CYBERSECURITY AND ARTIFICIAL INTELLIGENCE: HOW TO ALLOCATE LIABILITY BETWEEN THE STAKEHOLDERS?

Report on Session 186 of WSIS Forum 2019, Geneva Switzerland (8 April 2019)

Artificial Intelligence (AI) may be used by companies to enhance their security, as well as by attackers to better breach security. Today, white hat hackers are also skilled professionals with serious jobs and human resources teams. The number of players (manufacturers, operators, programmer, trainer, end-users) makes it consequently difficult to allocate liability.

On 8 April 2019, nine professionals from varied professional and regional backgrounds came together in the context of the 2019 WSIS Forum to discuss this complex issue. This panel exchange addressed four main questions:

1. identifying all stakeholders involved in cyberattacks;
2. allocating cliability between them in light of the current legal frameworks;
3. clarifying insurance options; and
4. analyzing regulatory challenges.

The main example used throughout the panel discussions, when illustrating concepts or issues, was that of an autonomous vehicle causing an accident.

Introduction

Dr. Yaniv Benhamou, PhD, Attorney-at-law, Lecturer (IP & Privacy), University of Geneva (YB) introduced the panel, which takes place within the framework of a joint research project on cybersecurity liability conducted by the University of Geneva and the Hebrew University of Jerusalem. He mentioned that discussing allocation of liability in relation to AI is important for three reasons:

- First, because of the multiple stakeholders potentially involved. When a liability issue arises with an autonomous vehicle, for instance, it may involve the driver, the vehicle's manufacturer, the software designers of all applications inside the vehicle, the general programmers, a third party, etc.
- Second, because of AI's rapid transformation combined with its dual impact on cybersecurity: AI development enables enhanced cybersecurity but, at the same time, allows for increased attacks of greater complexity and scope when used for wrong purposes, as appears from the graphic below.
- Third, because understanding the liability issues raised by AI requires a multidisciplinary approach implicating both private and public as well as technical and legal actors.

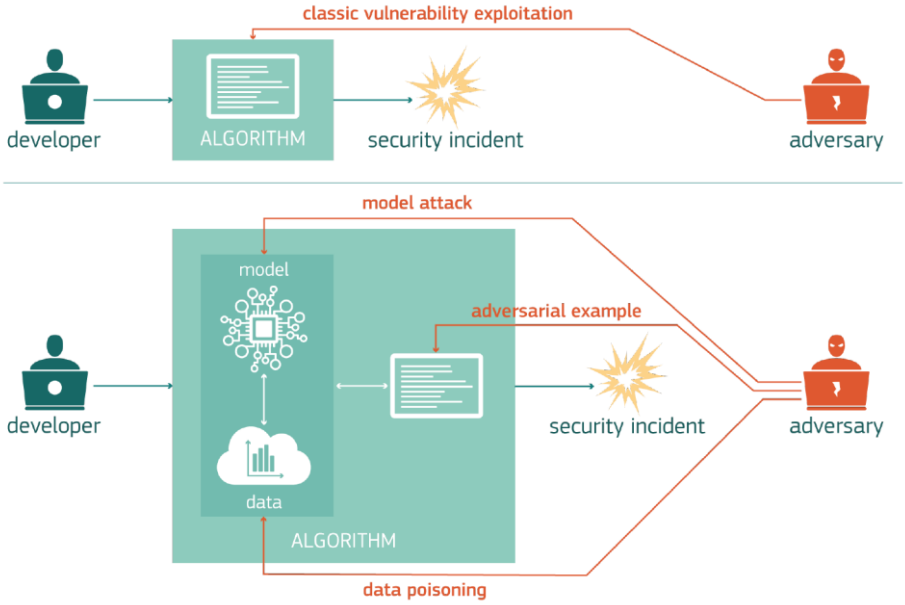

Figure 12. Paradigm shift in the cybersecurity of systems following the introduction of AI components

Figure taken from the report "EU, Artificial Intelligence, A European Perspective, 2018, p. 89

Broadly summarized, AI involves three main components: input, algorithm and output. First, various types of input (data) are fed into the AI system. Then, algorithms process and learn from this input, and eventually create output (new data). The way AI functions creates a so-called "blackbox" issue: since we do not always understand how AI processes the input and generates output, there remains mistrust in the process and it is difficult to allocate liability in the case of a problem. For instance, when a machine scans people's faces and issues an opinion as to their ethnicity, it is not always possible to identify which features were processed, the weight that was given to each feature in the global assessment, whether there are unethical bias in the processing, etc.

