

FINAL GROUP PROJECT

ARTIFICIAL INTELLIGENCE IN CYBERSECURITY

CLIENT

DR. FLORENT FREDERIX, DG CONNECT

STUDENTS

*YUMENG (LINKI) DING | YOU JING LEE | BANUREKHA MOHAN | SIMEON KAKPOVI |
NICHOLAS TAN JIT YANG | SIU SING (ROB) YAU*

ACADEMIC SUPERVISOR

DR. CHRIS COLERIDGE

This Final Group Project is substantially our own work and conforms to Cambridge Judge Business School's guidelines on plagiarism. Where reference has been made to other work this is acknowledged in the text and bibliography.

PROJECT AVAILABILITY: (please tick one box only)

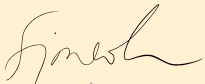

Available (to Cambridge University staff and students only)	<input checked="" type="checkbox"/>
Consult us (prospective readers will be directed to email you for permission to read your project) Our contact email address is:	<input type="checkbox"/>
Confidential until this date:	<input type="checkbox"/>



Final group project 2018-19 - Declaration and disclaimer

This Final Group Project is substantially our own work and conforms to Cambridge Judge Business School's guidelines on plagiarism. Where reference has been made to other work this is acknowledged in the text and bibliography.

This work has been undertaken as part of a student educational project and the material should be viewed in this context. The work does not constitute professional advice and no warranties are made regarding the information presented. The Authors, Cambridge Judge Business School and its Faculty do not accept any liability for the consequences of any action taken as a result of the work or any recommendations made or inferred.

Name	Signature
Yumeng (Linki) Ding	
Simeon Kakpovi	
You Jing Lee	
Banurekha Mohan	
Nicholas Tan Jit Yang	
Siu Sing (Rob) Yau	



Contents

1	Introduction	1
1.1	Project Context	1
1.2	Research Questions	2
1.3	Scope of Report	2
1.4	Summary of Findings	3
2	Review of Current and Future Applications	6
2.1	Cybersecurity in the EU's Context	6
2.2	Limitations of Current Cybersecurity Practices	6
2.3	Advantages Offered by Artificial Intelligence	7
2.4	Defining AI in Cybersecurity	8
2.4.1	AI Applications for Defensive Cybersecurity	9
2.4.2	Use of AI in Cyber-attacks	12
3	Methodology	15
3.1	Survey	15
3.2	Interviews	16
4	Impact and Maturity Framework	19

4.1 Evaluating the Potential Impact of AI-based Cybersecurity Applications	20
4.2 Evaluating the Maturity of AI-based Cybersecurity Applications	21
4.3 Results and Discussion	21
4.4 Limitations	22
5 Understanding Key Trends	24
5.1 Strengths	24
5.2 Weaknesses	26
5.3 Opportunities	29
5.4 Threats	30
6 Policy Recommendations	34
6.1 Focus Funding on Use-inspired Basic Research with Possible Benefits to AI in Cybersecurity	34
6.2 Create a Centralised Online Platform and Implementation Guides for ML-based Cybersecurity Applications	37
6.3 Address Information Asymmetries of the AI Cybersecurity Product Market Through a Labelling Scheme	39
7 References	42
8 Appendix	50
8.1 Appendix A: Interview Questions	50
8.1.1 Questions for academia	50
8.1.2 Questions for cybersecurity vendors	51
8.1.3 Questions for firms employing cybersecurity services	51
8.1.4 Questions for governmental bodies and policy think-tanks	51

8.1.5 General Questions (other)	51
8.2 Appendix B: Potential Impact Data	53

Abstract

This report ^[1] analyses the state of Artificial Intelligence (AI) applications in cybersecurity and how they can be leveraged by the European Commission to support existing efforts in that domain.

The report has four key products:

1. A comprehensive literature review of offensive and defensive applications of AI in cybersecurity.
2. A matrix evaluating the potential impact and maturity of the identified applications.
3. A SWOT analysis that compiles expert opinions ^[2] on the potential benefits and challenges firms face in implementing these applications.
4. A list of actionable recommendations that would enable the European Commission to lower adoption barriers for AI in cybersecurity.

¹This report is a capstone project for the MPhil in Technology Policy course at the Cambridge Judge Business School. The client, the European Commission, sought a comprehensive assessment of the use of Artificial Intelligence in the cybersecurity domain that can be leveraged to support cybersecurity efforts in the European Union.

²We are immensely grateful to the individuals and organisations who generously contributed their time and experiences to this project.

1. Introduction

1.1 Project Context

The rise of connectivity has underpinned the growth of societies and economies in the digital age but has also made them more vulnerable to cyber threats (EC, 2013b). Thus, cybersecurity has become inextricably linked to prosperity and security (EC, 2017b). Cybersecurity refers to measures that protect information systems against various threats. It seeks to safeguard the confidentiality, integrity, and availability of data (ECA, 2019). In the past few years, news headlines have reported private firms losing millions of consumer records in data breaches (Armerding, 2018), state actors using cyber-attacks to influence elections (Burt, 2019), and hackers using Internet of Things (IoT) botnets to cause internet outages (Agerholm, 2016).

Remarkably, 80% of companies in the European Union (EU) were victims of cybersecurity incidents in 2016 alone (EC, 2017b). In addition to recent cases of the ransomware WannaCry within the EU, this has elevated cybersecurity in the EU's political agenda (Barrinha and Farrand-Carrapico, 2018). The European Commission (EC) has long acknowledged the importance of cyberspace in promoting the growth of its economy and internal market. Hence, it has endeavoured to promote cybersecurity as part of its Digital Single Market (DSM) strategy (EC, 2013a).

Within the DSM strategy, the EC has also singled out AI as a strategic area that can be used to benefit EU citizens and the economy (EC, 2018a). As proof of that commitment, the EU is raising AI investments by 70%, up to 1.5 billion euros, under its Horizon 2020 research and innovation programme (ibid.). Furthermore, the EU has recently passed a law to improve data-sharing (EC, 2018d) and is leading the way in AI regulation by supporting the development of an Ethics Guidelines for Trustworthy AI (AI HLEG, 2019).

AI has been a transformative force in a wide range of domains, including healthcare, finance, and advertising (ENISA, 2019d). Although it can also be used to fill existing

gaps in cybersecurity (Mittu and Lawless, 2015), there is limited documentation of AI's level of development in various applications and of its critical issues.

