomProductCatalog/m3640104_IHL_SummaryGenevaConv.pdf [hereinafter AMERICAN RED CROSS].

[171] *Id.* at 3.

[172] *Id.* at 4.

[173] *Id.*

[174] *Id.*

[175] *Id.*

[176] *Id.*

This could pose a significant obstacle to the implementation of a CWMD. In the modern age, nations that make use of computer technology and the Internet have developed a dependency on this technology. Cyber-attacks targeted at civilian communications and Internet access might be seen as violations of the Articles of the Fourth Geneva Convention. Articles 79 through 135 require generally that civilians be permitted to live their lives without interference, unless security concerns require otherwise.[177] If such security concerns do exist, internment under humane conditions is acceptable.[178] However, in a cyberwar scenario there is no need for a physical military occupation in a targeted nation and thus no need for direct interference with civilian lives. With computer technology being such an integral component to the people of tech-savvy nations, a significant or total obstruction to those citizens' ability to use their technology without the requisite security concerns would violate the relevant Articles.

ii. Protocol I

Protocol additional to the Geneva Conventions of August 12, 1949, and Relating to the Protection of Victims of International Armed Conflicts (Protocol I) expanded the protections established for civilians and aid workers in international conflicts.[179] Here exist several provisions that have prohibitions that could directly obstruct the application of a CWMD during a wartime scenario. Of note, Articles 51 and 54 prohibit indiscriminate attacks on civilians and sources of food and water.[180] Water treatment facilities have seen a recent trend towards automation, making them viable targets for cyber-threat actors.[181] The same is true of food processing industries.[182] Articles 51 and 54 also prohibit the destruction of "other materials needed for survival."[183] One could make the argument that this could include power plants and oil pipelines in the 21st century.[184]

[177] *Id.*

[178] *Id.*

[179] *Id.* at 5.

[180] *Id.*

[181] Jim Turner, *Industry Trends Drive Need For Increased Automation, Information Technology*, WATER WORLD, http://www.waterworld.com/articles/print/volume-21/issue-12/automation-technology/industry-trends-drive-need-for-increased-automation-information-technology.html (last visited Sept. 21, 2016).

[182] Jenni Spinner, *Adept: automation fuels the future of food*, BAKERYANDSNACKS.COM (Dec. 17, 2013), http://www.bakeryandsnacks.com/Processing-Packaging/Automation-essential-to-food-manufacturing-success.

[183] AMERICAN RED CROSS, *supra* note 170, at 5.

[184] *See* Andrew Follett, *Reports: U.S. Very Vulnerable To A Mass-Blackout EMP Attack*,

Article 56 prohibits attacks on nuclear generating stations and dams.[185] Stuxnet proves that even facilities dealing in the processing of nuclear materials are not immune to cyber-attacks.[186] It stands to reason that dams are similarly vulnerable. A cyber-attack of sufficient sophistication designed to destabilize a facility of this kind could, in theory, have the same disastrous effects as the oil pipeline attack, but with a categorically higher potential to inflict harm.

Article 35 of Protocol I is also worth mentioning.[187] It prohibits the use of weapons that "cause superfluous injury or unnecessary suffering" and prohibits "means of warfare that cause widespread, long-term, and severe damage to the natural environment."[188] This would likely include such weaponry as the anti-personnel exploding musket ball outlawed by the Saint Petersburg Declaration. It can also easily apply to kinetic weapons of mass destruction. Nuclear weapons have the power to cause superfluous injury and unnecessary suffering, and they inflict the long-lasting environmental damage also prohibited in Article 35.

An artificially intelligent weapon system could conceivably inflict this same kind of damage by manipulating the control systems for nuclear weapons, but the AI itself – as intangible code in cyberspace – would be incapable of inflicting the prohibited harms directly. It could also be directed to avoid inflicting the prohibited harms, or even directed to avoid harming civilians in any instance. This creates a grey area in the established law. Weapons that do inflict the prohibited harms are, themselves, prohibited. It would be prudent to address whether the same restrictions should apply to weapons that can inflict the prohibited harms, but can be directed not to do so.

iii. Restrictions on Weapons of Mass Destruction

Implementation of 19th-century Russia's anti-personnel application of the exploding musket ball technology may very well have been contrary to the

THE DAILY CALLER (May 18, 2016), http://dailycaller.com/2016/05/18/reports-u-s-very-vulnerable-to-a-mass-blackout-emp-attack/ (discussing the risks posed to U.S. society if an EMP device was ever successfully executed, creating widespread power outages across the United States). *See also* Chris Rhodes, *What Happens When the Oil Runs Out?*, OILPRICE.COM (Jul. 29, 2013), http://oilprice.com/Energy/Crude-Oil/What-Happens-When-the-Oil-Runs-Out.html (examining how widespread the use of oil truly is worldwide. For example, oil is used not only for transportation, but for the production of clothes, clothes, computers, pharmaceuticals, and food).

[185] AMERICAN RED CROSS, *supra* note 170, at 5.

[186] Gary Brown & Keira Poellet, The Customary International Law of Cyberspace, 6 STRATEGIC STUD. Q. 126, 130 (2012) (describing how a U.S. led cyber-attack successfully disabled and damaged an oil pipeline in the Soviet Union in 1982).

[187] AMERICAN RED CROSS, *supra* note 170, at 5.

[188] *Id.*

laws of humanity, but firearms and their ordnance typically fall below the threshold to be categorized as weapons of mass destruction. 50 U.S.C. §2302 defines weapons of mass destruction as "any weapon or device that is intended, or has the capability, to cause death or serious bodily injury to a significant number of people through the release, dissemination, or impact of – (A) toxic or poisonous chemicals or their precursors; (B) a disease organism; or (C) radiation or radioactivity."[189]

The nuclear bombs dropped on Nagasaki and Hiroshima during World War II are two examples of weapons of mass destruction which utilized radiation.[190] Chemical and biological WMDs have also seen use throughout recent history. As early as the 1760's, British military officers in Colonial America used blankets infected with smallpox virus – *disease organisms* – to "extirpate" Native American populations.[191] And during World War I, chemical warfare was fairly commonplace with both Allied and Central Forces. Chlorine, phosgene, and tear-gas were used to great effect, and the United Kingdom even began developing an arsenic-based chemical weapon called the M Device before the war's end.[192] The proliferation of these devastating weapons has naturally been met with resistance, and world leaders have come together numerous times to address concerns over the continued use of weapons of mass destruction.[193]

III. Legality of Nuclear Weapons and Non-Proliferation

August 6th and August 9th, 1945, are the only two dates in the history of humankind where nuclear weapons have been utilized in armed, international conflict.[194] The international community has continuously debated the use of nuclear weapons ever since. The raw power nuclear weapons unleash, combined with the long-lasting and deleterious effects of nuclear radiation on hu-

[189] 50 U.S.C. § 2302 (2012).

[190] *Bombings of Hiroshima and Nagasaki – 1945*, ATOMIC HERITAGE FOUND., http://www.atomicheritage.org/history/bombings-hiroshima-and-nagasaki-1945 (last visited Sept. 23, 2016).

[191] Harold B. Gill, Jr., *Colonial Germ Warfare*, HISTORY.ORG, http://www.history.org/Foundation/journal/Spring04/warfare.cfm (last visited Sept. 23, 2016).

[192] John Simpkins, *Poison Gases*, SPARTACUS EDUC., http://spartacus-educational.com/FWWgas.htm (last visited Sept. 23, 2016).