From a legal point of view, **YB** reminded that allocating liability in the context of AI and cybersecurity implies a multifaceted analysis of various questions pertaining to intellectual property law, privacy law, database and trade secrets law, unfair competition and practices law, contract law and tort law – the latter being the angle under which the panel discussions were conducted.

1. Mapping the challenges and the stakeholders

This sub-panel first discussed how academic and technical experts use AI and machine learning (ML) techniques to enhance cybersecurity and fix vulnerabilities.

Prof. Stéphane Marchand Maillet, Department of Computer Science, University of Geneva (SMM) started by briefly explaining the technological features of AI used in cybersecurity.

Discussing what makes AI different from other disruptive technologies, he addressed the difference between AI and ML, which are often confused but are factually distinct. Whereas AI is an expert system made of rules encrypted into an algorithm, which then solves a problem or question, ML is different in that the algorithm is not as "smart" at the start, but learns and becomes smarter through processing of knowledge-rich data.

Dr. Olivier Crochat, Executive Director, Centre for Digital Trust (C4DT), EPFL (OC) added that strengthening the ML process and enhancing cybersecurity implies continuous training. For instance, every autonomous vehicle equipped with a ML system may react differently to an obstacle, based on the data to which it has been exposed so far and that has formed part of its "training". Experts at EPFL are working on ML on multiple fronts: they strive to enhance natural language processing, facial recognition and data analysis but also, on the cybersecurity front, they try to understand how to protect ML from malicious threats. However, notwithstanding these very important research areas, experts must also remember that cybersecurity failures may not only result from a vehicle's inability to anticipate, avoid and react to third party attacks on its systems, but also from flaws in the design of the vehicle, improper design of the algorithm, etc.

Mr. Steven Meyer, CEO and co-Founder of ZENData, Cybersecurity services, ZENData Geneva (SM) added that ML plays an important role in the cybersecurity industry. For instance, ML systems incorporated in antiviruses allow identification of abnormal behavior and help detect attacks.

Discussing the challenges of AI in cybersecurity, **SMM** notably identified the following.

- **AI's "blackbox" nature.** Whenever an issue arises, the relevant information that may help allocate liability can often be found in the data processed by the AI system, not the algorithm itself. This causes challenges of **interpretability**, e.g. the need to go back the chain of data processing to understanding which input data led to a specific (faulty) output. For instance, if an automatic vehicle suddenly brakes for no apparent reason and causes an accident, we would need to understand the reason behind this reaction: was it because it wrongly perceived an obstacle on the road? Because it misread a road sign? Etc. AI's blackbox nature is also tied to a **lack of transparency** in current AI systems. With Apple's face recognition, for instance, no one knows what is inside the algorithm and how easily it may be fooled.

- The **high number of potentially involved stakeholders** in AI-related issues further complicates allocation of liability. When an autonomous vehicle is executing the task of automatically locating itself within its environment, for instance, multiple stakeholders are already involved including the AI developer, the hardware manufacturer and software producer. All these stakeholders have a

role to play in ensuring that the detection process is reliable (the vehicle should not be fooled if a road sign is modified with duct tape, for instance). Other parties who may not be directly involved in the development and functioning of the vehicle's AI system also play important roles in ensuring safety and avoiding accidents. Notably, the driver should remain vigilant at all times; the local government should adapt the traffic logistics and rules to the new reality of automatic vehicles, which react differently from regular vehicles; the manufacturer of the vehicle should ensure its physical safety; insurers should offer adequate coverage; etc.