1.2 Research Questions

In this report, we explore how the EC can utilise AI to support cybersecurity efforts in the EU. Specifically, we:

1. Examine how has AI been used in cybersecurity so far
2. Assess the impact and maturity of these applications
3. Evaluate the key future trends and challenges of AI in cybersecurity
4. Craft recommendations to strengthen the EU's support for the adoption of AI in cybersecurity

The rest of the report is as follows: Section 2 examines the intersection of AI and cybersecurity, and reflects on current uses of AI by both attackers and defenders through a literature review. Section 3 describes the study's methodology. Section 4 presents a framework for evaluating various AI-based cybersecurity applications. Section 5 uses a SWOT analysis to discuss critical issues identified by 33 informed professionals. Section 6 draws on lessons from the previous sections to formulate viable policy recommendations.

1.3 Scope of Report

Although some may describe cybersecurity in broader terms, the discourse in this report is principally limited to the defence of networks and information systems. This discussion refrains from engaging certain areas such as disinformation, deep-fakes, and image recognition. Also, AI, like all algorithms, is susceptible to exploitation. Although the need for cybersecurity to secure AI applications in all domains deserves a lengthy discussion, considerations of AI vulnerability in this report are limited only to a cybersecurity context.

1.4 Summary of Findings

In this report, we examined how the advancements of AI in cybersecurity could contribute to EC's existing efforts in the field. To meet this goal, our team conducted an extensive literature review and compiled the opinions of numerous qualified experts on the subject.

We first examined current and future applications of AI in cybersecurity for both attackers and defenders. Defensively, we found that AI could be used for various applications of anomaly detection and workflow automation. Likewise, attackers can utilise AI to enhance payloads and to manipulate defensive systems.

Defensive Applications
Malware detection
Intrusion detection
User and Event Behavior
Phishing and spam detection
Vulnerability management
Threat Intelligence
Cybersecurity workflow automation
Offensive Applications
Automating Phishing
Enhancing Malware
Adversarial AI

We followed up by evaluating the impact and maturity of each of these applications using a matrix. Although some applications of AI are mature, applications in other categories, like vulnerability management and threat intelligence, lag but would provide immense value if developed.

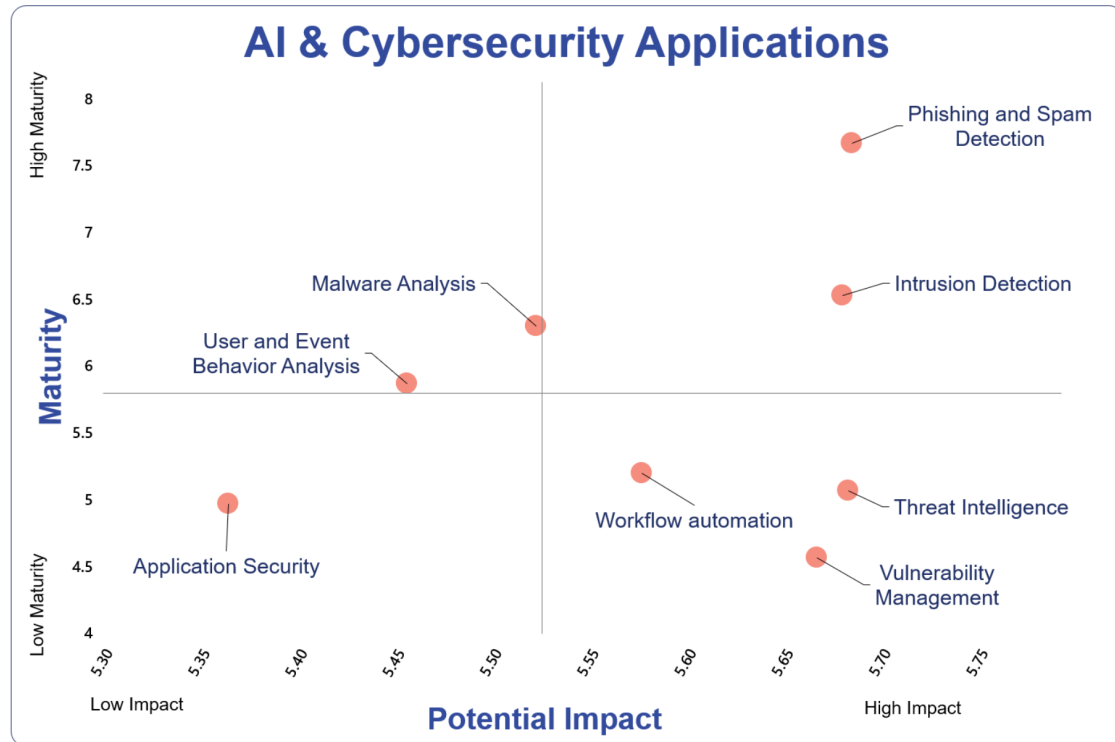


Figure 1.1: Matrix showing the impact and maturity of various AI applications in Cybersecurity

Building on the literature review and the framework, we performed a SWOT analysis using insights from 33 interviewees. Our analysis revealed that AI applications in cybersecurity are nascent, but offers tangible benefits to cybersecurity teams today. AI is also poised to enhance the speed, scope, and consistency of security analysis if deployed universally. Further, we identified various barriers faced by cybersecurity teams when adopting AI-based tools.

