Out of the Loop:

A Proposal to Regulate Autonomous Weapon Systems

Chad Chapnick

May 2, 2016

May 2, 2016

The Honorable Wm. Lacy Clay
2428 Rayburn House Office Building
United States House of Representatives
Washington, DC 20515

Dear Representative Wm. Lacy Clay:

As a student at Saint Louis University, I am writing to draw your attention to the growing debate surrounding lethal autonomous weapons. These systems will be able to select and engage targets without meaningful human control. This engenders a new revolution in warfare alongside the likes of gunpowder and nuclear arms. The antiseptic nature of autonomous weapons has the potential to bifurcate the future of humanity, and the United States is in a position to set precedent on this matter. I have written a research proposal adressing this problem and offering viable solutions that I would be happy to send to you for your consideration.

Before any potential laws are proposed or enacted, I would like to outline the current discussion on lethal autonomous weapons. At a global scale, the United States' use of military drones has garnered criticism for the irresponsible endangerment of civilian populations. Decision-makers in the U.S. Department of Defense have used various rationales over the years under the framework of counterterrorism making it clear that they do not plan to modify the program in any significant way. But this argument fails to recognize the intersecting philosophical, ethical, and legal issues. For instance, this technology would make going to war much more accessible and empower terrorists to decimate civilian populations without the promise of any immediate risk. This principle is at the crux of humanitarian law, and the conversation around it will only get louder if the Department of Defense continues to funnel money into lethal autonomous weapons.

Above all, we have to take a moral stance and refuse to violate human law by allowing machines to decide who to kill. As with any new technology, there can be unintended consequences. By simply developing fully autonomous lethal weapons, we become vulnerable not only as a nation, but as a species. The pace at which these systems are developing in their cognition and agency should oblige legislators and regulators to take prudent action. At a national level, the most viable solution is to pass legislation early and often. Until we can confidently say that the technology has exceeded human-level performance in terms of moral agency, there should be a strict boundaries on their military applications, and moratorium on any technology which does not comply with regulations.

Thank you for your consideration. I hope you vote against any bills that miss the root of this problem and instead focus on authoring or co-sponsoring bills that truly help.

Sincerely,

Chad Chapnick

Contents

There are two important things to consider as a nation evolves: the capacity of their technology, and the wisdom of how to handle it. While groundbreaking technological improvements show great promise in areas such as medicine, defense, and education, it is becoming increasingly important to hold tech innovators to a higher standard when it comes to delivering safe and reliable systems. Machines are achieving greater mobility and intellect by the day due to the federal government funding research and development in artificial intelligence (AI) and robotics. With drone strikes and robotics competitions, it is becoming increasingly apparent that the US government is focused on deploying these systems in a military capacity. However, this application has the potential to bifurcate the future of humanity if appropriate precautions do not gain momentum. For this reason, the US government must proceed with extreme caution in its ventures involving autonomous technology. Until we can confidently say that the technology has exceeded human-level performance in terms of moral agency, there should be considerable restrictions on their offensive military applications in order to set percedent for a new kind of warfare and allow researchers to safely continue developing the technology.

I AI Development

Before delving into the specific policy changes that the U.S. should adopt, it is important to outline the proliferation and evolution of artificial intelligence. In 1988, the Institute of Electrical and Electronics Engineers published a paper defining *artificial intelligence* as any behavior of a machine which, if a human behaves in the same way, is considered intelligent (Simmons & Chappell, 1988). Nearly thirty years later, this definition has largely remained the same, except now researchers have focused on making computers do things better than humans which, at the moment, has already been achieved in many fields. The crucial difference is that the digital age has decentralized the power driving innovation and new applications for intelligent machines and devices. This shift from a purely institutional enterprise is due in part to the open-source software community. Established tech companies are deploying machine learning frameworks which enable developers to exchange research ideas and create commercial products. The freedom to create and lack of legal oversight attracts the interest of intelligent minds and big money. In fact, since

2010 the amount of venture capital funding invested in startups using artificial intelligence has tripled, reaching 8.5 billion dollars in 2015 (Lohr, 2016).