- The **variety of threats** that may "fool" ML and AI systems is also concerning. For instance, a cyber criminal may tamper both with the output results of an AI system (by reverse engineering and modifying a ML algorithm to give different results, for instance) and with the input data that is fed to ML/AI systems – an activity called "data poisoning". In addition, AI is now used to mine social networks of employees and exploit their vulnerabilities in order to hack into companies. To avoid and counter such threats, **SMM** added that specialists and researchers must continue working reinforcing AI's behavior on two fronts: making the system more robust in its *performance* (ensuring that it makes less mistakes to strengthen security) and in its *resistance* to attacks and unpredicted cases that may cause disruption. In order to do so, we must ensure that AI systems are able to detect rare events and learn from them. Indeed, by identifying what is "normality" and what deviates from normality, AI can perform better and with less risks of failure. In other words, the more we understand why decisions are taken, the more we understand the system as a whole and can "close the door" on potential attackers.

Adding on the topic of threats, **SM** reminded the audience that most AI is used for good. However, he noted that both cybersecurity specialists and hackers now invest large amounts of energy and money in understanding and controlling AI cybersecurity systems. This has changed the nature of threats to which we risk being exposed in the future. For instance, only a few years ago, a hacker who had gotten into a person's email account and wanted to send a phishing email needed to spend time going through all of that person's contacts and messages, identify which targets were the most likely to be "good" victims, etc. With the development of AI, however, it is now possible for a hacker to scour through a mailbox and obtain this information automatically, in a few seconds. In addition, innovations such as open source software and cloud computing make it easier for criminals to both hack the engine (neural network) itself but also use ML at their advantage to scale up their attacks.

Concluding this first sub-panel, **YB** summarized the main vulnerabilities of AI as follows:

- (1) the use of AI by a cyber-attacker to penetrate a security system;
- (2) data poisoning (allowing a cyber-attacker to break through a system);
- (3) data design flaws; and

- (4) actions that target human vulnerability.

2. Civil liability: current legal challenges

Building on examples of recent fatal accidents implicating autonomous vehicles who had been confronted with unfamiliar sensory feedback or inputs their guidance systems could not interpret, speakers in this sub-panel attempted to identify who should bear civil liability in these circumstances and under which liability regimes. They also discussed the interface between the various potentially applicable liability regimes.

Mrs. Ria Kechagia, Legal Counsel, Scientific Collaborator, Department of Commercial Law, UNIGE/ C4DT, EPFL (RK) explained that the first legal regime that comes to mind in the context of an autonomous vehicle's accident is product liability which, in Europe, is addressed by the *Product Liability Directive*.[1] This Directive sets out basic rules concerning the manufacturer's liability for safety defects and the user's duty of care in taking all necessary measures to keep a product in good shape according to the manufacturer's instructions. However, these principles are difficult to apply in practice since the Directive was not drafted with AI-driven products in mind. As such, a new legal regime that better accommodates the recent issues raised by such technological developments is needed.

RK also discussed the fact that actual liability rules, when applied in the context of AI-driven products, entail important risks, costs and legal exposure for the manufacturer, which may hinder innovation. However, this issue could be circumvented if the manufacturer, after investing in improving the software found in its product, raises the price of its product to offset some of the costs to the consumer. RK believes that consumers may be ready to pay more for AI-driven products, knowing that their safety is increased and that the manufacturer remains liable if something goes wrong. In addition, insurance companies could play a role by guaranteeing the good quality of the product and decreasing the fees for the consumers.

From a technological viewpoint, **SMM** explained that understanding the chain of decisions that led to an erroneous decision is crucial to allow legal experts to allocate liability. Therefore, in order to establish liability, we must first verify if we can reproduce the "bug" that led to the accident. In practice, however, this is still extremely difficult to do. Even in cases where the algorithm itself is rather simple, the data fed into the algorithm is so diverse and ever changing (in the case of autonomous vehicles, one may think about inputs from cameras, sensors, lasers, microphones, etc.)

[1] Directive 85/374/EEC.

that it is often impossible to reproduce the environment in which the accident happened. For instance, something as trivial as a bird standing for a few seconds on a speed limit panel could be at the source of the vehicle's wrong decision.

OC added that the AI "blackbox" phenomenon is at the heart of the difficulty regarding allocation of liability. Certifications schemes could help vehicle manufacturers in pleading that they took all reasonable steps to prevent the accident. However, it would be more difficult to impose certification schemes on software developers.

In light of these interventions, **YB** summarized that the main issue today appears to be whether we can reproduce an accident or not. If not, we find ourselves in the middle of the AI "blackbox" and may be unable to allocate liability as it is impossible to identify the source of the problem.

