Artificial Intelligence Security Policy

Adeeb Ahmed

4th Year Undergraduate

PS 306-01

Dr. Shaprio

Wednesday, November 30th 2016

*[0] Weak AI – Artificial intelligence that does well at the limited range of tasks for which it was designed for. (spam filter, search engine, ad results)
*[1] Day 0 – The day when strong artificial intelligence is revealed to the public
*[2] Strong AI – Artificial intelligence with the ability to apply past experience to new problems areas and challenges. (autonomous driving, playing go)

Artificial Intelligence Security Policy

Cyber security has gained a lot of attention due to the rise in government/corporate hackings. Polices such as Cyber Intelligence sharing and Protection Act (CISPA) or Stop Online Piracy Act (SOPA) aim to aid authorities in capturing hackers violating privacy/security laws, but arguably give the government too much power.

The failure of these policies did nothing for matters of cyber defense, security, or piracy. Over the last few years the attention towards these policies, and the topic of cyber defense, has simmered down. This is an issue because as technology progresses AI (artificial intelligence) will introduce many ethical issues that would require policies for regulation. Taking all factors in to consideration, the government is not best suited to regulate technology alone.

The growth of AI has introduced problems, and will continue to introduce problems to matters of security. For instance, CAPTCHAS are a form of security used throughout the internet. Currently, it's possible to train a weak AI*[0] to solve these captchas with a high degree of accuracy. Though it still takes a while, it's only matter of time before the technology is perfected and creates vulnerabilities across the web. There are many more instances of security issues similar to this one.

Ideally, the government and tech giants should work together in forming cyber defense policies around AI security. However, legislation is too slow when it comes to matters of technology and it might just be wiser to leave certain decisions to the tech giants.

Currently the status quo on the situation is that major corporations have banded together to self-regulate (open source) AI research. “Researchers from Facebook, Alphabet, Amazon, Microsoft and IBM are looking at the practical consequences of AI, such as how it will impact transportation, jobs and welfare. The group doesn't have a name or an official credo, but its general goal is to ensure AI research focuses on benefiting people, not harming them (Conditt, 2016).” The pace of advancement in matters of cyber security is not enough to keep up with advancements in AI. AI and machine learning resources are being researched and released regularly. In addition, the major companies and the government research independently of each other.

An alternative solution would be requiring the government to work with companies, the open source community and vice versa. The collaboration between government and the tech space will help develop robust policies to ensure the safety of AI R&D. With global collaboration AI research can be truly be standardized and secure. Generally speaking, companies have their own best interest in mind and an entity to represent people should be involved.

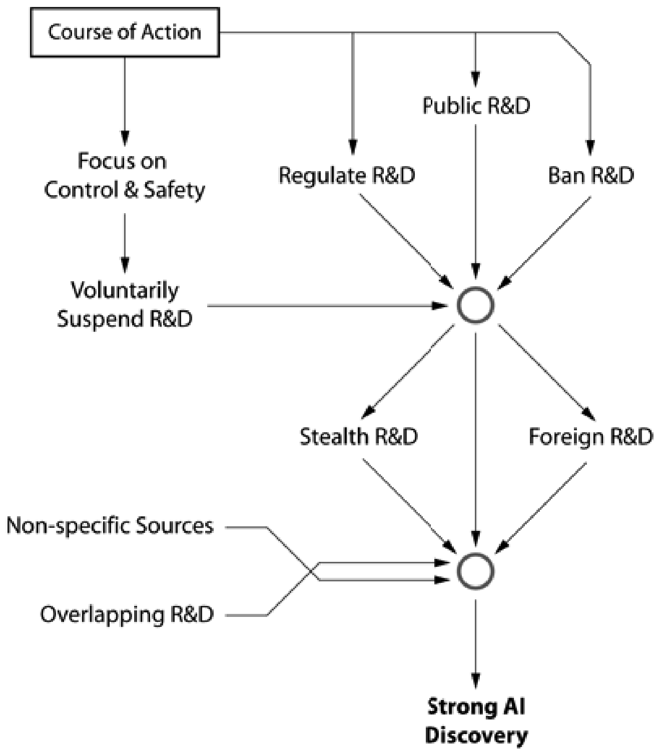
The extremist solution would be to ban AI research altogether until safety and security is assured from products of AI. The idea is the government should ban R&D research until they can come up with a way to safely regulate it. Stakeholders in favor of this solution fear that a major discovery in the field is only a matter of time, and that strict security measures need to be taken. Their greatest fear is the “AI-pocolyspe”: a doomsday scenario where AI finally overtakes humanity, enslaving/destroying the human race, and ultimately becoming the next dominant life form.