[193] *See Weapons of Mass Destruction*, N. ATL. TREATY ORG., http://www.nato.int/cps/en/natohq/topics_50325.htm (last updated June 28, 2016) (discussing various meetings throughout the course of the 20th century addressing use of Weapons of Mass Destruction).

[194] *See* Katherine Iliopoulos, *Hiroshima and Nagasaki Anniversary: A Call To Disarm*, CRIMES OF WAR (last visited Oct. 1, 2016) (discussing the atomic bombings of the Japanese cities of Hiroshima and Nagasaki, which ended World War II).

mans and the environment, make them a hotly contested subject.[195] Almost immediately after the United States deployed their nuclear weapons against Japan, the United Nations called for a full stop on the further development and use of nuclear weapons.[196] No such agreement has been achieved by the totality of the international community, but nuclear and non-nuclear nations alike continue to participate in the ongoing discussion.[197]

a. Treaty on the Non-Proliferation of Nuclear Weapons (1968)

The 1960's saw enough development in nuclear weapons and power technology to spur the international community to take action. For a time this technology was exclusively available to the global superpowers; the U.S.A., the United Kingdom, the Soviet Union, as well as France and the Peoples' Republic of China.[198] This exclusivity would be short-lived. The processes used to generate nuclear power circulated in academic journals, and the materials required for the process became cheaper.[199] Eventually even unstable and developing nations were able to develop this technology on their own.[200] By the end of the decade, enough nations either had developed or possessed the capability to develop nuclear weapons to give cause for concern.[201] As between the five superpowers, nuclear weapons acted as a sufficient deterrent by means of the doctrine of mutually assured destruction.[202] However, the doctrine of deterrence only works when the parties in conflict both have access to nuclear weapons

[195] *See generally The Effects of Nuclear Weapons*, CAMPAIGN FOR NUCLEAR DISARMAMENT (last visited Oct. 1, 2016) (stating that the world's knowledge of nuclear weapon technology comes from the atomic bombings that ended World War II, atmospheric nuclear testing, as well as nuclear accidents).

[196] *See The Nuclear Non-Proliferation Treaty (NPT), 1968*, ATOMIC HERITAGE FOUND., http://www.atomicheritage.org/history/nuclear-non-proliferation-treaty-npt (last visited Oct. 1, 2016) [hereinafter *NPT 1968*] (providing an overview of the Nuclear Non-Proliferation Treaty which successfully helped slow the development of nuclear weapons during the Cold War).

[197] *Id.* (concluding that although the international community of nations has not universally adopted the NPT, it is a more desirable regime than independent state actors continuing to develop unfettered access to nuclear weapons).

[198] DEP'T OF ST.: OFF. OF THE HISTORIAN, TREATY ON THE NON-PROLIFERATION OF NUCLEAR WEAPONS 1 n.1 (2010), http://www.state.gov/documents/organization/141503.pdf [hereinafter DEP'T OF ST. TREATY].

[199] *NPT 1968*, *supra* note 196.

[200] *See* Akhilesh Pillalamarri, *India's Nuclear-Weapons Program: 5 Things You Need to Know*, THE NAT'L. INTEREST (Apr. 22, 2015), http://nationalinterest.org/feature/indias-nuclear-weapons-program-5-things-you-need-know-12697 (discussing India's post-colonial development of nuclear weapons in its conflicts with Pakistan, and the subsequent imposition of sanctions against India).

[201] *NPT 1968*, *supra* note 196.

[202] *Id.*

systems; the doctrine is compromised when many smaller nations with "volatile border disputes" possess the same technology.[203]

1n 1968, members of the United Nations signed the Treaty on the Non-Proliferation of Nuclear Weapons, and it entered into force in 1970.[204] Nuclear powers and non-powers alike agreed to its provisions.[205] Although it contains eleven Articles, the objectives of the treaty can be distilled into three pillars. The first pillar establishes the non-proliferation objective through Articles I, II, and III.[206] It prohibits the transfer of nuclear weapons and technology by nuclear-capable states. This first pillar also prohibits non-nuclear states from receiving or developing nuclear weapons and technology.[207] The second pillar establishes the objective of promoting peaceful uses for nuclear technology through Article IV.[208] This pillar promotes the development of nuclear technology for peaceful purposes and encourages international cooperation to that effect.[209] The third pillar encourages the disarmament objective through Article VI; it requires its signatories to work in good faith towards ending the nuclear arms race and striving for disarmament.[210]

The treaty was initially designed to remain in effect for twenty-five years, but it was extended indefinitely in 1995.[211] The number of nations who agree to be bound by this treaty has grown from 43 at the time it entered into force, to nearly 190, and it is "the most widely adhered to nonproliferation or arms control agreement in history."[212] As such, it has been largely successful in achieving its nonproliferation objectives and in maintaining global security. [213]

It is not, however, a perfect solution. Several obstacles highlight how some threats can slip through the cracks. One of the most significant obstacles is outright non-compliance. While this is a problem faced by most regulations, the stakes are decidedly higher when the regulation is a nuclear non-proliferation treaty as opposed to regulations on, for example, withholding employee income tax. Some well-known nations who have or continue to disregard the

[203] *Id.*

[204] DEP'T OF ST. TREATY, *supra* note 198, at 1.

[205] *See Treaty on the Non-Proliferation of Nuclear Weapons*, UNODA, http://disarmament.un.org/treaties/t/npt (last visited Oct. 31, 2016) (listing the signatories to the NPT, including non-nuclear nations such as Afghanistan, Austria, and Belgium).

[206] DEP'T OF ST. TREATY, *supra* note 198, at 1-2.

[207] *Id.* at 3.

[208] *Id.*

[209] *Id.*

[210] *Id.*

[211] *Treaty on the Non-Proliferation of Nuclear Weapons (NPT)*, UNODA, http://www.un.org/disarmament/WMD/Nuclear/npt (last visited Sept. 23, 2016).

[212] DEP'T OF ST. TREATY, *supra* note 198, at 4.

[213] *Id.*

treaty's objectives are Iran, Iraq, and North Korea.[214] Article X of the treaty allows for signatories to withdraw under certain circumstances.[215] Abuse of this withdrawal clause also presents compliance concerns. For instance, a signatory that is already in violation of the treaty might abuse the withdrawal clause to avoid sanctions.[216] This clause might also allow a compliant signatory to withdraw from the treaty and still remain in possession of otherwise-prohibited nuclear materials.[217]

b. International Court of Justice Advisory Opinion

The Non-Proliferation Treaty (NPT) provides very useful guidance in the development of nuclear technology and for disarming nuclear weapons. While it does prohibit certain uses and applications of nuclear technology, the language of the NPT avoids directly addressing the issue of whether or not the possession, use, and threat of using nuclear weapons is legal. In 1996, the International Court of Justice (ICJ) released an advisory opinion to address this vital concern.

c. International Law and Cyberspace

The international laws of war advance several critical humanitarian objectives.[218] Such objectives include the reduction of suffering of combatants and the protection of civilians in the theater of war. Another is halting the proliferation and prohibiting the use of weapons that cause unnecessarily severe damage to the environment and injury to combatants and civilians. There are other critical objectives advanced by the international law of war, but for the purposes of this discussion, the two described above provide an adequate framing device to analyze the issue of weapons of mass destruction in cyberspace.

i. On Cyberwar

Presently there is no cyberspace equivalent to the Geneva or Hague Conventions, or the Saint Petersburg Declaration. The international community has

[214] *Id.* at 9-10.