Discussing other potentially applicable liability regimes, **Dr. Asaf Lubin, Lecturer at Yale University, Cybersecurity Postdoctoral Policy Fellow at the Fletcher School of Law and Diplomacy, and an Affiliate at the Berkman Klein Center at Harvard University (AL),** added that statements from Volvo's CEO (2015) and the Managing Director of Volvo Australia (2017) seem to suggest a move toward strict liability for manufacturers when an accident occurs and the vehicle is in full autonomous mode. This is also the position adopted by some scholars. Indeed, the manufacturer is best placed to contract out liabilities with other stakeholders along the supply chain. However, adopting a general strict liability model, ignores the possibility for complex real life scenarios and technological issues, which are still likely to arise.

The 2018 U.K. *Automated and Electric Vehicles Act* adopts a different model, under which the driver should be held liable for any fault during autonomous driving, and will be indemnified by his insurance company. The insurance company may then litigate with the manufacturer and other stakeholders for the costs of the claim.

Mrs. Limor Shmerling Magazanik, Managing Director , Israel Tech Policy Institute (LSM) added that the legal profession generally expects to find a definitive answer in allocating liability. However, this is not always possible in the cybersecurity field. It is often very difficult to ascertain who is actually behind an attack, both in real-time and in retrospect. The allocation of liability may also vary depending on whether the vehicle at issue is fully or partially autonomous. Policymakers may simply need to accept that some crucial information might always be missing, and adapt their policies and legislation to this reality. She agreed with the other panelists that as of today, it appears that the solution has been to impose strict liability to compensate for the difficulties in allocating liability. However, we must remember that regulation must always be adapted to the current state of

innovation. If regulation is too strong and/or too early, it might cause a "chilling effect" in the field of AI-driven products that will kill innovation. It is important to keep a balance between all interests at stake.

LSM noted that she sees some similarities between cybersecurity liability legislation and the GDPR. Indeed, the vehicle's manufacturer is akin to the GDPR's "controller" (e.g. entity that is closest to the consumer, initiates the whole scheme, has the ability to make choices and may monitor what happens). Other actors such as software developers are akin to the GDPR's "processors" who also have legal obligations (notably to report to the controller in case of issues), albeit of a lighter scope.

Finally, **LSM** argued that solutions other than legal reforms should also be taken into consideration. Review boards, for instance, are a "softer" way to deal with the risks and challenges posed by new technologies. These boards may notably allow multiple stakeholders (e.g. governments, academia, industry and civilians) to work together in a collegial environment and try to find solutions without fearing strict legal penalties. In addition, because AI raises a whole set of new problems just like the Internet did some twenty years ago, it is important to continue to learn from recent use-cases and to discuss them thoroughly.

AL added that some cybersecurity liability issues will inevitably end up before courts. He gave the example of the *Flynn v. FCA US LLC* (3:15-CV-00855-SMY-RJD) now pending before the District Court for the Southern District of Illinois (scheduled for discovery in August and hearings to begin in October, 2019). The case concerns reports that certain Fiat-Chrysler cars (which have since been recalled) can be hacked through their infotainment system (the UConnect system). These reports vividly demonstrated how hackers could exploit security vulnerabilities and remotely control vehicles as they were being driven, with potentially devastating consequences. Although there has never been a reported case of any car actually being hacked, a complain was filed on behalf of a putative class of consumers who purchased Chryslers. Even though there was no injury in fact, members of the class argued they overpaid for cars which they thought were secure. According to the plaintiffs theory of damages, purchasers reasonably expected and paid for information security when they purchased Chryslers, and are therefore entitled to recover that percentage of the sales price attributable to information security because of the security flaws in the UConnect system. Similar overpayment theories have been attempted in data breach cases, including the Target class action and Cahen, but have been rejected by courts on the grounds there is no reasonable basis to believe consumers considered data security when they made their purchasing decisions. However, in the vehicle security context, judges might adopt a different approach.

While this case is not about automated vehicles but rather mere internet-connected vehicles, it does open the door for Courts to set certain expectations around the cybersecurity protections that car manufactures are required to provide their consumers, and cases like this will ultimately push regulators to move forward. **LSM** agreed and noted that it is essential that judges who preside cybersecurity trials be trained in technology to fully understand the problems at issue and help establish viable solutions. One idea could perhaps be to create specific courts for such technologies; however, this will likely be a very complicated process.