By far the most important product of Artificial Intelligence is not the problems that it can solve, rather it’s the security gaps created by AI systems. As of today the general public is somewhat uneducated when it comes to matters of computer science. People generally refer to science fiction movies they've seen of AI threating to eradicate the human race. The only people who really understand what AI truly is, are computer scientists. This needs to change. Congressional representative Susan Brooks presented the Computer Science Education Act (CSEA) of 2013 to congress. The act was developed on the premise that "elementary and secondary computer science education gives students a deeper knowledge of the fundamentals of computing, yielding critical thinking skills that will serve students throughout their lives in numerous fields (Brooks 2013)." With fundamental knowledge of computer science, the general public can safely adapt to the use of AI systems in future generations.

Though the bill was not passed, it's a step in the right direction for AI security. The bill was again revised in 2015 by congressional representative Robert Casey Jr. It's likely the bill will stick around and eventually be admitted as law. This is where I believe collaboration with tech giants would be beneficial. Prior and post the CSEA, companies were funding massive campaigns that demanded computer science be included in common core education. The tech giants have the funds to enforce the bill once signed into law, and by doing so will decrease the financial burden of the education budget. The rate at which the US (and society) becomes computer science literate will translate into the level of security we establish, come Day 0*[1]. The more familiar a society is with computer science and its concepts, the more secure they'll be.

The extremists' solution of banning AI R&D might keep us safer slightly longer, but only by delaying the inevitable. According to Dustin Juliano, author of aisecurity.org, "it is crucial to understand that strong AI*[2] is not out of reach because we lack a certain kind of technology or instrumentation (Juliano 7.1)." For example, in particle physics, complex and expensive equipment is required to detect and measure certain particle interactions. However, in computer science, strong AI is algorithmic. It is a puzzle in the form of a computer program; all of the building blocks already exist, we need only arrange them correctly. Essentially, it doesn’t matter what we do, at this point the arrival of strong AI is just a matter of time.

Figure 7.1: The illusion of choice to restrict or pursue strong artificial intelligence research. All paths lead to SAI discovery.



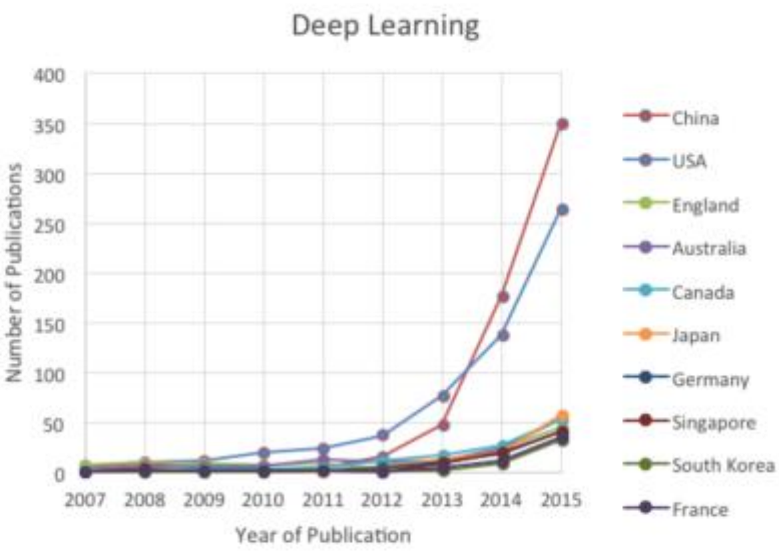
From figure 7.1 we can see that too much focus on control and safety might lead to a voluntary suspension of AI R&D which would slow down progress in the field. Presenting the opportunity for a major discovery to foreign counties such as China. The embargo of AI R&D would not stop people from researching AI.