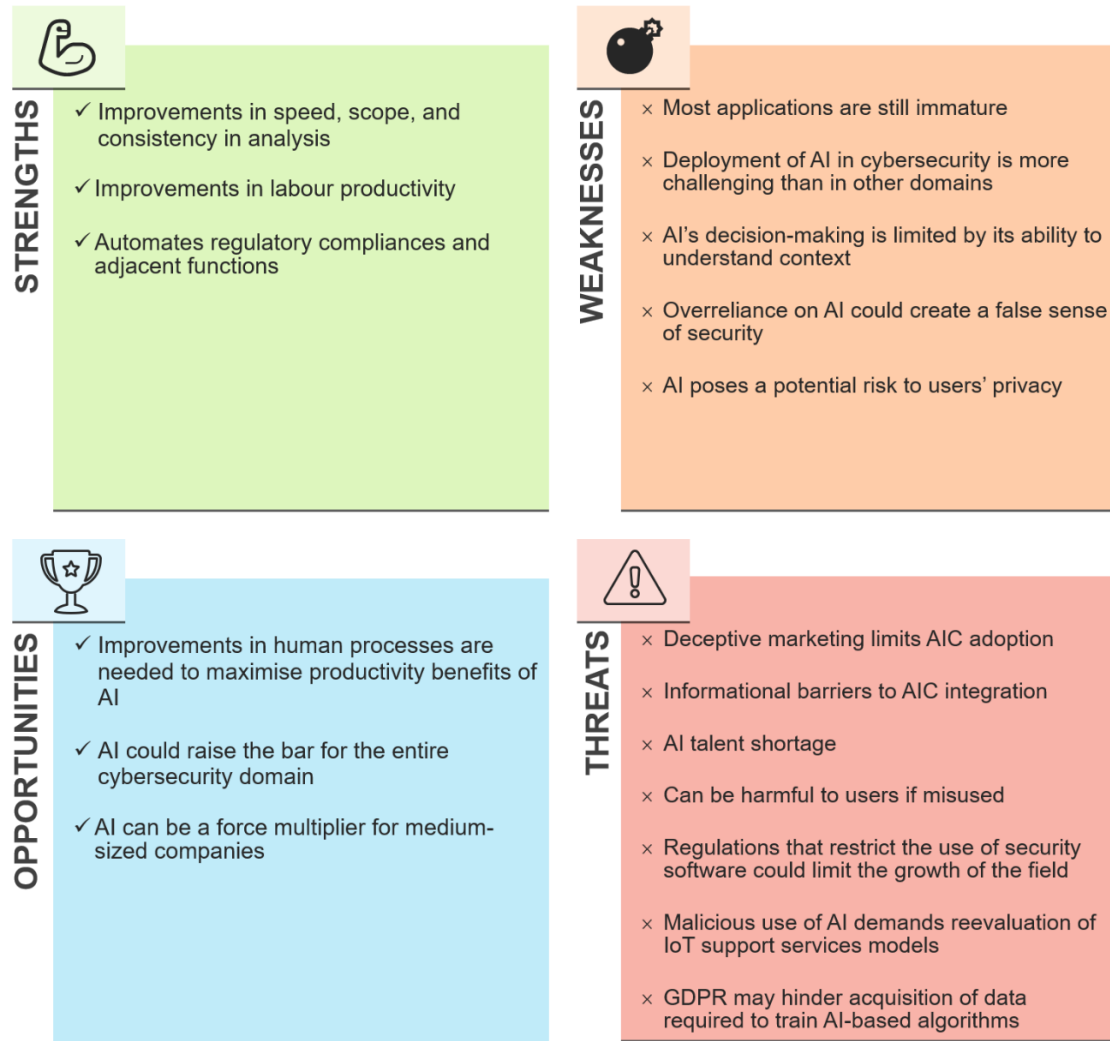


Figure 1.2: SWOT Analysis of AI in Cybersecurity

Three of these are not addressed by existing policies, including: (1) Difficulties adopting AI-based cybersecurity tools, (2) Exaggerated vendor claims and (3) Technical immaturity. Subsequently, we proposed three actionable policy recommendations to address these barriers:

1. Provide funding for basic use-inspired research in AI with possible benefits to cybersecurity, to promote technical maturity of applications.
2. Create a centralised electronic platform with detailed implementation guidelines to eliminate knowledge barriers
3. Create a quality labelling scheme to address current information asymmetries

2. Review of Current and Future Applications

Through an extensive literature review, we examined gaps in the current cybersecurity landscape that can be addressed by AI. Further, we identified defensive and offensive applications of AI in cybersecurity.

2.1 Cybersecurity in the EU's Context

Within the EU, the security of networks and information systems is critical to the economic and social goals of the DSM (European Parliament and Council of the European Union, 2016). As a result, the EU has endeavoured to strengthen its digital autonomy and create the world's safest digital environment (ECA, 2019). Specifically, it plans to prepare member states against cyber-attacks, build a culture of security in critical sectors, and enhance cooperation among member states (ENISA, 2016).

To that effect, the EU has undertaken several critical measures, chiefly founding the European Union Agency for Network and Information Security (ENISA). In addition to providing support to EU member states, ENISA oversees the implementation of the EU certification framework for cybersecurity products, which aims to improve the EU cybersecurity market. It also helps implement the Network and Information Security Directive (NIS Directive), which was enacted to increase cybersecurity competency across member states (EC, 2013a).

2.2 Limitations of Current Cybersecurity Practices

Despite increased spending on cybersecurity measures, there are two key challenges associated with cybersecurity that are difficult to address: (1) the growing attack

surface of organisations and sophistication of cyber-attacks, and (2) the cybersecurity skills shortage.

Increased attack surfaces and sophistication of cyber-attacks

The rising sophistication and volume of cyber-attacks have made firms more vulnerable to cyber threats. A 2018 survey of IT professionals cited these two factors as the primary reasons for increased susceptibility to cyber threats (Baker, 2018). For example, adversaries are engineering new malware strains to self-propagate through networks without the need for human operators (Cisco, 2018). They are also learning to use legitimate techniques and services, such as encryption and cloud services, to evade detection, and are taking advantage of unprotected gaps in defenders' networks, namely unmonitored IoT devices (ibid.). Likewise, the increasing adoption of technologies such as IoT devices, smart technologies, coupled with the implementation of company practices such as bring your own device (BYOD), has exponentially increased the attack surface through which adversaries can target users and firms (Yeboah-Boateng and Boaten, 2016).

Cybersecurity skills shortage

The increased focus on cybersecurity by governments and firms has generated a high volume of demand for cybersecurity professionals. However, the limited pool of cybersecurity professionals is unable to meet this demand, resulting in at least 1 million unfilled cybersecurity jobs globally (Cisco, 2015). This shortage may persist, leading to an estimated 3.5 million vacant cybersecurity jobs by 2021 (Feiman, 2019).

While joint public and private partnerships can seek to mitigate this deficiency by training additional cybersecurity professionals (EC, 2017c), they are unlikely to be meet demand in the long run. To bridge the shortfall in the cybersecurity workforce, firms and cybersecurity providers will need to rely on automation increasingly.

2.3 Advantages Offered by Artificial Intelligence

According to the EC, "AI refers to systems that display intelligent behaviour by analysing their environment and taking actions - with some degree of autonomy - to achieve specific goals" (EC, 2019, p. 3). This involves the notion of rationality, which gives the system the ability to carry out a wide range of intelligent functions,

that encompasses decision-making and optimisation of resources in pursuing certain goals with certain criteria (EC, 2019). AI has been applied to a plethora of problems in various industries and offers a wealth of possibilities when applied to cybersecurity challenges.

