A Business Analysis On AI Security [*]

Yuting "Summer" Yue[†]

January 11, 2018

1 Overview of AI Security

Neural networks have become pervasive in various domains. They are computing systems inspired by human brains that can progressive improve their performances on certain tasks by observing a large number of example inputs and outputs. Neural networks bring convenience to people's lives in a wide range of fields, from browsing social media websites to launching a rocket to Mars. Applications of neural networks include Facebook's image tagging and Tesla's autonomous cars.

However, in industries such as finance [1], healthcare [2] and autonomous vehicles [3], insecure algorithms can pose extreme security risks. Prone to adversarial perturbations, neural networks can often be fooled. For example, a facial recognition system used for the authentication can be fooled by an attacker to evade recognition or impersonate an-other individual[4]. Autonomous cars can be fooled by slightly modified stop signs which may lead to traffic accidents. If an algorithm that issues loans to people with eligible credit card history misclassifies non-eligible applicants, a large number of loans could be mistakenly issued to attackers. We refer to this phenomenon as "adversarial attacks".

2 History

Although developers have been concerned with cybersecurity since the 20th century, AI security that can prevent adversarial attacks into machine learning systems is an emerg-ing field. The fact that neural networks are prone to adversarial attacks was discovered for the first time in a research paper by Google in 2013[5]. The paper referred to this phenomenon as an "intriguing property of neural networks".

In recent years, other researchers also have started looking into this field in AI secu-rity. Numerous attack and defense methods have been proposed. In 2016, Papernot's group performed a real-world and properly-blinded evaluation on their attack methods. They were able to successfully attack systems hosted by MetaMind, Google and Amazon and yield an error rate above 80%[6].

[*] The technical report and experiments can be found at its GitHub page

[†] This research paper is supervised by Sangdon Park (sangdonp@cis.upenn.edu), Prof.Insup Lee (lee@cis.upenn.edu) and Prof.James Weimer(weimerj@seas.upenn.edu).

3 Players and Drivers of The Field

3.1 Large Technology Companies

As more and more neural network systems are deployed into production, large technology companies such as Google, Microsoft and Amazon will need to put more emphasis on the security of these systems. A small breach in a system may lead to disastrous profit loss for these companies. If a flaw is discovered in a model of Tesla's self driving cars such that the system thinks that stop signs are yield signs, many customers will become very reluctant to purchase the product, which would lead to billions of profit losses for the company.

In addition, large tech companies are major targets for hackers. People tend to perceive large companies' products as 'more reliable'. Therefore, if a hacker is able to hack into a company's system, he or she can gain more credentials and media coverage. For example, the IPhone X has a new face authentication feature. Users are now able to unlock the IPhone by simply looking at the camera. The neural network system Apple developed can compare the current photo to their recorded photos of the owner, and verify if they are the same person. Not long after the product was released, a researcher in Vietnam has demonstrated how he fooled the face recognition ID software on using a mask made with a 3D printer, silicone, and paper tape[7].

Due to the reasons stated above, large tech companies have the incentives to make their AI systems more secure. They can do this by internal R&D as well as acquiring smaller research companies. To prepare for the era of autonomous ride sharing systems, Uber Technologies Inc. already poached 40 researchers and scientists from Carnegie Mellon University in 2015, a university known for its Computer Science department[8]. Some of the researchers specialize in AI security research. This trend will endure with other technology companies as well.

3.2 Research Institutions

Research institutions in the US and Canada have been the major current players in the field. As machine learning becomes a prominent research topic, relevant research in AI security has been deemed important. The Machine Intelligence Research Institute, a research nonprofit in California wrote in their blog that 'a risk-conscious security mindset' is one of the primary goals for the institution and the field[9]. Additionally, the AI lab in University of California, Berkeley also has AI security as one of its major focuses.

AI security is deemed as a noble cause for many researchers in the field. As AI becomes increasingly powerful, a security flaw would become a more serious problem. Therefore, academic research in this field are motivated by the grander goal of benefiting the humanity by preventing technological disasters.

