No.12
18th October 2018

HardenStance Briefing

Trusted research, analysis & insight in IT & telecom security **PUBLIC/UN-SPONSORED**

AI in Cyber Security and Cyber Warfare

HardenStance attended KPN's "Alert Online" on 'Artificial Intelligence & Cyber Warfare' in Rotterdam on October 9th. The key take-aways around AI in cyber security are presented first, followed by those relating to broader ethical issues around AI.

- Cyber adversaries are not leveraging machine learning or AI yet – but they will.
- Nation states will likely be the first actors to leverage AI in cyber attacks.
- There are critical ethical concerns that arise with the use of AI in cyber warfare.

The use of AI in Cyber Security

- Mikko Hypponen, F-Secure's Chief Research Officer, told delegates at the KPN seminar that adversaries are not yet using AI to drive attacks. Moreover, he said he was hopeful that adversaries won't start using AI for "at least another two years."

Adversaries are deploying more capabilities designed to drive defensive AI algorithms into learning incorrectly.

- Hypponen said he was hopeful that the vast majority of talent with machine learning or AI skills applying themselves to cyber threats will continue to be drawn to the defender's side in the near term. People with these skills are even more highly sought-after than many cyber security professionals, he said, and most cyber criminals would prefer a highly-paid job.
- When AI-powered attacks do come, examples of the form they take will include polymorphic malware that knows how to change itself and email worms that adjust their message themselves to improve their efficacy.
- Hypponen predicted that nation states will be the first to use machine learning or AI in an offensive capacity. He went on to make the case for inter-governmental rules of engagement around what's allowed and what's not (although one could argue that global rule-making is in the process of breaking down and fragmenting).
- In the meantime, adversaries are having to contend with the increased level of AI deployment by defenders. Hence, adversaries are deploying more capabilities designed to drive defensive AI algorithms into learning incorrectly.
- A variety of different machine learning or AI algorithms are already powering a lot of cyber security products in the field. Many newer generation spam filters use these technologies. A subset of other random examples that HardenStance is aware of include Cylance's CylanceOPTICS (EDR); Fortinet's FortiWeb (WAF); Palo Alto Networks Magnifier (behavioural analytics); and XM Cyber's HaXM (automated APT simulation) platforms.

Broader Ethical Issues with AI

- A number of speakers at the event pointed to the need for appropriate regulation to harness the power of AI in support of human goals. At a general level, Mikko Hypponen stated "we have to define the rules [of AI] while we can". In the context of the use of AI by nation-states in warfare he said that "it must be a human being that decides whether another human being lives or dies."
- Jelle van Haaster from the Dutch Ministry of Defence reminded delegates of Vladimir Putin's public remark in 2017 that "the one who becomes the leader [in AI] will be the ruler of the world." Van Haaster added: "in warfare, rapid decision-making is

critical. AI can help you win. How do we exercise control without compromising the ability to win?" While calling for the clear benefits of AI to be harvested, including in the cyber security domain, KPN's CISO, Jaya Baloo also called for "a Geneva Convention-type equivalent" to limit the use AI in warfare. She also called for an AI community to maximize the benefits and minimize the risks from AI.

- Professor Bob de Wit from Nyenrode Business University raised some eyebrows – and some nervous laughter – by suggesting that "we should not over-estimate the capabilities of judgement that human beings have. If you establish an algorithm of sophisticated rules you might arrive at [outcomes] that are better than those of human beings. If you look at human history, I'm not sure that human beings can consider ethics to be their core competence."

KPN's CISO, Jaya Baloo, called for a "Geneva-Convention type equivalent" to limit the use of AI in warfare.

- Arguing for transparency in AI-enabled systems, Matthijs Pontier from the Dutch Pirate Party, whose platform centres on digital civil rights, argued that "if machines make decisions over us, they should be able to understand human ethical decision-making, take that into account, and be able to explain their decisions to us." He also expressed concern that automation of warfare risks lowering the risk threshold for relatively powerful countries to start wars against weaker ones.
- Making a similar case, Professor Ibo van der Poel, Professor in Ethics and Technology at Delft University of Technology said transparency requires that people need to be able to understand in layman's terms how AI algorithms learn so that they can be adjusted to avoid undesirable outcomes. The test case of "Loomis vs Wisconsin" that he cited is not encouraging here.
- In October 2017, Saudi Arabia became the first country in the world to confer citizenship status on a robot. The robot - "Sophia" – declared that "she" was apparently "very honoured and proud for this distinction." We might reasonably wonder whether the rights of various types of humans – women, journalists, pro-democracy activists, for example - might not merit more attention to their rights than "Sophia" or other robots. The European Union appears divided on this question. The European Parliament believes sophisticated, autonomous, robots should be granted "personhood". The European Commission does not.
- Several speakers and delegates pointed to the incompatibility between a controlled evolution in AI that is democratically harnessed to ethical ends and the commercial race to be first among the world's tech giants. Gerard Smit, CTO Benelux, IBM, nevertheless pointed to the Partnership on AI. This counts many of the world's largest tech companies among its members. It seeks to be "an open platform for discussion and engagement about AI and its influences on people and society." ■

You can replay the full 3 hour livestream of this excellent KPN event here

HardenStance will also be chairing the "Stopping the breach – Securing future networks with the help of AI and virtualisation" session at FutureNet World. This event will be in London on March 26th - 27th and is endorsed by ETSI and the TM Forum.

- HardenStance received no payment – direct or "in kind" – for publishing this Briefing.
- **Contact HardenStance's Principal Analyst:** patrick.donegan@hardenstance.com
- HardenStance received no payment – direct or "in kind" – for publishing this Briefing.
- **Register here** for **free email notifications** whenever new IT and telecom security content is made available by HardenStance.
- **www.hardenstance.com**

HardenStance Ltd Disclaimer of Warranty and Liability

HardenStance Ltd has used its best efforts in collecting and preparing this report. HardenStance Ltd does not warrant the accuracy, completeness, currentness, noninfringement, merchantability or fitness for a particular purpose of any material covered by this report.

HardenStance Ltd shall not be liable for losses or injury caused in whole or part by HardenStance Ltd's negligence or by contingencies beyond HardenStance Ltd's control in compiling, preparing or disseminating this report, or for any decision made or action taken by user of this report in reliance on such information, or for any consequential, special, indirect or similar damages (including lost profits), even if HardenStance Ltd was advised of the possibility of the same. The user of this report agrees that there is zero liability of HardenStance Ltd and its employees arising out of any kind of legal claim (whether in contract, tort or otherwise) arising in relation to the contents of this report.