AI can enable improved responses to cyber threats

The use of AI can allow cybersecurity teams to tackle highly sophisticated and complex cyber threats with greater ease and efficiency. AI makes it possible to sift through copious amounts of data within an organisation to identify vulnerabilities and malicious code as well as any other anomalous patterns and threats present (Golden and Johnson, 2017). This allows attacks to be detected much earlier (Mackinnon, 2018). Furthermore, machine learning enables cybersecurity teams to automate tasks. They can design algorithms that learn iteratively from existing data without additional human input (Mackinnon, 2018).

Mitigating the cyber-talent shortage

The automation of cybersecurity work by AI allows the efforts of professionals to be directed away from mundane, labour-intensive tasks to those that are more strategic and require problem-solving (Golden and Johnson, 2017; Mackinnon, 2018). According to a 2017 study by Golden and Johnson, nine out of ten cybersecurity professionals believe that AI can help fill the gaps left by skill shortages (Golden and Johnson, 2017). A recent report by the Ponemon Institute (2015) suggests that savings from AI automation in cybersecurity tasks can amount to 21,000 working hours or USD\$1.3 million annually on average.

2.4 Defining AI in Cybersecurity

When discussing the applications of AI in the field of cybersecurity, it is essential to remember that AI is a broad field centred around machines making autonomous decisions (West, 2018). Most uses of AI in cybersecurity have been in applications of machine learning (ML), which is one of many subsets of the greater AI field. In ML, machines are taught to use data to make decisions which have not been explicitly preprogrammed (Crowdstrike, 2018). ML can be further broken down into Shallow Learning and Deep Learning, both of which have been applied to cybersecurity. The

former requires domain expertise to identify relevant data attributes, while the latter uses a multi-layered model of the data to select features autonomously (Apruzzese et al., 2018). Still, applications of AI outside of ML exist and are becoming more common, including the use of natural language processing (NLP), and expert systems.

2.4.1 AI Applications for Defensive Cybersecurity

The following section highlights eight of the most commonly cited applications of AI to support defensive cybersecurity efforts.

Malware detection

Machine Learning can be used to identify malicious files. Traditional antivirus (AV) systems relied primarily on matching unique signatures to find malware. However, this meant that AV researchers would need to dissect and analyse the code of new malware strains to generate signatures to update their engines (Cylance, 2017). As the number of malware samples increased exponentially in the late 2000s, it became impractical to create general signatures for all new variants. The application of ML allowed AV engines, without being fed any new data, to determine whether newly encountered files are malicious, and to filter them accordingly. Hundreds of research papers have been published on the use of both shallow and deep ML techniques for malware detection (Apruzzese et al., 2018). In addition, the most popular endpoint protection products on the market boast of using ML as part of their malware detection solution (Infosec, 2018).

Spam and phishing detection

Much has been written on the use of ML for email filtering. ML can be used to detect phishing emails, which constitute the top vector for malware delivery and a frequent vector for financial fraud (Pham, 2018; Infosec, 2019). For over 18 years, Google has used ML techniques to filter emails (Newman, 2018). More recently, Gmail announced that its ML framework, Tensorflow, enabled the platform to block an additional 100 million spam messages daily (Vincent, 2019).

Intrusion detection

ML can also be used to identify anomalies in network communications. Intrusion detection systems (IDSs) are often deployed to identify unwanted activity on a host or network (Rozenblum, 2001). Much like malware detection, conventional IDSs rely on static signatures crafted by analysts. By applying ML techniques to IDSs, security teams can identify threats that would have otherwise been missed by static signatures or human analysis (Kostas, 2018). This can be accomplished either by looking for patterns that deviate from normal usage, or by looking for known patterns of attacks (Tsai et al., 2009). Anomaly-based methods of identifying intrusions are more likely to detect zero-day attacks, or attacks that have never been seen before (Kostas, 2018). In a survey of IT practitioners, respondents claimed, on average, to detect 63% of zero-day exploits when using AI-enabled solutions (Ponemon Institute, 2015).

User and entity behaviour analysis

Much like intrusion detection, user and entity behaviour analytics (UEBA) technologies function by creating baseline profiles of normal behaviour for users and devices on a network and flagging any activity which is atypical of these generated profiles (Howarth, 2018). The premise of UEBA technologies is that they can detect signs of malicious activity even when the adversary has already gained access to the network. If used effectively, UEBA solutions can help companies detect insider threats, find compromised accounts, identify brute-force attacks, and detect unauthorised changes in user permissions (Brook, 2018). Beyond protecting corporate networks, the use of UEBA can be useful for other tasks requiring anomaly detection, including fraud detection in financial services or defending industrial control systems (Polyakov, 2018). A 2018 Gartner report anticipates that spending on stand-alone UEBA solutions will amount to \$352 million by 2020 and that UEBA technologies will be embedded in 80% of threat detection solutions by 2021 (Sadowski et al.).

Vulnerability management

Vulnerabilities in software often exist due to subtle errors made by software developers. The bugs that lead to these vulnerabilities are usually found when developers manually tread through the source code, or when those bugs are publicly reported to the Common Vulnerabilities and Exposures database (Russell et al., 2018). Properly trained ML systems can be well suited for finding vulnerabilities in software in a

more automated fashion. This would allow software developers to patch bugs before they are exploited by attackers (ibid.). Of course, ML tools can likewise be used by hackers to find vulnerabilities to exploit, especially against IoT devices that are often poorly patched (Schneier, 2018). This autonomous race to find vulnerabilities in devices may become critical as Gartner predicts 20 billion connected devices will be in use by 2020 (van der Meulen, 2017).

Threat intelligence and workflow automation

The use of AI can help cybersecurity analysts automate threat intelligence activities. AI-based tools could help analysts take better advantage of open-source intelligence (OSINT) data from various sources, including blog posts and social media (Vadapalli, Hsieh and Nauer, 2018). For example, NLP could be used to derive intelligence from social media posts to help analysts track vulnerabilities and threat campaigns (ibid.). Analysts can also use supervised learning methods to sort and prioritise data and alerts, and unsupervised learning methods to help connect actors, events, and activities (Ettinger, 2019). This would be especially useful for understaffed teams with limited time and resources.