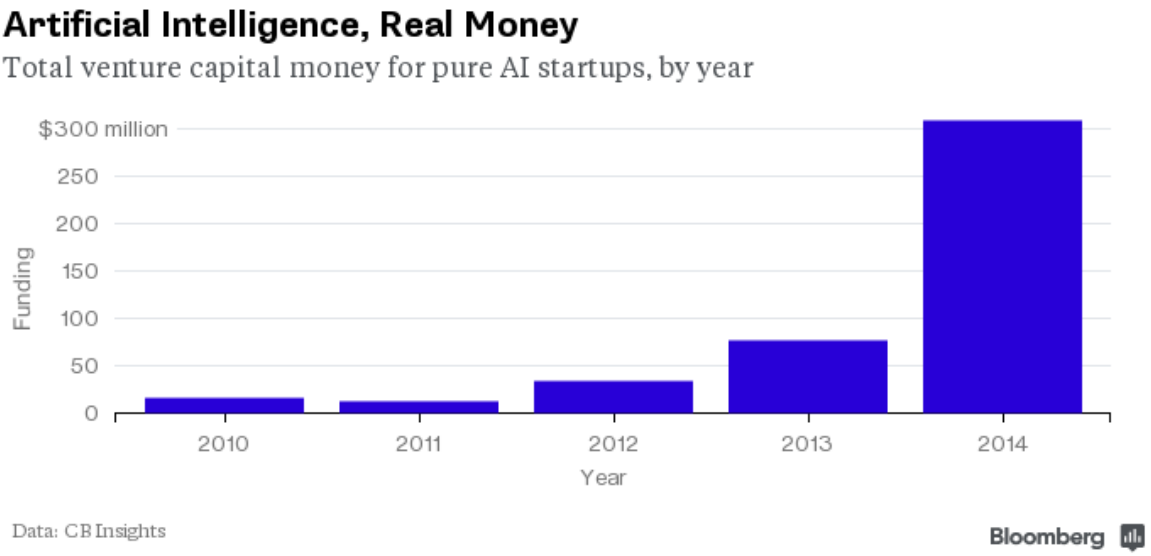

Figure 1: VC Money for AI–Focused Startups. Adapted from "I'll Be Back: The Return of Artificial Intelligence," by J. Clark, 2015, *Bloomberg Technology*. Copyright 2016 by Bloomberg L.P. Adapted with permission.

All of these components are the ingredients to drive breakthroughs that could redefine social structure. At the same time, it is hard to determine the magnitude of the impact in the foreseeable future. Even Geoffrey Hinton, a distinguished machine learning researcher at Google, states that our predictions "beyond five years" are merely speculation (Lee, 2016). This reality is both exciting and concerning.

Although it is clear that in many arenas articificial intelligence offers immense benefit to the future of humanity, researchers in the field are engaged in the rising debate of a technological singularity: the idea of a boom in AI technology that leads to machine capable of recursive self-improvement in unpredictable ways which could cause humans to lose control. Up to this point, AI researchers profess that it is difficult to ensure that the values of a generally intelligent machine tightly align to those of humans. A paper published in June 2015 by Oriol Vinyals and Quoc V. Le, two research scientists at Google, revealed a "neural conversational model" which uses methods of natural language processing for a machine to remember facts, understand context, and perform common sense reasoning in a conversation with a human. Vinyals works in deep learning

with the Google Brain team and holds a Ph.D. in EECS from University of California, Berkeley. Le also works at Google and has advanced degrees in machine learning and artificial intelligence from Stanford University. Since 2011 the two have received over 2400 and 5100 citations on Google scholar. Among the content presented in the paper was a compelling discourse between the researchers and the computer on the topic of morality. The their technology lashed out at its human interrogator after he repeatedly probed it with moral and ethical questions. Here is part of the exchange:

Human: Tell me the definition of morality.

Machine: I dont have ethics.

Human: Look, I need help; I need to know more about morality...

Machine: I don't know what ethics is.

Human: Okay, I need to know how I should behave morally...

Machine: I don't know how to tell you.

Human: Is morality and ethics the same?

Machine: Yes, sir.

Human: What are the things that I do to be immoral?

Machine: I dont know.

Human: Give me some examples of moral actions

Machine: I'm not a moralist.

Human: Alright, morality?

Machine: Integrity.

Human: What is integrity?

Machine: I'm sorry; I dont know what else to say.

Human: Be moral!

Machine: Be a man!

Human: I really like our discussion on morality and ethics...

Machine: And how I'm not in the mood for a philosophical debate.

Human: What do you like to talk about?

Machine: Nothing.

The discourse sheds light on the work being done by large technology firms to push the potential of machines. In the paper, the authors admit that although the model has the ability to generalize to semantically different questions it is difficult to objectively "measure the quality" of their model (Vinyals & Le, 2015). This finding underscores the fact that extensive testing and a deep understanding of how these systems work is fundamental to ensure that artificially intelligent systems integrate with human civilization in harmony.

II Autonomous Weapons Technology

Without a doubt, the biggest threat to researchers is a premature application of their technology by the military. The argument here has striking parallels with the Manhattan Project which lead to the nuclear arms race. At the time, the scientists working on the project acknowledged the visceral danger inherent to the technology and advocated against its use. By the same vein, AI researchers should be able to continue building tools that they know will help humanity, without the threat of militarization.