3.3 Policy Makers

Policy makers have not yet strictly moderated AI companies who are deploying products that implement machine learning algorithms, but they will as more products get deployed.

For example, Lyft partnered with self-driving technology company Aptiv to offer rides in its robot taxis during CES in Las Vegas in early 2018. This is a start for autonomous cars to become more significant in our everyday lives. As self driving cars increasingly hit the roads, the policy makers will need to set a security guideline on these vehicles. AI has also been used in weapons. Lethal autonomous weapons systems (LAWS) are able to operate and select who to kill autonomously without human judgments. Policy makers have the responsibility to determine if such weapons should be legal.

4 Factors That Restrain the Field

4.1 The Slow-down of AI Development

The field of AI security only exists because neural networks are used in software products. The development of AI security is highly correlated with the development of artificial intelligence itself. It is possible for a newer and better technology, that can achieve everything a neural network can achieve in the future, to be invented. In that case, the effort to make neural networks safer would be no longer needed.

4.2 The Performance VS Security Tradeoff

In addition, there is often a tradeoff between security and performance. For example, an image classifier can be either made more accurate or more secure. Small companies who are greedy may opt for the former, since they are not normally the target for hackers. The presence of the performance-security tradeoff may slow down the development of AI security.

5 What can managers control

5.1 R&D

It is most crucial for managers at large scale technology companies that use machine learning is to invest in research. As mentioned before, a lack of security in AI systems may lead to serious profit losses. In addition, it may also undermine the company's brand image, which would lead to further profit losses. By doing AI security research internally, companies can convince customers to rely on their systems, as well as prevent malicious attacks from hackers.

5.2 Catch up with enabling technologies

Enabling technologies for AI security include computing power in the distributed systems, storage and hardware. If these enabling technologies become more advanced, research and development in AI security will become more efficient. Therefore, it is important for managers to catchup up with enabling technologies as well.

5.3 Relationship with Policy Makers

Managers should negotiate sooner rather than later with policy makers about security guidelines for their systems, in order to develop systems according to the rules. If a company only discovers AI security policies after the products are developed, it is costly to redevelop their products.

6 The Future

Currently, Google and some academic institutions dominate research in secure AI systems. In the future, other large and medium sized technology companies will also become motivated to develop secure AI systems, in order to build a better brand image and to prevent unnecessary losses caused by hackers. Therefore, the field of AI security has a long way to shift from research focused to development focused. The government will also play a role in setting necessary security guidelines for companies deploying machine learning systems.

References

[1] Xin-Yao Qian and Shan Gao. Financial series prediction: Comparison between precision of time series models and machine learning methods. *arXiv preprint arXiv:1706.00948*, 2017.

[2] Matt Fredrikson, Somesh Jha, and Thomas Ristenpart. Model inversion attacks that exploit confidence information and basic countermeasures. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, pages 1322–1333. ACM, 2015.

[3] Ahmad EL Sallab, Mohammed Abdou, Etienne Perot, and Senthil Yogamani. Deep reinforcement learning framework for autonomous driving. *Electronic Imaging*, 2017(19):70–76, 2017.

[4] Mahmood Sharif, Sruti Bhagavatula, Lujo Bauer, and Michael K Reiter. Accessorize to a crime: Real and stealthy attacks on state-of-the-art face recognition. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pages 1528–1540. ACM, 2016.

[5] Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, and Rob Fergus. Intriguing properties of neural networks. *arXiv preprint arXiv:1312.6199*, 2013.

[6] Nicolas Papernot, Patrick McDaniel, Ian Goodfellow, Somesh Jha, Z Berkay Celik, and Ananthram Swami. Practical black-box attacks against deep learning systems using adversarial examples. *arXiv preprint arXiv:1602.02697*, 2016.

[7] Reuters. A researcher says he fooled face id on apple's iphone x with this diy hack. *Fortune*, 2017.

[8] Mike Ramsey. Carnegie mellon reels after uber lures away researchers. *Wall Street Journal*, 2015.

[9] Muehlhauser. Why ai safety? *MIRI blog*, 2015.