[215] Treaty on the Non-Proliferation of Nuclear Weapons art. 10 July 1, 1968 21 U.S.T. 483 729 U.N.T.S. 169.

[216] DEP'T OF ST. TREATY, *supra* note 198, at 12.

[217] *Id.*

[218] *See generally War and international humanitarian law*, ICRC.ORG, https://www.icrc.org/eng/war-and-law/overview-war-and-law.htm (last visited Oct. 16, 2016) (providing an overview of international humanitarian law which "aims to limit the effects of armed conflicts for humanitarian reasons.").

come together to address cybercrime and while there has been no shortage of scurrilous cyberspace activity to warrant such attention in recent years, cyberwar does not appear to have generated the same volume of concern.[219] The existing law of war evolved around traditional, kinetic warfare and will not always perfectly govern cyberspace, but application of existing law combined with developing international custom can provide valuable guidance.

Earlier in this article, war was considered generally to be the sum of actions taken by one sovereign nation against another with the intent to undermine the integrity of the targeted sovereign.[220] War as a legal concept does not have one all-encompassing definition, but guidance from multiple legal bodies provides that war may be "associated with a State's use of force to vindicate its rights … under international law."[221] Although the United Nations Charter prohibits the threat or use of force in instances of aggression, use of force is authorized in other scenarios such as self-defense.[222] However, in effectuating a self-defense response, the principles of necessity and proportionality remain in effect.[223]

As for war in cyberspace, some have argued that there can be no such thing.[224] Thomas Rid, a professor and scholar on international law and war, and his supporters argue that harmful actions in cyberspace alone amount to sabotage, albeit efficient and aggressive sabotage, but not grievous enough to rise to the level of war.[225] However, if a cyber-attack that results in physical property damage that leads to loss of life would be an exception to this rule and rise

[219] Andreas Zimmermann, *International Law and 'Cyber Space'*, 3 EUR. SOC'Y OF INT'L L. 1, 4, 6 (2014) (discussing the lack of international law regarding the governance of cybersecurity due to the fact that present cyber-attacks are still seen by the international community as non-violent).

[220] *See generally The Peace of Westphalia and Sovereignty*, BOUNDLESS.COM, https://www.boundless.com/world-history/textbooks/boundless-world-history-i-ancient-civilizations-enlightenment-textbook/the-rise-of-nation-states-1052/nation-states-and-sovereignty-1053/the-peace-of-westphalia-and-sovereignty-1055-17653 (last visited Nov. 2, 2016) (describing the Treaty of Westphalia which recognized nations as equal sovereigns that have control over their own domestic affairs without the interference of other outside powers. This system of international relations did not end the concept of war, but largely reduced war to conflicts between nation state actors).

[221] U.S. DEP'T OF DEFENSE, DEPARTMENT OF DEFENSE LAW OF WAR MANUAL 18 (2015), http://www.defense.gov/Portals/1/Documents/pubs/Law-of-War-Manual-June-2015.pdf.

[222] *Id.*

[223] Harold Hongju Koh, *International Law in Cyberspace*, 54 HARV. INT'L L.J. 1, 4 (2012), http://www.harvardilj.org/wp-content/uploads/2012/12/Koh-Speech-to-Publish1.pdf.

[224] Gal Beckerman, *Is cyberwar really war?*, BOS. GLOBE (Sept. 15, 2013), https://www.bostonglobe.com/ideas/2013/09/15/cyberwar-really-war/4lffEBgkf50GjqvmV1HlsO/story.html.

[225] *Id.*

to the level of an act of war in the view of cyberwar skeptics.[226] This view is not without merit; the law of war has evolved alongside the traditional, kinetic theater of war for centuries. Opponents to the notion of cyberwar advise caution before officially pursuing cyberspace as the newest theater of war in which to conduct offensive operations.[227] As recent cyberattacks have demonstrated, the integrity of existing defensive and security measures is wanting, and focusing on offensive strategies at the expense of shoring up domestic defenses continues to leave those systems vulnerable.[228]

Detractors to the idea of cyberwar present legitimate reasons to reexamine the current push towards offensive cyber-operations, particularly when defense and security are sacrificed or neglected in favor of offensive capabilities.[229] However, the detractor argument is rooted in traditional, kinetic views on warfare. Such maxims as "code can't explode" are technically correct, but to downplay the significance of hostile cyber-operations sets a dangerous precedent.[230] Technology continues to advance at a sometimes startling rate. More and more features of everyday life are becoming automated and networked together. This is true in both the civilian and military sectors, as discussed in the introduction.[231]

Unless the progress of technology makes an abrupt about-face and sprints off in a direction other than developing networked and remotely-accessible systems and tools, technological interconnectivity will continue to touch more and more facets of everyday life. The potential for harm exists now, and the avenues through which cyberspace threat actors can inflict that harm will only increase with time; that is to say, even if cyberwar cannot exist in the present, technological advancement will make it so in the not too distant future. The three-pronged artificial intelligence program proposed above could address both the current state of cyber-defense and the growing desire for offensive cyber-weapons.

Another theory advanced by Jeppe Teglskov Jacobsen argues that, while cyberattacks are valid and potentially dangerous, the predisposition of State actors is to resort to conventional, kinetic responses in the face of cyberattacks, thus decreasing the likelihood of an actual cyberwar scenario.[232] This detraction

[226] *Id.*

[227] *Id.*

[228] *Id.*

[229] *Id.*

[230] *Contra id.* (demonstrating how minimizing the threat of cyber warfare helps countries from retaliating against a cyber attack with nuclear weapons).

[231] *2017 DOD budget calls for 15 percent increase in military cyber spending,* MILITARY AEROSPACE ELECTRONICS, http://www.militaryaerospace.com/articles/2016/02/cyber-security-dod-budget.html (last visited Sept. 23, 2015).

[232] Jeppe Teglskov Jacobsen, *The Cyberwar Mirage and the Utility of Cyberattacks in War: How to Make Real Use of Clauswitz in the Age of Cyberspace* 12 (Danish Inst. for

of cyberwar holds more water than Rid's absolute rejection. It holds up particularly well in certain examples; for instance, a technologically integrated State would have fewer options to launch a cyberattack against a State with less technological sophistication, which might make conventional means more attractive. As between technologically sophisticated States, conducting hostilities through cyberspace does become a viable option. Cyberattacks are relatively inexpensive, faster, and do not require risk of harm to human soldiers, and the advantages of cyberwar will only grow in number as technology advances.[233]

ii. On Cyber-Weapons

In theory, an artificially intelligent weapons system could inflict massive casualties in both military and civilian populations, severely damage the environment, and irreversibly compromise the integrity of a targeted State. International Humanitarian Law suggests that the use of such a weapons system would be universally condemned by the international community.[234] Certainly the wanton and indiscriminate annihilation inflicted by this AI weapon would be in violation of even the most basic principles of the law of war; its use would never be explicitly authorized, and even its consideration would spark heated debate.

In the same vein that some voices argue against the existence and possibility of cyberwar, some voices argue that cyber-weapons categorically cannot be classified as weapons of mass destruction by definition. Jeffrey Carr, CEO of Taia Global Inc., is one such proponent of this theory; his particular interpretation rests on the ability of cyber-weapons to directly inflict physical damage or injury. Carr proposes that "the potential effect of… a cyber weapon – is in direct proportion to how much a given population relies upon the network that the weapon subverts or destroys."[235] This is an appropriate measure with which to gauge the efficacy of a cyber-weapon; to use an example, the temporary interference a DDoS attack would create in the computer systems that control a dam would not be nearly as profound as the resulting catastrophe inflicted by a computer virus specifically designed to open the dam above a populated area.