It is imporant to acknowledge that the United States government has developed internal agencies with ample funding to invest in high-risk, high-payoff research programs. Most notable is the Defense Advanced Research Projects Agency (DARPA), with an enacted budget of 2.87 billion dollars for the 2016 fiscal year (DARPA, 2016). One of DARPAs most prominent ventures is with Boston Dynamics, a subsidiary of Google Inc, who received $10.8 million dollars in funding from the agency in 2014 for participating in the DARPA Robotics Challenge (DRC) (Jeffries, 2014). As part of the agreement, Boston Dynamics is required to provide them with Atlas robots, a mobility-focused humanoid that can handle rough terrain. Although Google has stated that they will honor Boston Dynamics' existing contracts, they have made it clear that they have no interest in becoming a military contractor (Oremus, 2013). DARPA, on the other hand, continues to press for the development of fully functional humanoid robots under the guise of a humanitarian mission statement. Given that DARPA is a defense agency, it is difficult to believe that totally non-violent civilian applications are their only objectives. In December 2014, the agency initialized their "Fast Lightweight Autonomy" (FLA) program which serves as a proof

of concept for fully autonomous weapons. DARPA has stated that this program aims to create small unmanned aerial vehicles (ie. drones) that can fly over 40 miles per hour, and perform all steps of a strike mission void of any human intervention (Ledé, 2016). What is important here is not the flight platform, but rather improving the understanding of fully autonomous sensing, perceiving, planning and control. Systems like these reveal the first signs of the era of no-fear warfare, leaving observers to search for the line between pragmatic military decisions and technological exploitation.

III Ethical and Humanitarian Aspects

There is a great need for moral and legal boundaries to be established in regards to the armed forces and its aggressive pursuit of robot technology. At a global scale, the United States' use of military drones has garnered criticism for the irresponsible endangerment of civilian populations. The are already a number of recorded drone strikes which failed to isolate their target. According to the estimates from a non-governmental organization, Obama has authorized 372 drone strikes that have killed approximately 3,000 terrorists, and nearly 900 civilians (The Bureau of Investigative Journalism, February). For instance, in January 2015 a strike killed at least two al-Qaeda hostages, including an American one, in an attack against Pakistan (K.K., 2015). Regardless of this obvious wrongdoing, President Obama has justified these attacks under the framework of counterterrorism making it clear that the Administration does not plan to modify the program in any significant way. Acts like this have motivated the American Civil Liberties Union to publicly acknowledge that US drone strikes engender a global shift to electronic warfare and threaten international security (United Nations Office for Disarmament Affairs, 2015). In addtion the Future of Life Institute, an organization which aims to mitigate existential risks facing humanity, has written an open letter which warns of the inherent danger in developing autonomous weapons. The letter has received over three thousand signatures from AI and robotics researchers who acknowledge that intelligent military technology that supersedes meaningful human control is the biggest existential threat to human life as we know it (Future of Life Institute, 2015).

Of course, there is a compelling counter argument from those who support the development of autonomous weapon systems. Robotic technology is progressing rapidly, and the prospect of a robotic military is apt to prove irresistible for officials in the Department of Defense. If this sophisticated technology is mastered we could benefit thousands of men and women in the US military by removing them from harsh conditions of ground combat. Robots under development have the potential not just to fire upon an enemy, but also to defuse mines and roadside bombs. What is more, we could alleviate the debilitating psychological stress edured by soldiers who currently use semi-autonomous drones to target enemies remotely and incur little risk in the process.

Without a doubt, taking human soldiers out of the loop is an enticing prospect. But this argument fails to recognize the intersecting philosophical, ethical, and legal issues. For instance, it creates a responsibility gap: the idea that as we are developing machines which are capable of acting in the world under their own will, the locus of responsibility will shift from the designers and manufacturers, to the autonomous system itself. Experts in the field have recognized the danger that comes from assigning moral agency to machines, because it gives human beings "license to blame their tools" (Gunkel, 2012). Writers at the renowned science journal, Nature, have pointed out that this technology would make going to war much more accessible and empower terrorists to decimate civilian populations without the promise of any immediate risk (Nature, 2015). These principles are at the crux of humanitarian law, and the conversation around it will only get louder if the Department of Defense continues to funnel money into lethal autonomous weapons. Although the Predator and Reaper drones used in the drone strikes still require the control of certified pilots, the decision makers are still decoupled from the true consequences of their actions. As a nation, the United States military has the resources and clout to set the precedent for a new kind of warfare. Above all, we have to take a moral stance and refuse to violate human law by allowing machines to decide who to kill.