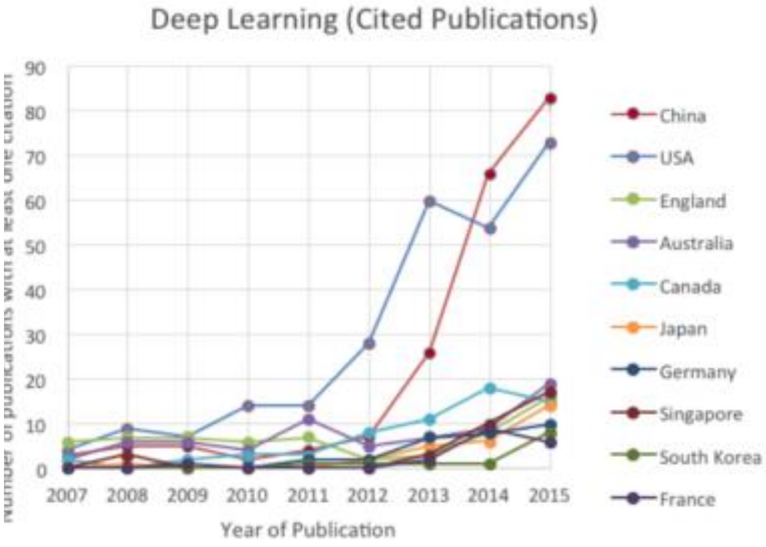
The ban of R&D would lead to stealth and foreign reliance on any strong AI breakthroughs. It would also create the issues of overlapping research since research would have to be done off record. Thus delaying the inevitable strong AI discovery for no greater purpose.

Economic Analysis

Goals	Impact Category	Policy Alternatives		
		Independent entities	**Global Collaboration**	**Government Bans AI research**
Safety & secure implantation of AI to society	Educate public about AI	**Slow progress**	**Fast progress**	**Negligible progress**
Maintain leadership in AI research	# of AI R&D publications/yr	**~260**	**~260**	**0**
	# of AI R&D publications cited/yr	**~72**	**>72**	**0**
Social/Economic well being	Job growth	2,700	**>2700**	**0**
	Median pay	**$110,620**	**>$110,620**	**0**
Fiscal expense	Cost to educate public about AI	N/A	**Decrease**	**N/A**

From congress.gov: As of 11/29/2016 a CBO Cost Estimate for this measure has not been received.





Data for these figures were obtained from a search of the Web of Science Core Collection for "deep learning" or "deep neural net*", for any publication, retrieved 30 August 2016.

Political Analysis

Actors	Motivations	Beliefs	Resources
Interest Groups			
Tech Giant Panel	The safe advancement of AI technology	Safely Integrate strong AI to technology products	Corporate budget Best tech talent Corporate influence
National Science and Technology Council (NSTC)	The safe and secure implementation of new science and technologies	Safely Integrate strong AI to technology products through collaboration	NSTC budget NSTC R&D teams National influence
Open source community	Transparency, safety, security, community	Safely Integrate strong AI to technology products through updates done by a global community	Global awareness and security Decentralized
Uninformed population fearing AI	Fear, lack of knowledge, loss of job	AI cannot be safely integrated	Public opinion Votes
Malicious Hackers	Greed, poor upbringing, politics, religion, etc	Using their skills to negatively impact other actors	Open source projects and public knowledge
Unelected Officials			
John P. Holdren – NSTC chair/Assistant to president	Economic growth and improvement	AI Technology has opened up new markets and new opportunities	Professional advisor to the president, can influence AI policies
Eric Horvitz – Microsoft AI researcher	Educated AI related decisions must be made on gov level	Increase AI education within all levels of government	Created the “One hundred year study”
Elected Officials			
Susan W. Brooks	Lack of CS education in common core	Computer science education is essential for the future	Congressional vote
Robert P. Casey Jr.	Lack of CS education in common core	Computer science education is essential for the future	Congressional vote

In terms of AI research, the US was the leader up until 2013, when they were surpassed by China in number of publications. This is the same year the CSEA bill was created. Although correlation does not equal causation, it's interesting to see what might be a motivating factor for including computer science education at an elementary level.

Via the independent entity policy, Median pay is expected to increase due to the increase in demand for AI researchers and the current number of AI researchers staying rather constant. This is a product of congress that still need to vet and pass the CESA. By the time the bill is passed and signed into law, the median pay for an AI researcher is anticipated to increase exponentially.

Since the bill is not yet passed, there is no budget/estimate regarding the costs of the program. Though the costs would decrease if the congress collaborates with the tech panel to assist with funding. The bill still needs to be approved before any financial actions are taken. Socioeconomically there is positive job growth in the field, expecting around "2700 new jobs by 2024 (BL&S 2015)." Via the collaboration policy, many more jobs can be created from developing systems of checks and balances for strong AI.

Political actors such as Brooks & Casey above are making the appropriate steps in improving AI security from a ground level. The impact computer science concepts can have on youth are exponential. However, to get the bill passed true collaboration will be required from all parties. The biggest threat to AI security is the lack of responsiveness to progress in the field. With the current legislation system its feared that congress is too slow to respond to breakthroughs in AI.