Int'l Studies, Working Paper, 2014:06, 2014),
https://www.ciaonet.org/attachments/25101/uploads.

[233] *Id.*

[234] Brown, *supra* note 186, at 127 (illustrating how there is a law of war during traditional war and the likelihood that if a cyber attack causes effects similar to traditional war it is condemnable by the international community).

[235] Jeffrey Carr, *The misunderstood acronym: Why cyber weapons aren't WMD*, 69 BULL. OF THE ATOMIC SCIENTISTS 32, 32 (2013).

In those scenarios, the DDoS attack would be an ineffective cyber-weapon because it had minimal impact on a network that is heavily relied upon, while the virus would be an extremely effective weapon. However, the interference caused by a DDoS attack on popular news website during a time of crisis would be extremely effective at disrupting vital communications. Another measure of a weapon's effectiveness, according to Carr, is the weapon's ability to directly kill or injure human beings.[236] Presently, no cyber-attack is attributed with the direct death of a human being.[237] This is one requirement found in any definition of weapons of mass destruction.[238] 50 U.S.C. referenced earlier in this section, and 18 U.S.C. referenced by Carr both define weapons of mass destruction: "any weapon that is designed or intended to cause death or serious bodily injury through the release, dissemination, or impact of toxic or poisonous chemicals."[239] Under these definitions, even if a cyber-weapon played a part in inflicting widespread loss of life, the cyber-weapon itself would still not attain classification as a weapon of mass destruction because it was not the direct cause of harm.

Relying on existing legal definitions and concepts is useful when dealing with issues that involve cyberspace, but these existing definitions and concepts will come up short of providing critical guidance when an issue wholly exists in cyberspace. Although changes to law happen slowly it does happen, and sometimes change happens quickly. The foundations of space law were established just one year after the launch of *Sputnik*.[240] Cyberspace is not dissimilar to outer space in the sense that it is a new frontier. Just as the international community has agreed to the prohibition of certain weapons systems in outer space, the same attitude can and should be applied to cyberspace.[241]

IV. Developing International Cyber Law

Treaties are not the only means of establishing international law. Customary international law, which "develops from the general and consistent practice of states if the practice is followed out of a sense of legal obligation," can be as binding on states as any treaty.[242] The foundations of a customary international regime on cyberspace have already begun to develop. Responses to large-scale

[236] *Id.* at 33.

[237] *Id.*

[238] *Id.*

[239] *Id.*

[240] Brown, *supra* note 186, at 128.

[241] Treaty on Principles Governing the Activities of Sates in the Exploration and Use of Outer Space, Including the Moon and Other Celestial Bodies, Jan. 27, 1967, 18 U.S.T., 2410, 610 U.N.T.S. 205.

[242] Brown, *supra* note 186, at 126.

cyberattacks, perpetrated by one state against another, help set this foundation. Noteworthy examples include the cyberattacks allegedly launched by Russia against Estonia in 2007 and Georgia in 2008.[243] Also in 2008, a computer worm dubbed "agent.btz" infiltrated the computer systems of the Department of Defense and was designed to transfer classified information to foreign computers.[244] In 2010, a cyberattack launched by Chinese hackers pilfered intellectual property from Google and compromised user G-Mail accounts.[245] The Stuxnet incident is another example.[246]

While the cyberattacks themselves are the problem, customary international law is borne through the responses to such actions. Oddly enough, there has been a "lack of protest from nations whose systems have been degraded in some way by obnoxious cyber activity."[247] A lack of protest to these actions creates a permissive regime under customary international law. The same can be said of espionage. While individual states have laws against spying, "there is no international law prohibiting espionage or insisting it violates sovereignty."[248] Even with strict laws prohibiting spying, many nations still conduct espionage despite the risks, which has in turn created a permissive regime within the international community. This is not to say that espionage is considered legal activity, but the permissive regime does not take action against such activities either.[249] The same appears to be true of cyber-operations.

Although little is prohibited in a permissive regime, the developing customary cyberspace law is controlled by certain existing principles. Aggressive cyberattacks designed to cause kinetic damage are "covered by the law regarding the use of force and armed attack."[250] While not necessarily rising to the level of use of force, cyberattacks designed to compromise the integrity of nuclear control systems would also likely be prohibited under customary international law given the large body of law aimed at nonproliferation and disarmament.[251] Due to the interconnectivity of the global economy, and the indiscriminate harm such a collapse would generate, cyberattacks launched against one nation's financial system would also likely be prohibited under customary international law.[252]

[243] *Id.* at 130-31.
[244] *Id.* at 131.
[245] *Id.*
[246] *Id.* at 131-32.
[247] *Id.* at 132.
[248] *Id.* at 133.
[249] *Id.*
[250] *Id.* at 137.
[251] *Id.* at 138.
[252] *Id.*

This permissive regime is further guided by existing principles of international law. In the context of wartime scenarios, "the logical approach is to take what guidance exists to govern more conventional warfare and determine whether it can be applied to cyberspace activities."[253] In 2012, Harold Koh delivered remarks to the USCYBERCOM on how international law already controlled certain cyberspace activities.[254] Koh explained that cyberattacks could constitute a use of force under Article 2(4) of the UN Charter, and that a state's right of self-defense under Article 51 of the Charter could be triggered by a cyberattack.[255] Koh also discussed issues with attribution, proportionality and necessity when responding to cyberattacks.[256] Attribution is critical to the developing customary international law on this matter because, if a cyberattack cannot be attributed to a nation and is instead attributed to a terrorist organization or gang not connected to or acting as a proxy of another nation, the perpetrated acts and state response have no precedential value for customary law.[257]

V. International Law and Artificial Intelligence

This article has discussed the theoretical capabilities of an artificially intelligent weapons system capable of learning and acting without direct input from a human operator. The damage such a weapons system could generate could potentially rival the damage inflicted by the detonation of a nuclear bomb, if not surpass it. Given the extensive body of law that aims for the non-proliferation and dismantling of these kinds of weapons, one might ask why an artificially intelligent weapons system should even be considered at all. It stands to reason that a CWMD would immediately run afoul of existing international legal principles, and would probably lead to the creation of new law specifically forbidding their use or development.

As we have seen, legal principles exist that can safely govern certain aspects of cyberspace today. But these same principles may not provide suitable guidance in the future as technology continues to rapidly evolve. Existing legal principles are grounded in the physical world and, when it comes to war, the law has thus far evolved around traditional, kinetic warfare. Cyberspace is different than the physical world. In some instances it makes sense and is useful to apply existing legal regimes to cyberspace. But cyberspace opens a universe of new possibilities, and those possibilities grow in number and complexity with each technological leap forward.

[253] *Id.* at 127.

[254] Koh, *supra* note 223.

[255] *Id.* at 4.

[256] *Id.*

[257] *Id.* at 6-7.

Cyberspace experts appear to agree that a cyberattack of sufficient sophistication and scale can lead to a loss of lives, and that such an attack is likely to happen in the not-too-distant future.[258] Because cyberspace is fundamentally different from the physical world we inhabit, the approach to cyber-weapons should take those differences into consideration. The law does change. That change can come from trying cases in court, legislative review, or international customs. The law governing war, and legal principles that are currently used to govern cyberspace are not immune to this change, nor should they be. Existing principles can serve as guidance in the creation of new legal principles to govern actions and weapons in cyberspace.