3. Insurance

The speakers in this third sub-panel exchanged on various aspects of cyber insurance, a tool that is becoming more and more important in helping businesses limit their liability in case of cyber attacks.

AL mentioned that a certain number of issues need to be addressed before the cyber insurance market can play a meaningful role in increasing cyber posture and cyber hygiene for policy holders. First, we must consider the insurability of certain types of cyber risk resulting from intentional torts or criminal activity, such as the indemnification of fines for ransomware payments or GDPR violations. In May 2018 DLA Piper and Aon reviewed the insurability of GDPR fines across Europe and found that GDPR fines were only insurable in two countries (Finland and Norway). Most countries did not allow for the insurance (including UK, France, Italy, and Spain) whereas in some countries there was significant ambiguity around such insurance.

Second, we must consider the application of the war exclusion provision in insurance policies, introduced long before the advent of cyber attacks, and determine whether cyber insurance should cover indemnification of state sponsored cyber attacks and acts of cyber terrorism. This is the issue standing at the heart of the Mondelez case which concerns damages caused by the NotPetya virus, allegedly originating from the Russian government, to the American snack company. Mondelez, who was greatly affected by the virus and covered by a Zurich general insurance policy, requested to be indemnified, but Zurich refused alleging the policy's "war exclusion" is applicable and that based on reports by the intelligence community the incident is a war-like event. This case exemplifies the evidentiary and interpretative issues likely to plague the cyber insurance market in the years to come. For example, the Court in *Mondelez v. Zurich* will be required to address the question of whether the NotPetya attack can be attributed to Russia (U.S and U.K. intelligence community statements have yet been corroborated by any specific evidence that can be presented in court). Note that some insurance policies have moved away from the generic war exclusion to a specific "state sponsored" and "cyber terrorism" exclusions which will have to be interpreted in Courts in the years to come.

Because of the ambiguity surrounding the insurability of these cyber risks, businesses are now more hesitant to simply rely on silent cyber coverage (e.g. cyber damage merely being covered under general insurance policies without a specific reference to cyber). Concurrently, there is an increase in the conclusion of explicit (or affirmative) cyber insurance policies by businesses hoping to secure better coverage for these types of risk. The parties to these polices may try to ensure, to the extent possible, in their negotiation that the wording leave not room for such ambiguity.

AL added that another concern companies have in adopting cyber policies is whether the insurance companies have the necessary expertise around cyber risk. He mentioned the recent combined approach set forth by Cisco, Apple, Aon and Allianz that integrates technology, services, and cyber insurance into a single product. This is one way insurers try to elevate that risk. The problem here is that these companies might have conflicting interests, and it is unclear whether all of them will be subject to the same fiduciary obligations that insurers hold towards their consumers.

Regulators should also consider encouraging companies to disclose their cybersecurity standards in a uniquely tailored cybersecurity policy that is separate from their traditional privacy policies. Such cybersecurity policies will not only help mature the market and alleviate information asymmetries, but they could also be interpreted as binding contracts for consumers, at least in those jurisdiction which have recognized the privacy policies as having similar effect.

Finally, the emerging "security guarantees" option provided by information security companies offers its own example of both a promise and a risk. These are products provided by IT security companies (such as ransomeware monitoring services) which are coupled with some guarantee by the company in the case of a hack. For example, beginning in 2016 SentinelOne began offering up to 1,000$ per infected endpoint (and up to 1$ Million per company) in warranty protection against ransomeware attacks. On the one hand these guarantees allow for some indemnification to be provided in the case of an incident directly by the cybersecurity expert, which is a positive thing. On the other hand, these companies are not insurers and yet they offer guarantees which are basically insurance products. This is problematic from a regulatory and oversight standpoint, as these companies are not subject to the requirements that other insurers must meet (*e.g.* obligations around transparency, reporting, fiduciary duties, and training obligations).