In the long term, cognitive computing platforms offer the possibility of automating much of the work of threat investigations. Cognitive computing systems attempt to mimic human thought using self-learning algorithms. The most popular of these is IBM's Watson, which learns to "reason" by processing vast amounts of data (Marr, 2016). Cognitive computing technologies have been applied to various challenges in government, health-care, and financial services, albeit with mixed results (Dignan, 2017). When it comes to cybersecurity, vendors of these technologies claim it will eventually be able to consume data from millions of information sources and produce actionable threat intelligence tailored for individual companies. Moreover, they suggest that cognitive computing technologies will be able to identify threats and vulnerabilities, and propose appropriate mitigations. This would substantially reduce the time needed for security analysts to investigate those threats (Koslowski and Felle, 2018).

Other Applications

Beyond the applications described above, many other applications of machine learning and other AI techniques exist including classifying malicious Uniform Resource Locators (URLs), securing mobile and web applications (i.e. Application Security),

and strengthening encryption algorithms (Blackledge, Bezobrazov and Tobin, 2015; Webroot, 2017; Shraven, 2018).

2.4.2 Use of AI in Cyber-attacks

As explored in the previous sections, the application of AI in cybersecurity offers myriad benefits. However, it is also essential to consider the possibility of AI being exploited for offensive cyber applications, intensifying current and future cyber threats. The use of AI to bolster cyber-attacks reinforces the need for AI to be used by defenders to level the playing field.

Phishing and spam

AI has the potential to increase the scale and efficiency of phishing attacks exponentially. This is because AI offers the ability to trawl through data quickly and complete tasks that would otherwise be done by humans. When exploited for offensive applications, AI can allow cybercriminals to increase the scale and rate of attacks, as well as the number of possible targets (Jowitt, 2018a). For instance, AI could be used for automated spear-phishing, where NLP can be used to develop personalised emails that are highly convincing to unsuspecting victims (Brundage et al., 2018). Online criminals have already employed basic ML algorithms to increase the efficiency of attacks, and are capable of profiling targets much more efficiently (Jowitt, 2018a). This is in contrast with traditional spear-phishing attacks, which require attackers to craft targeted-phishing messages manually (Jowitt, 2018a). This could enable attackers to launch large-scale attacks with relative ease, which in turn would reduce the time and costs they incur (Lewis, 2018; Ismail, 2019). Standard cybersecurity measures would also struggle to cope with the scale of such attacks.

Malware

AI algorithms can be applied to existing malware variants, improving their ability to avoid detection by standard cybersecurity measures (Mackinnon, 2018). The potential for new types of AI-enhanced malware has been documented in literature, namely (1) Polymorphic malware and (2) Malware capable of identifying human targets.

AI can be used to speed up polymorphic malware, allowing it to change its code

frequently to evade detection (Ismail, 2019). Over the past year, organisations have reported the use of such advanced malware, which are capable of adapting behaviour to prevent identification (Brundage et al., 2018; Ismail, 2019). These malicious files may render conventional cybersecurity and network protection tools like blacklisting useless, as security scans would have difficulties detecting them.

In addition, AI enables the creation of malware that can identify human targets based on specific characteristics. At the 2018 Black Hat security conference, IBM revealed DeepLocker, an AI-based malware sample capable of hiding and remaining dormant within trusted carrier applications, such as video conference software, until a specific target is reached (Menn, 2018). The malicious file can then enable its malicious intent by identifying the target through facial recognition, geolocation or voice recognition (Stoecklin, 2018). The approach is akin to a sniper's attack, where the malware is created explicitly for a specific target and is difficult to detect otherwise (Jowitt, 2018b; Menn, 2018). Although malware like DeepLocker has been demonstrated purely in a research context, it exemplifies the types of targeted malware that could be developed by advanced actors in the future using AI.

Manipulation of defensive AI systems (adversarial AI)

AI allows cybercriminals to bypass AI-based defensive systems by poisoning these systems with adversarial data. Given that AI and ML are highly dependent on training datasets, this reliance can be exploited to allow hackers to bypass the very security measures designed to protect the system. For instance, hackers can deceive the AI models used for cybersecurity by tampering with the datasets used (Loucks, Schatsky and Davenport, 2018). As it is almost impossible for the AI model to detect such adversarial inputs and to differentiate them from the correct training dataset, this makes it possible for hackers to bypass the cybersecurity system and infiltrate the network (Kobie, 2018).

Research has shown that AI can be manipulated when training datasets are tampered with (Eykholt et al., 2017; Polyakov, 2018). Neural networks, in particular, have intrinsic blind spots that can be exploited by introducing perturbations and distortions. For instance, a research team showed that the introduction of black and white stickers on stop signs could disrupt the AI in driverless cars and cripple their ability to recognise the stop sign, even though the perturbations are easily recognisable by the human eye (Eykholt et al., 2017). A team at Google went one step further and developed a Universal adversarial patch capable of undermining an AI system when added to any scene (Brown et al., 2017). While these experiments have largely been

conducted on AI in driverless cars, they nonetheless show that deep neural networks are indeed vulnerable to adversarial inputs that introduce slight perturbations.

In the context of cybersecurity, this means that AI models used for cyber protection can be misled to yield inaccurate results. This not only highlights the limitations of AI in cybersecurity, but also emphasises that intrusions will only become more viable with the help of AI, putting organisations at greater risks.

3. Methodology

Through the literature review, we have highlighted the importance of AI applications in cybersecurity and provided an overview of AI uses for both defensive and offensive cyber purposes. However, gaps do exist in the current literature, as limited information is available on the maturity of various AI-based cybersecurity tools. Most reports only provide an unstructured overview of AI applications and do not discern between speculative and real applications. Further, we found that key future trends, opportunities and threats of AI applications in cybersecurity are mostly unexplored. This necessitates the use of alternative research methods to achieve the remaining objectives of this paper.

3.1 Survey

We surveyed 30 informed cybersecurity professionals to evaluate their perceived maturity of AI applications in cybersecurity. This data was used to construct an impact and maturity matrix in the next section.

Respondents were asked to rate the applications of AI/ML on a scale of 1 to 9, based on their perception of maturity. They were provided with a set of descriptions to ensure that their understanding of various applications is uniform (Table 1). The rating system is modelled after the Technology Readiness Level (TRL) scale, created by the UK government, which enables uniform, consistent evaluation of technical maturity across various types of technology (GOV.UK, 2014). The descriptions of the ratings were modified to better cater to the research, where 1 indicates a speculative application and 9 represents an application being proven and effective.