The existing international humanitarian law and treaties regulating weapons and conduct in war provide guidance on how a CWMD could be handled. A CWMD cannot be used aggressively without reason, and in accordance with self-defense it must respect the principles of proportionality and necessity. One major advantage of a CWMD is the ability to control and direct otherwise indiscriminate harm. Some of the scenarios provided earlier in this article describe potential widespread damage an AI weapons system could inflict on a civilian population. Unlike conventional WMDs, a CWMD would not only inflict indiscriminate harm in contravention of international humanitarian law. A CWMD could be designed to, or instructed to exclusively target military objectives, even going so far as to minimize enemy military casualties. An AI such as Infiltrator or Sentinel could even be tasked with improving civilian network infrastructure to the extent possible for the purpose of securing civilian access to vital information and communications. Perhaps most significantly, when military objectives are achieved by means of devastating weapons such as bombs and missiles, the damage and injury can linger for years. With a CWMD, once the military objectives are achieved, the weapon could simply be turned off.

DANGERS AND NECESSITY OF ARTIFICIALLY INTELLIGENT CYBER-WEAPONS

I. Speed of Technological Advancement

Moore's Law is a theory that describes the rate of technological advancement. It generally states that the rate of advancement will increase exponentially over time, while that exponential growth factor will depend upon the indus-

[258] *Cyber Attacks Likely to Increase*, PEW RES. CTR. (Oct. 29, 2014), http://www.pewinternet.org/2014/10/29/cyber-attacks-likely-to-increase/.

try.[259] The technology of various industries develops according to demand.[260] This can be seen in practice in communications; telephones have advanced from fixed, corded models to wireless devices capable of performing a multitude of tasks in just over a decade.[261] Some industries see faster advancement than others, and computer technology sees the most rapid advancement of any industry.[262]

Cyber-threat actors are taking advantage of this evolution. The increasing sophistication of technological tools enables more dangerous cyberattacks, and allows cyber-threat actors to inflict increasingly greater damage. Logic would suggest that technological evolution should create a level playing field, since both cyber-threat actors and their victims are afforded essentially the same technology. That does not appear to be the reality. While the physical hardware is widely available to both sides of the cyber-threat issue, cyber-threat actors continually develop new methods to bypass government and corporate defensive measures in what is referred to as the innovation gap.[263] Insidious innovations are not necessarily required to bypass existing cyber defenses. New technology can offer avenues for infiltration by cyber-threat actors by simply creating new opportunities for access.[264]

[259] David L. Chandler, *How to predict the progress of technology*, MIT NEWS (Mar. 6, 2013), http://news.mit.edu/2013/how-to-predict-the-progress-of-technology-0306.

[260] Karehka Ramey, *Technological Advancements and Their Effects on Humanity*, USE OF TECH. (Nov. 12, 2012), http://www.useoftechnology.com/technological-advancements-effects-humanity/.

[261] *Id.*

[262] Max Roser, *Technological Progress*, OURWORLDINDATA, https://ourworldindata.org/technological-progress/ (last visited Sept. 23, 2016).

> [T]he capabilities of many digital electronic devices are strongly linked to Moore's Law and are improving exponentially … [Between 1975 and 2009] … the power and speed of computers increased exponentially; the doubling time of computational capacity for personal computers was 1.5 years … Considering the time since the introduction of the IBM 350 in 1956, the growth rate of storage capacity … since 1980 … has been very steady and at an even higher rate than the increase of computer speed … The price for a given product quality has been exponentially decreasing over 110 years. And over the last six decades the energy demand for a fixed computational load halved every 18 months.

Id.

[263] Robert Winters, *Technological Advances in Computer Crime*, CRIM. JUST. FOCUS (July 20, 2013), https://cjfocus.com/2013/07/20/computercrimeadvances/.

[264] This kind of arms race is far from new. With the rise of cyberspace, the technologies and actions involved in cybercrime and cybersecurity are more complex than ever before, especially so in regards to artificial intelligence. *See* Sam McQuade, *Technology-enabled Crime, Policing and Security*, 32 THE J. of TECH. STUD. 32, 36 (2006) ("Crime and policing/security are technologically competitive enterprises that are inextricable, dynamic and co-evolving. Criminal innovations drive policing and security innovations, and by extension each perpetually co-evolves the other throughout time and society… [T]he gamut of simple-to-complex physical and social technologies used by these enterprises are dynamically intertwined, and they become more complex over time and distance subject to broad social, cul-

As social media continues to develop, and more personal information is used in online transactions, cyber-threat actors have developed avenues of infiltration that negate almost every possible counter-intrusion measure.[265] The use of botnets allows external parties to gain control of a computer to use for their own purposes, usually for financial gain.[266] Ransomware and crypto-locker programs allow an external party to lockdown or encrypt the systems of another network, and the encryption keys can be held for ransom or blackmail purposes.[267] While most of these actions revolve around financial profiteering, it is not hard to imagine similar tactics being used against military networks or other networks that control vital civilian infrastructure. A response is needed that can do more than keep up with the rising frequency and sophistication of cyberattacks. AI may present a solution that can swiftly and intelligently respond to sophisticated cyberattacks.[268]

a. AI Development

AI has also seen great development in a relatively short span of time. There have even been vast improvements made since this article was started just over one year ago. IBM's Watson has seen application in the medical field to assist in cancer treatment.[269] Recently, Watson has been tasked with matching cancer patients with clinical trials. "It's hoped that the technology will boost the number of patients given access to cutting edge treatments, as currently, many eligible patients miss out on trial opportunities."[270] Meanwhile, Berg, a U.S. biotech company, has also used artificial intelligence to help to assist in develop-

tural, political and economic conditions and constraints.").

[265] *See* Wajeb Gharibi & Maha Shaabi, *Cyber Threats in Social Networking Websites*, INT'L J. OF DISTRIBUTED & PARALLEL SYS., Jan. 2012, at 119, 125 ("Information security professionals, government officials and other intelligence agencies must develop new tools that prevent and adapt to the future potential risks and threats [for new technology in general and social websites in particular.]").

[266] Winters, *supra* note 263.

[267] *Id.*

[268] *See* Selma Dilek, et al., *Applications of Artificial Intelligence Techniques To Combating Cyber Crimes: A Review*, INT'L J. of ARTIFICIAL INTELLIGENCE & APPLICATION, Jan. 2015, at 21, 22, 25, 32-33 (explaining AI applications have seen some promising, if limited, use in cyber-defense, i.e., Artificial Neural Nets, Intelligent Agents, Artificial Immune Systems, Genetic Algorithms, and Fuzzy Sets already provide a number of advantages over traditional cyber defense strategies including adaptability, robustness, learning capability, and versatility – and future AI cyber-defense applications will need to continue to compensate for human limitations).

[269] Kitty Knowles, *IBM Watson is helping to cure cancer*, THE MEMO (June 2, 2016), http://www.thememo.com/2016/06/02/ibm-watson-cancer-cures-treatments-artificial-intelligence-ai/.