Mrs. Līga Raita Rozentāle, Director of Governmental Affairs for Cybersecurity Policy, Microsoft (LRR) shared the perspective of a private company on cyber insurance and AI-related issues. She noted that AI not only has a positive impact on cybersecurity, but may also have, in some cases, a negative impact which may increase risks for private companies. For instance, AI allows cyber attackers to manipulate information more easily. The market for items such as fake sensors is

also growing which suggests that more incidents may happen in the future. Companies must take insurance into account when addressing their cybersecurity issues, and **LRR** suggested that insurers could consider "rewarding" good cybersecurity measures by decreasing premiums for companies reaching a certain level of cybersecurity and taking proactive measures.

4. Regulatory challenges and next steps

The last sub-panel exchanged on regulatory challenges related to AI and cybersecurity, and attempted to outline the most relevant next steps in tackling this complex issue.

Mrs. Stephanie Borg Psaila, Interim Director of DiploFoundation and the Geneva Internet Platform (SBP) addressed the issues faced by policymakers. According to her, capacity development is essential to keep up with AI; indeed, judges and policymakers need to have at least a basic understanding of AI if they want to regulate it.

SBP identified a few of the policy issues related to AI, cybersecurity and the allocation of liability that need to be addressed, including the use of AI by cyber criminals (for instance, criminals using AI to avoid detection) and protection of data (e.g. the necessity of adopting data-centric models in systems development). On the technical side, it is important to improve AI systems, since it is crucial to understand when AI is right or wrong if we want to allocate liability.

Mrs. Katarzyna Gorgol, Adviser Digital Affairs and Telecommunication, Delegation of the European Union and other International organizations in Geneva (KG) discussed regulatory challenges and next steps from a European perspective. In the European Union, the *Cybersecurity Act*, was recently adopted by the Members of the European Parliament and should enter into force at the end of May. This Act notably establishes the first EU-wide cybersecurity certification framework, thus ensuring a common cybersecurity certification approach in the European internal market. This certification framework is meant to be open, inclusive and transparent, and the EU Commission will work with industry standardization bodies to that effect. ENISA, the European Union Agency for Cybersecurity, will also work on certification schemes. This Act indirectly touches upon questions of liability since compliance will be taken into account when allocating liability for cyber incidents.

As for the actual European *Product Liability Directive*, as previously explained by **RK**, it mostly applies to movable and tangible products and as such is not well adapted to new technologies. The extent to which new technologies such as autonomous vehicles fall under this legislation is debated, and the European Commission is working on a report on the challenges posed by AI to this legal framework.

On a more general level, the European Commission has formed two expert groups in the areas of AI and liability and new technologies. It also issued, on the same day at this panel (April 8, 2019) a communication titled "Building Trust in Human-Centric Artificial Intelligence". This document will have practical applications in helping allocate liability since it provides guidance to evaluate AI applications according to seven principles, including accountability of algorithms, explainability and transparency.

Conclusion

In the final minutes before closing of the panel, SM stressed that questions of ethics relating to the use of AI must also be addressed, especially with regard to potential input biases. In the United States, for instance, AI is used to guide the police in identifying where the next terrorist attacks may happen; however, experience has shown that this inevitably leads to discriminatory biases against some sectors of population. The analysis pertaining to allocation of liability for events involving AI needs to include these important ethical considerations.

The panel concluded with a question from a member of the audience, who mentioned that although the debate on liability of autonomous vehicles is relevant, we should first address the most pressing geopolitical issues such as the use of AI as an offensive weapon. Panelists agreed that this topic was highly relevant, but noted that it raises different political questions in which the industry is not directly involved. The industry however still calls for responsible state behavior and will continue to put pressures on governments to ensure that they address these questions quickly and efficiently.

To further the necessary exchanges on the allocation of liability in cybersecurity matters, YB invited the audience to attend the next edition of the Geneva Cybersecurity Law & Policy Conference, organized in the framework of the research project between the University of Geneva and the Hebrew University of Jerusalem mentioned above, on **June 20, 2019** at the University of Geneva.[2] Researchers working on these topics are also invited to participate in the Geneva Internet L@W Research Colloquium held on **June 21, 2019**.[3]

Justine Ferland, attorney-at-law and researcher, University of Geneva

[2] More details available at: https://www.unige.ch/droit/cybersecurity-liability/

[3] More details available at : https://www.unige.ch/droit/pi/research/research-colloquium/research-colloquium2019/