Application	Description
Malware detection	involves the identification of malware, as well as the determination of the function of a malware sample
Intrusion detection	Refers to the monitoring of network traffic to search for suspicious activity or threats, issuing alerts when such activities are identified
Phishing and spam detection	Involves the scanning and filtering of phishing and spam emails
User and Event Behavior	Refers to the process that takes note of the normal conduct of users, and detects any anomalous behaviour or instances when deviations from these “normal” patterns occur
Application security	Primarily refers to the use of ML for identifying or blocking malicious requests to an application
Vulnerability management	Refers to the continuous process of identifying, testing, mitigating and reporting vulnerabilities in applications, systems, software, and IoT devices
Threat Intelligence	Involves the identification, collection and analysis of information about current or potential attacks that threaten an organisation or system
Cybersecurity workflow automation	Refers to the automation of common cybersecurity tasks normally done by human analysts (e.g. triaging alerts)

Table 3.1: Common applications of AI in cybersecurity and their descriptions.

3.2 Interviews

To gain a better understanding of critical future issues on the use of AI in cybersecurity, we have also interviewed 33 professionals across more than 10 different countries who were informed on AI and cybersecurity (Figure 1). These were technology managers, engineers, consultants, professors, and lawyers who have had relevant experiences. The interviewees were classified into general groups based on their organisations:

1. Cybersecurity vendors
2. Firms employing cybersecurity products

3. Academics, NGOs, and think-tanks
4. Governmental bodies

The interviews were conducted using the 7 Questions interview technique. This technique is valuable for capturing strategic insights from multiple parties (GOfS, 2017). It facilitates the exploration of multiple future strategic issues on a topic. We used it to identify a range of opinions on key issues from our diverse stakeholders. Additionally, policy considerations and recommendations for the EC would need to take a holistic approach in examining the impacts of said policies to the different stakeholders to ensure that the policies achieve the goals of effectiveness, efficiency and equity (Morestin, 2012).

Two primary considerations shaped the geographical spread (Figure 1) of the interviewees: 1) the need to focus on EU interviewees to enable more accurate policy considerations for the EC; 2) the proportion of global cybersecurity funding. As the United States contributes 75% of global cybersecurity funding (CBIR, 2016), having a sizable proportion of US interviewees was crucial for understanding the possible considerations and challenges of adopting AI-based tools cybersecurity, from a country that is at the forefront of cybersecurity developments.

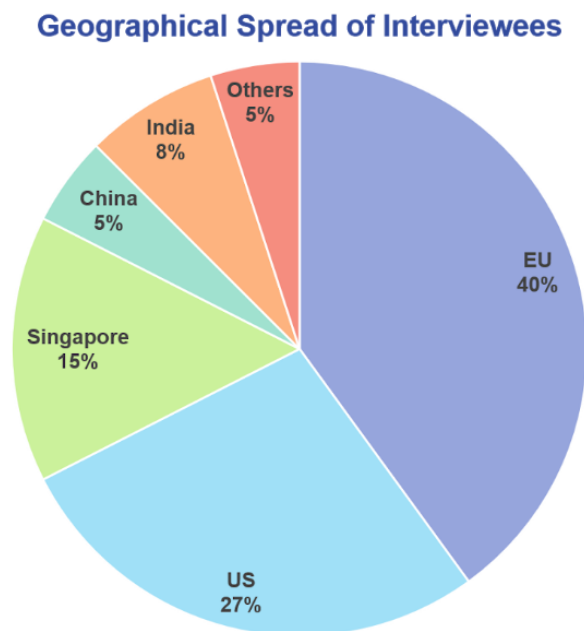


Figure 3.1: Geographical Spread of Interviewees

These interviews were conducted either physically or virtually and lasted between 30 minutes to an hour. Interviewees were prompted with a targeted set of questions

but were given the flexibility to expand on specific areas of interest or expertise as appropriate. The interview questions and the details of the interviewees are presented in Appendix A and Annex A, respectively.

4. Impact and Maturity Framework

The literature review presented in section 2 illustrated the potential benefits and harms of AI in cybersecurity. Based on the literature review alone, however, it is difficult to differentiate the applications that are still in their infancy from those which can provide immediate value. Moreover, it would be constructive to analyse the relative potential impact of each application. The intersection of these criteria may highlight fertile areas of focus for policymakers.

Accordingly, this section quantifies the relative impact and maturity of the eight general applications of AI in cybersecurity. As mentioned previously, these were identified as the most commonly cited applications in the literature review.

They are:

1. Malware analysis
2. Phishing and spam detection
3. Intrusion detection
4. Vulnerability management
5. Threat intelligence
6. UEBA
7. Application security
8. Workflow automation

4.1 Evaluating the Potential Impact of AI-based Cybersecurity Applications

We define the potential impact of AI applications as the sum of time-savings and cost of incidents averted. Based on our literature review, these are the most practically quantifiable benefits. Cost-savings include the amount of money that would be saved on losses incurred from cyber-attacks in the respective areas of cybersecurity. They are drawn from a study by Accenture and Ponemon Institute (2017) on the cost of cybercrime. Time savings are based on the number of labour hours saved weekly on various tasks through the use of AI. They are obtained from a study conducted by Ponemon Institute (2018) on the value of AI in cybersecurity. These labour hours were converted into monetary values using an average cybersecurity wage rate of US\$62.50 an hour (consistent with the 2018 Ponemon report). The final impact score is standardised by taking the logarithm of the combined time and costs savings.

$$\begin{aligned} \text{Potential Impact} = & \log(\text{Time savings [in dollars]} \\ & + \text{Cost of incident averted [in dollars]}) \end{aligned} \quad (4.1)$$

$$\begin{aligned} \text{Time savings} = & \text{Weekly labour hours saved} \times \text{Weeks in a year} \\ & \times \text{Cybersecurity Wage Rate} \end{aligned} \quad (4.2)$$

Three assumptions underpin the values used to approximate the impact of each AI application. Firstly, we assume there is only one way to avert a particular cyber threat. This means that the applications are non-overlapping and that there are no spillover benefits between each application. Secondly, we assume that each application can avert most if not all cyber-attacks relevant to the respective application, to generate the cost savings equal or close to that of the losses attributed to the attack vector. Finally, we assumed that the benefits generated by each application are only mapped onto the cyberthreats averted and that there are no other quantifiable benefits or externalities besides the prevention of a cyber-attack.