[270] *Id.*

ing treatments for cancer.[271] The response suggested by the AI was a new drug – BPM31510, "which tries to reverse the Warburg effect - the phenomenon in which cancerous cells change their energy supply.[272] After early clinical trials, data from 85 patients showed signs the approach could kill tumors.[273]

The legal profession has also recently seen a promising early application of artificial intelligence. A program developed by IBM called ROSS is capable of translating complex legal concepts into 'plain English,' which it accomplishes by reading through the *entire body of law* and returning a cited answer and topical readings from legislation, case law and secondary sources.[274] Ross also minimizes the time it takes to narrow down results from a thousand to only the most highly relevant answers.[275] In addition, it keeps up-to-date with developments in the legal system, specifically those that may affect attorneys' cases.[276]

A technology company called Cylance is already using AI to defend against cyberattacks.[277] Dubbed CylanceProtect, the AI is designed to:

> [extract] millions of unique characteristics from the code [of a file] and [analyze] them against trained statistical models to determine their intention. Rather than relying on hash comparison or post-run behavior heuristics to determine what to do, Cylance evaluates objects in less than 100 milliseconds, early in the run time process. That way, if the object is determined to be malicious, execution can be stopped.[278]

CylanceProtect can protect against a diverse array of malicious programs, including "spear phishing, zero-day malware, privilege escalations, scripts, and malicious programs…, and eliminates the need for antivirus and intrusion detection and prevention systems."[279]

In addition to the MarIO program described in the introduction, Google's DeepMind AI has recently completed the notoriously challenging Atari game, *Montezuma's Revenge*, by using a digital analogue of incentive.[280] AI has also

[271] James Gallagher, *Artificial Intelligence 'outsmarts cancer'*, BBC NEWS: HEALTH (June 8, 2016), http://www.bbc.com/news/health-36482333.

[272] *Id.*

[273] *Id.*

[274] Cecille De Jesus, *AI Lawyer "Ross" Has Been Hired By Its First Official Law Firm*, FUTURISM (May 11, 2016), http://futurism.com/artificially-intelligent-lawyer-ross-hired-first-official-law-firm/.

[275] *Id.*

[276] *Id.*

[277] Katherine Noyes, *This company uses AI to stop cyberattacks before they start*, INFOWORLD (June 9, 2016), http://www.infoworld.com/article/3081532/security/this-company-uses-ai-to-stop-cyberattacks-before-they-start.html.

[278] *Id.*

[279] *Id.*

[280] James Vincent, *Watch Google's AI master the infamously difficult Atari game Montezuma's Revenge*, THE VERGE (June 9, 2016), http://www.theverge.com/2016/6/9/11893002/google-ai-deepmind-atari-montezumas-revenge.

been recently tasked with screen-writing and playing board games like Chess and Go against globally ranked professionals.[281] Admittedly, artificial intelligence has fared better in some of these pursuits than others.[282] While these developments are very promising, the activities performed and challenges solved by the artificial intelligence typically involve clear, stable rules, so that the outcome can, with varying degrees of accuracy, be predicted, often relatively early in the game._Regardless, this research will enable future artificial intelligences to accomplish more complex tasks and make significant contributions to human civilization. However, this progress may come at a price.

i. Dangers and Solutions

Some popular names in the fields of Science, Technology, Engineering, and Mathematics (STEM) have made various ominous claims regarding AI research. These names include the world-renowned astrophysicist Stephen Hawking, and Elon Musk of SpaceX and Tesla fame.[283] Some of their comments warn of catastrophe, war, and human extinction at the cold, metal hands of our own creations.[284] Unless these men, and others like them who make similar claims, know something the rest of us don't about artificial intelligence, this rhetoric is purely speculative and no more based in reality than HAL or Ultron. While the research on artificial intelligence has shown a promising future, the examples of AI we see today still act in accordance with programming; no actions performed by modern artificial intelligence today can be attributed to a cognitive process resembling human thought.[285] That being said, there are dangers associated with the use or misuse of any technology. AI is no exception, and it can still pose dangers to the integrity of cyberspace.

[281] Steven Borowiec, *AlphaGo seals 4-1 victory over Go grandmaster Lee Sedol*, THE GUARDIAN (Mar. 15, 2016), https://www.theguardian.com/technology/2016/mar/15/googles-alphago-seals-4-1-victory-over-grandmaster-lee-sedol.

[282] HAL 90210, *This is what happens when an AI-written screenplay is made into a film*, THE GUARDIAN (June 10, 2016), https://www.theguardian.com/technology/2016/jun/10/artificial-intelligence-screenplay-sunspring-silicon-valley-thomas-middleditch-ai.

[283] Sam Shead, *Eric Schmidt dismissed the AI fears raised by Stephen Hawking and Elon Musk*, BUSINESS INSIDER: TECH (June 10, 2016), http://www.businessinsider.com/eric-schmidt-dismissed-the-ai-fears-raised-by-stephen-hawking-and-elon-musk-2016-6?r=UK&IR=T.

[284] *Id.*

[285] *See* John McDermott, *Futurist: Computers to Outsmart Humans by 2029*, INC. (June 27, 2012), http://www.inc.com/john-mcdermott/futurist-ray-kurzweil-computers-outsmart-humans-by-2029.html (mentioning Futurist Ray Kurzweil's prediction that computers will have the capacity to pass the Turing test, a test that measures a computer's reasoning relative to a human's, by 2029).

1. Runaway Train

There are already some examples of malicious computer code spreading throughout the internet, presumably beyond the control of its creators and outside the original scope of the programs. Stuxnet is perhaps the most notorious example. A brief examination of Stuxnet's origins and capabilities was mentioned in the introduction.[286] After its original objective had, presumably, been achieved, Stuxnet managed to spread beyond the Iranian nuclear facilities at Natanz and would go onto infect thousands of computers in over one hundred countries.[287] Stuxnet's unexpected proliferation is believed to have been the result of an error in the malware's operating code.[288] Stuxnet was designed to spread through USB devices and, while this method allowed it to successfully infect its targets, the coding error also allowed Stuxnet to jump back out of the isolated Iranian system via USB drive and onto a computer connected to the Internet where it spread unchecked.[289] The version of Stuxnet discovered in the wild only targeted computers running software developed by a German company, Siemens, and would become inert if no such software was discovered.[290] The Stuxnet saga demonstrates the need for strict oversight when deploying cyber-weapons capable of self-propagation, and highlights a very plausible danger associated with those weapon systems.

Sentinel- and Infiltrator-type AI constructs would require the ability to learn and self-propagate. For a Sentinel-type AI, the learning ability would enable it to better defend against a wide variety of cyberattack methods, and continue to defend against new threats as they emerge. Similarly, an Infiltrator-type AI would operate most efficiently by learning which intrusion methods were most successful against corresponding defense strategies. The learning functionality would not need to operate like that of a human; instead, the AI's learning ability would allow the constructs to best adapt to their environments in accordance with their programming. An Index that contained every approved and prohibited function as well as the types of systems with which the AI could and could not interface would allow an Infiltrator-type AI to operate with maximum efficiency.

Problems could arise if an Index item was listed improperly, or if a prohibited system possessed characteristics of a system approved for autonomous infil-

[286] Finkle, *supra* note 38.

[287] Vivian Yeo, *Stuxnet infections spread to 115 countries*, ZDNET (Aug. 10, 2010), http://www.zdnet.com/article/stuxnet-infections-spread-to-115-countries.

[288] David E. Sanger, *Obama Order Sped Up Wave of Cyberattacks Against Iran*, N.Y. TIMES, June 1, 2012, at A12.