Unelected actor John Holdren is the chair of the National Science and Technology Council (NSTC) and advisor to the president. Directly under John is Afua Bruce, the Executive

Director of the NSTC who oversees the subcommittee of machine learning and artificial intelligence. This subcommittee is led by Ed Feltan, the Deputy CTO of the Unites States. In a letter to the public domain John states that "NSTC's Subcommittee on Machine Learning and Artificial Intelligence, which was chartered in May 2016, is to provide technical and policy advice on topics related to AI, and to monitor the development of AI technologies across industry, the research community, and the Federal Government (NSTC 5)."

I envision a collaborative system composed of the tech giants panel, US NSTC, and open source entities such as OpenAI and/or the 100-year study panel. OpenAI is a non-profit artificial intelligence research company whose mission is to build safe AI, and ensure AI's benefits are as widely and evenly distributed as possible. Organizations such as OpenAI can serve voice of the people by collecting public donations for research or policy funding. When a AI related policy passes in the house, the tech giant panel/open source entities should be consulted for full or partial funding, if there is a lack of funds (which there usually is). This will accelerate the implantation of such policies which in turn will increase net security.

In addition to this, tech giant panel members can meet/contact Mr. Holdren, to discuss strong AI discovery management. As can open source entities. The purpose of these meeting would be to inform Mr. Holdren of any immediate catastrophe that should be relayed to the president. This allows the president to take quick executive action if required. Thus improving net security.

Inversely, Mr. Holdren, and the AI team under his command, would be able to meet with the tech panel/open source entities to securely coordinate AI research. This collaboration will safely accelerate the discoveries required for strong AI. Taking that thought even further, why stop at the tech panel/open source entities? Collaborating with other governments such as China would be ideal for ensuring AI security because the AI needs of the Chinese will differ from needs in the US. Collaboration at the global level will validate strong AI's function the way they are intended and nothing more.

To ensure AI's safe integration to society, Microsoft AI researcher Eric Horvitz led in the development of the "100-year study" which is "a long-term investigation of the field of Artificial Intelligence (AI) and its influences on people, their communities, and society" (100yr Study report 1). It considers the science, engineering, and deployment of AI-enabled computing systems. Within the study group there is a Standing Committee that selects and oversees the One Hundred Year Study Panel. The Study Panel publishes their findings every five years to publicly assess the current state of AI. The Study Panel reviews AI's progress from the prior report, describes the technical and societal challenges found, and design secure AI systems compatible with human cognition.

There are a lot of things we can do to prepare for the arrival of strong AI. Even by taking all the precautions possible there is still a chance for an "AI-pocolypse" scenario to occur. While it's counterproductive to ban AI research, it could be argued that it's even riskier to proceed forward with AI in general. With the rise of hacktivist groups, cyber terrorism, corporate hackings, and now artificial intelligence, it's safe to say the cyber security space is rapidly growing.

Overall, the security of AI systems needs to be managed by a global community. There will need to be many complex series of checks and balances to prevent malicious actors attempting to reduce the world to an "AI-pocolyptic" zone. The best prescription to remedy matters of AI security would be Policy II (Global collaboration). A world where humans live peacefully with strong AI is attainable through transparency and collaboration on a global scale.

References

ARTIFICIAL INTELLIGENCE AND LIFE IN 2030 (Rep. No. 1). (n.d.). Retrieved https://ai100.stanford.edu/sites/default/files/ai_100_report_0831fnl.pdf

Conditt, J. (2016, September 01). Tech's biggest names are working to regulate AI research. Retrieved November 30, 2016, from https://www.engadget.com/2016/09/01/facebook-google-microsoft-amazon-ibm-ai-panel/

Holdren, J. P., & Bruce, A. (2016, October 12). *PREPARING FOR THE FUTURE OF ARTIFICIAL INTELLIGENCE* (United States, National Science and Technology Council). Retrieved November 23, 2016, from https://goo.gl/7QvNAu

H.R. 2356, 113th Cong., 1 (2013) (enacted). Retrieved November 30, 2016, from https://www.congress.gov/bill/113th-congress/house-bill/2536/text

Juliano, D. (n.d.). AI Security. Retrieved November 30, 2016, from http://aisecurity.org/ref/

Summary. (2015, December 17). Retrieved November 30, 2016, from http://www.bls.gov/ooh/computer-and-information-technology/computer-and-information-research-scientists.htm