4.2 Evaluating the Maturity of AI-based Cybersecurity Applications

We define maturity as the state of development of an AI application, which is determined by its current state of usability in the field. This measure acts as an instrument to separate applications that are speculative from those that are proven and effective. Actual data on maturity is elusive, given that most research papers do not reflect the viability of such technologies in non-theoretical settings, and vendors have incentives to overstate their capabilities to attract customers. As such, we obtained this data through a survey of the experts detailed in the methodology. As a caveat, we assumed that respondents, despite their diverse specialisations, had a shared understanding of each application and notion of maturity according to the description provided.

4.3 Results and Discussion

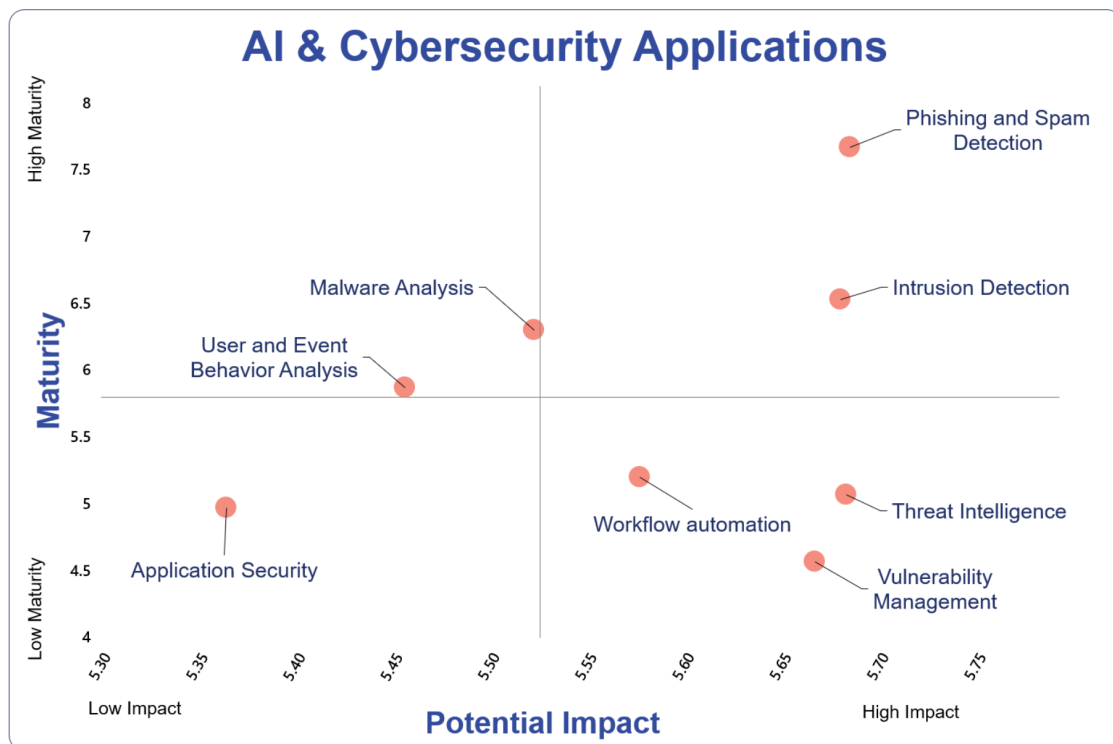


Figure 4.1: Matrix showing the impact and maturity of various AI applications in Cybersecurity

As the matrix serves to gauge the relative impact and maturity of various AI

applications in cybersecurity, the “high” and “low” indicators shown are relative measures for comparison (Figure 2). The lines in the graph indicate the average values of impact and maturity, respectively.

According to the matrix, phishing and spam detection and intrusion detection are applications of a high impact and high maturity. These are useful applications for organisations to pursue if they wish to derive high value in the short run. Although threat intelligence and vulnerability management could provide high value for cybersecurity teams, they are perceived as currently underdeveloped and may benefit from further investments in research.

It is important to note that each application represents a broad umbrella of applications so implications should be drawn contextually. After all, the value of a tool for a specific organisation still depends on a broad set of factors including its size, industry, and portfolio of cybersecurity defence mechanisms. For example, an e-commerce company may derive more value from the use of AI for application security than for threat intelligence, contrary to what the matrix suggests.

4.4 Limitations

We acknowledge that limitations do exist in this framework, due to some of the assumptions made about the data obtained. First, due to the limited data available on cost-savings associated with each application, we were unable to account for other benefits accruing to each application, which may not pertain directly to a cyberthreat. Such benefits include the money saved from the deployment of assets and capabilities such as preventive measures, as well as the associated administrative costs.

Next, each of the applications may have overlapping benefits, so the impact of applications may be lower (or higher) than the value they were assigned. For instance, phishing and spam detection may also help prevent infections by malware by reducing the likelihood of users clicking and enabling malicious attachments in email. Likewise, intrusion detection and threat intelligence may also have spillover benefits in helping to detect malware. Hence the time and cost-savings generated cannot be attributed to malware analysis alone.

Finally, a certain level of ambiguity does exist in the ratings of maturity level. These ratings are based on individual interpretations of the field in general, and the applications associated. Researchers and experts specialising in a particular application may be more optimistic about its maturity than others that may not

be as informed. However, this would likely have been addressed by averaging the ratings, especially given that the experts were not chosen systematically within the field, and most likely have varied experiences.

5. Understanding Key Trends

Following the literature review, this section consolidates the main points raised in the interviews to provide a better understanding of the key trends of AI-based cybersecurity tools. We use a SWOT analysis to examine the strengths, weaknesses, opportunities, and threats of AI in cybersecurity (See Figure 3).

In this model, strengths and weaknesses refer to the inherent advantages and disadvantages of AI applications in cybersecurity, respectively. Opportunities and threats are the elements in the external environment, which can be beneficial or harmful to AI-based cybersecurity applications (Emet Gürel and Merba, 2017).

5.1 Strengths

Improvements in speed, scope, and consistency of analysis

As the cybersecurity threat landscape becomes increasingly complex, and security teams monitor data from a wider array of sources, automation will become a necessity. In the future, the value of AI will be crucial for security teams to monitor the burgeoning quantities of data generated by their organisations. AI algorithms will enable cybersecurity teams to decrease the time to discover security incidents and the time needed to resolve security concerns. The use of AI to process large data sources can also enable security teams to detect threats and incidents that would have otherwise remain unnoticed. Finally, the use of automated anomaly-detection techniques will provide organisations with the ability to apply consistent methods of threat detection that might be immune to human lapses.