[289] *Id.*

[290] Michael Joseph Gross, *A Declaration of Cyber-War*, VANITY FAIR (Mar. 2, 2011), http://www.vanityfair.com/news/2011/03/stuxnet-201104.

tration. In that instance, an Infiltrator AI could determine that it was authorized to autonomously perform an action on an approved system, when in reality the action or the target would have required handler oversight. Infiltrator would not be disobeying orders or acting of its own accord, it would be operating according to its programming, even if the outcomes are contrary to what its handlers intended. This could lead to a situation where Infiltrator would spread unchecked through networked systems potentially across the world, not because it was acting out of malice, but because it was operating in accordance with its programming. This could lead to bigger problems if Infiltrator began executing sabotage and disruption subroutines while embedded in unapproved networks.

For artificial intelligence with learning capability, the runaway train scenario does not depend on unexpectedly-developed maliciousness, but rather undesirable actions taken in accordance with preprogrammed parameters as a result of stimuli in the environment. This is already an issue of notable concern for Google, which has begun developing a “kill-switch” for their artificial intelligence programs that operate self-driving automobiles.[291] Google’s kill-switch would operate to prevent a self-driving vehicle from breaking traffic laws – specifically, a vehicle that breaks a traffic law to avoid a crash may learn to break other traffic laws to achieve its objectives (reaching a destination).[292] Applied to Infiltrator, a handler-operated interruption system could be used to prevent Infiltrator from executing prohibited functions, or interfacing with prohibited systems as it searched for access points, vulnerabilities, and learned new infiltration strategies. Constant surveillance and handler direction over an artificial intelligence construct operating in cyberspace would be vital to avoid a runaway train situation. Quickly identifying program bugs and other errors in coding would also be required to prevent a construct like Infiltrator from operating outside of its parameters.

2. Mirror, Mirror

Exact numbers vary, but corporate espionage and theft of trade secrets costs the global economy hundreds of billions of dollars every year.[293] For individual

[291] Tom Risen, *Google Seeks Kill Switch for Rogue Artificial Intelligence*, U.S. NEWS & WORLD REPORT (June 8, 2016), http://www.usnews.com/news/articles/2016-06-08/google-seeks-kill-switch-for-rogue-artificial-intelligence.

[292] *Id.*

[293] Ellen Nakashima & Andrea Peterson, *Report: Cybercrime and espionage costs $445 billion annually*, WASH. POST (June 9, 2014), https://www.washingtonpost.com/world/national-security/report-cybercrime-and-espionage-costs-445-billion-annually/2014/06/08/8995291c-ecce-11e3-9f5c-9075d5508f0a_story.html.

American enterprises the numbers are drastically smaller – in the tens of billions range – but the damage dealt by the theft of trade secrets can still be crippling.[294] In 2012, Chinese nationals were arrested in connection with an attempt to steal cellular-glass insulation technology from an American company.[295] The value of the company's secret was estimated at $272 million.[296] Technology pilfered by a tech-savvy competitor can be used to devastating effect against the original creator of that technology. It stands to reason that an AI weapon system would also be susceptible to reverse engineering, but the consequences of such an event could be felt on a global scale; instead of corporate trade secrets or sensitive government files, a reverse engineered AI could provide cyber-threat actors with a devastating new weapon.

Sentinel- and Infiltrator-type AI constructs would both require the ability to embed their code on systems and survive efforts to wipe their components, a function possessed by the suspiciously-advanced malware developed by the enigmatic Equation Group.[297] This function would enable the AI to automatically resurrect itself at a later date, or upon prompting from a handler. These AI would also require the ability to completely purge any trace of their presence from a compromised network to prevent unauthorized access to their code. No security measure is perfect, however, and it is possible that a situation could arise where some unauthorized person, organization, or even a sovereign entity could gain access to the AI code.

Sentinel- and Infiltrator-type AI constructs would need to operate in tandem, cooperating with one another to achieve their offensive and defensive objectives. This would require a sort of friend-or-foe identification protocol which would prevent Sentinel from blocking Infiltrator access, or prevent Infiltrator from attacking a Sentinel-protected computer. The friend-or-foe ID system would also prevent unnecessary complications in operations which require stealth, speed, and participation from both constructs.

Admiral Michael Rogers, commander of U.S. Cyber Command, believes that a devastating cyberattack will claim human lives by 2025.[298] While artificial intelligence could provide the protection needed to ward off attacks of this kind, it could also act as the aggressor. Reverse-engineered AI counterparts to

[294] Christopher Munsey, *Economic Espionage: Competing For Trade By Stealing Industrial Secrets*, FED. BUREAU OF INVESTIGATION (Nov. 6, 2013), https://leb.fbi.gov/2013/october-november/economic-espionage-competing-for-trade-by-stealing-industrial-secrets.

[295] *Id.*

[296] *Id.*

[297] Kaspersky Press Release, *supra* note 37.

[298] Patrick Tucker, *Major Cyber Attack Will Cause Significant Loss of Life By 2025, Experts Predict*, DEF. ONE (Oct. 29, 2014), http://www.defenseone.com/threats/2014/10/cyber-attack-will-cause-significant-loss-life-2025-experts-predict/97688.

Sentinel and Infiltrator – let's call them Cloak and Dagger – could inflict damage on multiple fronts simultaneously to devastating effect.

If Cloak and Dagger were reverse-engineered from Sentinel and Infiltrator components, these analogue constructs could make use of this friend-or-foe protocol and spread through protected networks undetected and without resistance. A Sentinel construct guarding the computer networks of a multinational, billion-dollar corporation might not recognize Dagger as an intruder and permit it access. Infiltrator might not attack a critical military system guarded by a Cloak construct using Sentinel's friend-or-foe protocol. Cloak and Dagger could spread through otherwise protected networks and cause damage while Sentinel and Infiltrator would essentially ignore them.

3. Terminator

Technology does not evolve in a vacuum; it is driven by demand. War demands the technology to best protect a sovereign's interests, whether that protection necessitates defense or offence. Recent years have seen a push towards automating military force, and AI and robotics research appear poised to intersect in a big way in the near future. Unmanned aerial drones already see significant use in the theater of kinetic warfare. Drones can accomplish a limited range of military objectives without endangering the lives of troops on the ground. Most drones still rely on human operators, who can direct the craft from half a world away. Until recently the idea of autonomous, robotic soldiers has been relegated to the realm of science fiction.[299] That might not remain the status quo for much longer.[300] Autonomous weapon systems already see application in the military; some examples are described in the introduction. These weapon systems still rely on varying degrees of human operation, but few (if any) of them utilize autonomous robotics.

As robotics technology advances, direct human oversight might not be necessary. Advanced locomotion – and accompanying advanced AI – would reduce or eliminate robots' reliance on humans for troubleshooting, maintenance, or physical assistance navigating terrain in the field. As an autonomous, robotic military future becomes possible, many voices in the international community have spoken out against hastily applying these technologies in the theater of

[299] Tia Ghose, *Ban Killer Robots Before They Take Over, Stephen Hawking & Elon Mush Say*, LIVE SCIENCE (July 27, 2015), http://www.livescience.com/51664-stephen-hawking-elon-musk-ai-weapons.html.