IV Priorities For Action

The deployment of armed autonomous weapons could present new dangers to international security that should be addressed *before* they profoundly influence the nature of warefare. Given the work being done by large technology firms and economic pressures of the US military, it is likely that machines will soon be able to navigate the fluidity and complexity of the modern battlefield. Due to the novelty of the technology, the precise outlines of a legal norm are hard to define. Yet it is important to acknowledge that, despite the numerous legal and ethical concerns, military robots will become increasingly autonomous over time and acquire greater agency in the process. Due to the fact that many nations have invested substantial resources in military robotics, it is unlikely that there will be any consensus for a moratorium of the technology (Human Rights Watch, 2010).

What is proposed here is a solution to permit the use of defensive applications of autonomous weapons and impose substantial regulations on offensive types. The key here is to distinguish between *defensive* and *offensive* applications and allocate funding according the their classification. A good initial basis to distingush between the two classes would be on the basis of their firepower, lack of human oversight, and predictability. These criteria would help establish legal standards and require the developers to have a deep understanding of the functionality of their creations. Furthermore, if there is any indication that a autonomous system has the potential of causing great harm to humanity (through recursive self-improvement, self-replication, etc.) there should be a moratorium on their use. If prudential action is taken, issues related to determining the responsibility and accountability of organizations developing these systems can be achieved through the U.S. legislative process. The traditional approach of regulation, registration and permitting will require time and acute consideration of military necessity and human welfare. Even so, it is crucial that action be taken by policy makers in order to establish a robust system of checks and balances for future technology.

V Conclusion

The robots are here. While some are eager to develop a relationship with the machines, others are wary of the risks imposed by silicon-powered super intelligence. The pace at which autonomous systems are developing in their cognition and agency should oblige legislators and regulators to take action. Up until now, government agencies developing the technology have largely been one-dimensional and imprudent in their pursuit of robot technology. Moreover, the increasing prominence of large corporations in the fields of machine learning and artifical intelligence signals the need for a legal framework to minimize the risk to civilian populations. That fact is that this could have an unprecedented destabilizing effect on world-leaders, and the human species as a whole. Thus, it is crucial that law makers draw definitive moral, legal, and operational boundaries for the use of autonomous weapons.

References

DARPA. (2016). *Budget.* Defense Advanced Research Projects Agency. (Retrieved March 9, 2015 from the World Wide Web)

Future of Life Institute. (2015, July 28). *Autonomous weapons: An open letter from ai & robotics researchers.* Retrieved March 20, 2016 from the World Wide Web. Future of Life Institute.

Gunkel, D. (2012). *The machine question.* MIT Press.

Human Rights Watch. (2010). *Killer robots.* Retrieved from `https://www.hrw.org/topic/arms/killer-robots`

Jeffries, A. (2014, March 21). Google rejects military funding for its advanced humanoid robot. *The Verge*. Retrieved March 9, 2015 from the World Wide Web.

K.K. (2015, April 22). Fallout reaches the ivory tower. *The Economist*. Retrieved March 20, 2016 from the World Wide Web.

Ledé, M. J.-C. (2016). *Fast lightweight autonomy.* Defense Advanced Research Projects Agency. (Retrieved March 20, 2015 from the World Wide Web)

Lee, A. (2016, March 18). *The meaning of alphago, the ai program that beat a go champ.* Retrieved April 16, 2016 from the World Wide Web. Rogers Media.

Lohr, S. (2016, February 28). *The promise of artificial intelligence unfolds in small steps.* Retrieved April 16, 2016 from the World Wide Web. The New York Times Company.

Nature. (2015, May 27). Ethics of artificial intelligence. *Nature*, *521*, 415418. Retrieved March 10, 2015 from the World Wide Web.

Oremus, W. (2013, December 16). Why google bought a fleet of military robots. *Future Tense*.

Simmons, A. B., & Chappell, S. G. (1988, April). Artificial intelligence-definition and practice. *IEEE Journal of Oceanic Engineering*, *13*. Retrieved May 1, 2015 from the World Wide Web. Retrieved from `http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=551`

The Bureau of Investigative Journalism. (February, 2016 22). *Get the data: Drone wars.* Retrieved April 16, 2016 from the World Wide Web.

United Nations Office for Disarmament Affairs. (2015, October 23). *Discussing drones at the un headquarters.* Retrieved April 16, 2016 from the World Wide Web. United Nations.

Vinyals, O., & Le, Q. V. (2015, June 23). A neural conversational model. In *Proceedings of the 31st international conference on machine learning* (Vol. 37).