[300] Steven Metz, *Crossing the Rubicon: The Inevitable Emergence of Military Robots*, WORLD POL. REV. (June 10, 2016), http://www.worldpoliticsreview.com/articles/19026/crossing-the-rubicon-the-inevitable-emergence-of-military-robots.

war.[301] The Convention on Conventional Weapons held at the United Nations in April, 2016 concluded their deliberations by "call[ing] on all states to adopt a prohibition on the development, production, and use of fully autonomous weapons."[302]

Even without the looming concerns of autonomous military robots, military forces continue to become increasingly more reliant upon cyberspace to conduct their operations. If autonomous robots do see wide application in armed forces, this reliance will require the highest level of cybersecurity to ensure the automated component could not be compromised. Computers are not the only vulnerable targets anymore – power plants, self-driving cars, and even some pacemakers are susceptible to cyber-threat actors, as discussed earlier in the article. Any other autonomous technology, including a mechanized military force, would share at least some of the same vulnerabilities.

If technological automation becomes the norm for military applications, the already-high bar for cybersecurity standards will only increase. A cyber-threat actor could theoretically kill a person by tampering with their remotely-accessible pacemaker. Tampering with an entire army, or repurposing its soldiers for sinister purposes, raises the potential for harm even higher. The runaway train danger exists here too. Faulty programming or undesirable learning patterns could result in autonomous soldiers failing to distinguish between valid targets and non-combatants. Ensuring that artificially intelligent autonomous military applications are adequately protected from cyber-threat actors, requiring close supervision by human handlers, and developing a failsafe kill-switch option would be vital before deploying such technology in the field.

II. Necessity

Before the United States deployed its nuclear weaponry against Japan in World War II, the U.S. was drawn into the war after the attack on Pearl Harbor. Some cybersecurity experts believe that a cyberspace equivalent to Pearl Harbor is on the horizon.[303] A cyberattack on such a scale would likely result in significant human casualties. In the decades following the first deployment of nuclear weapons, nuclear power saw widespread proliferation across the globe. The possession of nuclear weaponry by multiple nations gave rise to the doc-

[301] Will Knight, *Military Robots: Armed, but How Dangerous?*, MIT TECH. REV. (Aug. 3, 2015), https://www.technologyreview.com/s/539876/military-robots-armed-but-how-dangerous.

[302] *Killer Robots and the Concept of Meaningful Human Control*, HUM. RTS. WATCH (Apr. 11, 2016), https://www.hrw.org/news/2016/04/11/killer-robots-and-concept-meaningful-human-control.

[303] Elisabeth Bumiller & Thom Shanker, *Panetta Warns of Dire Threat of U.S. Cyberattack*, N.Y. TIMES, Oct. 11, 2012, at A1.

trine of deterrence by way of the threat of mutually assured destruction.[304] A nuclear war between two nuclear power states would have devastating and far reaching consequences on such a high order of magnitude that no nation would willingly incite a nuclear conflict by firing first – with North Korea possibly being an exception.[305] A CWMD could give rise to a new doctrine of deterrence wherein no state would willingly incite a cyberwar against an AI-capable nation out of fear of the ensuing conflict wreaking havoc across cyberspace.

The cyber-Pearl Harbor concept has its share of detractors as well, and some voices go so far as to claim that the notion of a cyber-Pearl Harbor is a myth.[306] Detractors are quick to point out that using historical acts of physical war amounts to little more than hyperbole when using such events to describe actions taking place in cyberspace.[307] This argument uses the same logic as the "code can't kill" line of reasoning purported by detractors of cyberwar as a concept. That is not, however, the crux of the detractor argument against the occurrence of a cyber-Pearl Harbor. Instead, detractors point to the difficulty of attribution – being able to identify the identity of the threat actor – and the scale of cyberattacks to date.[308]

Furthermore, current cyberespionage strategies appear to be working just fine for cyber-threat actors who continue to commit cybercrimes against the United States. China pledged in 2015 to assist the United States in protecting cyberspace from crime.[309] However, the country is still suspected to be responsible for "years of large-scale cyber attacks that officials say have cost [the United States] billions of dollars in stolen intellectual property and compromised networks[.]"[310] Russia has also launched cyberattacks against U.S. net-

[304] *Cold War: A Brief History: Nuclear Deterrence*, ATOMIC ARCHIVE, http://www.atomicarchive.com/History/coldwar/page15.shtml (last visited Sept. 23, 2016).

[305] *How potent are North Korea's threats?* BBC (Sept. 15, 2015), http://www.bbc.com/news/world-asia-21710644.

[306] *See generally* Henry Farrell, *Cyber-Pearl Harbor is a myth*, WASH. POST (Nov. 11, 2013), https://www.washingtonpost.com/news/monkey-cage/wp/2013/11/11/cyber-pearl-harbor-is-a-myth.

[307] Henry Farrell, *The hack on the U.S. government was not a 'cyber Pearl Harbor' (but it was a very big deal)*, WASH. POST (June 15, 2015), https://www.washingtonpost.com/blogs/monkey-cage/wp/2015/06/15/the-hack-on-the-u-s-government-was-not-a-cyber-pearl-harbor-but-it-was-a-very-big-deal.

[308] Sharon Weinberger, *Cyber Pearl Harbor: Why hasn't a mega attack happened yet*?, BBC (Aug. 20, 2013), http://www.bbc.com/future/story/20130820-cyber-pearl-harbor-a-real-fear.

[309] Jane Perlez, *Xi Jinping Pledges to Work With U.S. to Stop Cybercrimes*, N.Y. TIMES, Sept., 22, 2015, at A11.

[310] Bill Gertz, *China Continuing Cyber Attacks on U.S. Networks*, WASH. FREE BEACON (Mar. 18, 2016), http://freebeacon.com/national-security/china-continuing-cyber-attacks-on-u-s-networks.

works, such as the email system used by the Joint Chiefs of Staff.[311] Launching a cyberattack on a scale large enough to cause physical damage and inflict human casualties may be possible by a powerful nation-state actor, but such an event would most likely draw the cyber-threat actor into a physical conflict or even physical warfare. It does not behoove any nation states currently acting in the capacity of a cyber-threat actor against the United States to escalate their operations to the point of provoking a kinetic military confrontation.

However, simply because an event has yet to take place, or because an action would lead to serious repercussions for the actor, does not make an unlikely event impossible, or even improbable. The corporate and government espionage objectives of China and Russia might not be shared by the more ideological governments of Iran or North Korea. The objectives and missions of sovereign nation-states change as leaders with new beliefs and attitudes rise to power. What stands to reason today may not be so tomorrow, and so it is imperative to prepare for that tomorrow.

Artificial intelligence provides a solution to the cyberwar problem. It can be used as an effective tool for both offense and defense, and it can intelligently respond to multiple threats simultaneously with unparalleled swiftness. Strict human handler oversight is necessary, from early development to field operations, to ensure that AI constructs operate as desired and achieve desired outcomes. Human oversight would need to evolve alongside the AI, determining how much – or how little – control would be required in any given situation. Simulating offensive and defensive strategies before engaging the AI in field operations would allow an AI weapon system to be implemented in such a way as to avoid undesirable outcomes and to minimize unnecessary damage if offensive strategies are required. Risk to human lives in armed kinetic conflict could be significantly reduced, and much of the intangible damage and interference caused by an AI weapon system in cyberspace could be stopped or even reversed. International laws that govern and restrict the use of kinetic weapons in traditional warfare can apply, to an extent, to the use of cyberweapons in cyberwarfare. By using existing legal concepts where applicable to conduct in cyberspace, artificial intelligence can provide both safety and security when accessing cyberspace and a highly effective weapon to use against those who would threaten it.

[311] *Id.*