Currently, machine-learning algorithms are used to automate narrowly defined cybersecurity tasks. Although AI is currently not a universal fix for cybersecurity problems (and likely never will be), it will be used by security teams to solve well-scoped problems that require pattern recognition. Although many ML algorithms

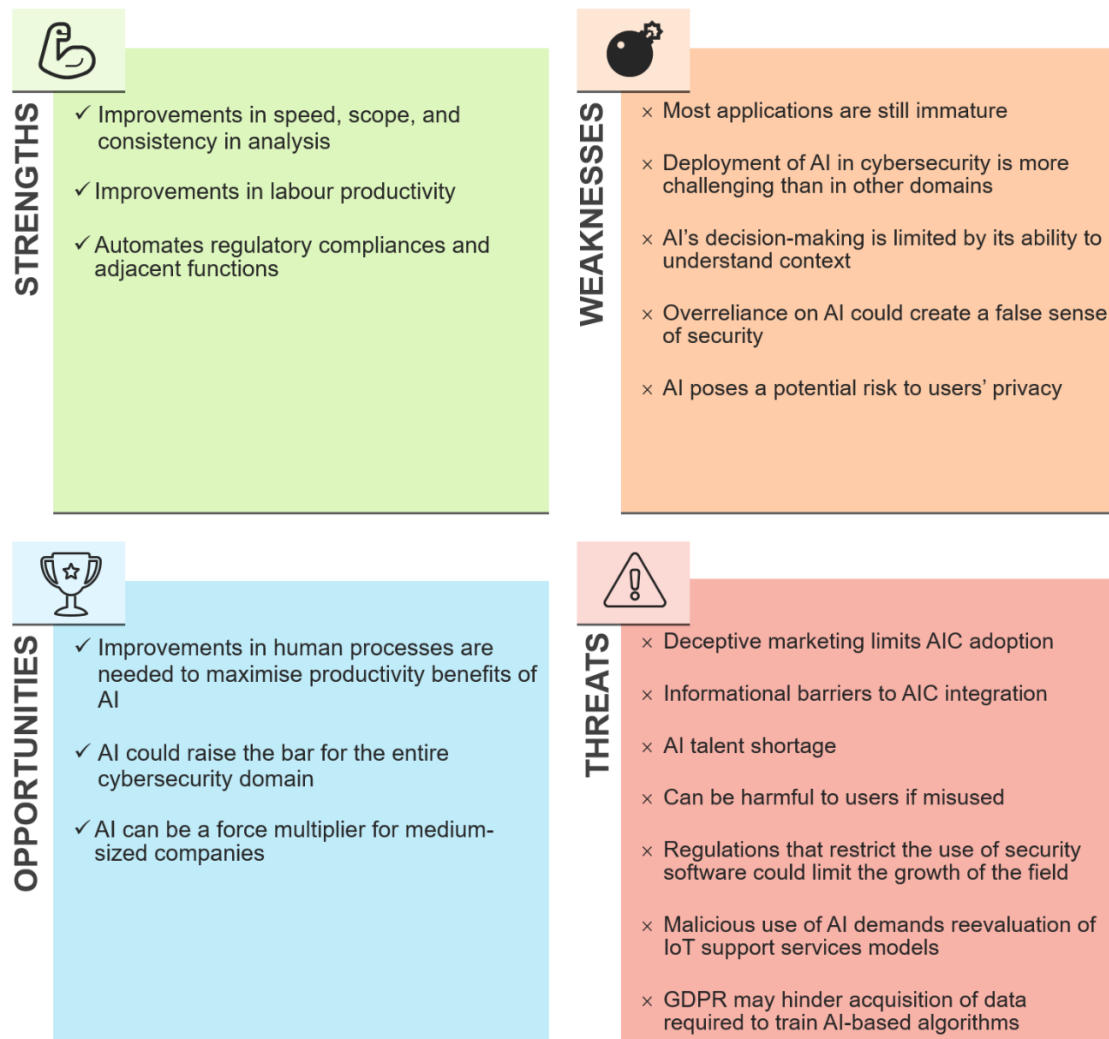


Figure 5.1: SWOT Analysis of AI in Cybersecurity

do not give conclusive (“yes” or “no”) answers, they output probabilities of events occurring. This can enhance decision-making by cybersecurity analysts.

Improvements in labour productivity

The use of AI could help fill the aforementioned cybersecurity workforce gap. AI can compensate for the shortage of personnel, skills, and resources. Currently, most organisations over-utilise human resources on tasks that require little human problem-solving abilities (Deloitte, 2018). AI applications can allow organisations to complete security tasks much faster than with human labour. AI applications can also free up labour from monotonous and repetitive jobs. Hence, the limited skills and talents can be reallocated to tasks that require critical thinking and strategising.

Automating regulatory compliance and adjacent functions

Security teams can use AI to automate repetitive tasks related to regulatory compliance and security audits. This reduces the need for human intervention and manual operations, and decreases the difficulty and complexity of complying with regulatory requirements. Examples of some possible applications include: (1) Synthesising regulation and policy; (2) Conducting regular and termly assessments; (3) Assigning and allocating cybersecurity responsibilities; (4) Conducting penetration tests.

5.2 Weaknesses

Nascent technology

Many applications of AI in cybersecurity are still immature. Although there is much discussion around the use of machine learning for cybersecurity functions, a majority of applications are still in their infancy. Machine learning tools are currently employed and valued by many organisations. However, the deployed uses of ML represent only a small subset of specific cybersecurity tasks which require pattern recognition. Deployed algorithms have yet to reach the level of autonomy often sensationalised by vendors and the media.

Unique technical limitations to AI in cybersecurity

The deployment of AI algorithms in cybersecurity is more challenging than in other domains. False-positive rates often need to be extremely low for AI applications to be useful. When ML algorithms are used in many applications of cybersecurity, they incur a cost for false alarms and missed detections. For example, each alert for anomalies in an intrusion detection system would require a human analyst to investigate the causal event. The investigations of false alarms represent a loss of productivity because they displace resources from other critical security activities. Even if a company decides it can tolerate high false positive rates, analysts are likely to suffer from alert fatigue. That is, they may be lulled into ignoring alerts after investigating too many false alarms (Kohgadai, 2017).

Besides the high cost of alarms, research by Sommer and Paxson (2010) identified additional characteristics that make ML difficult to apply in intrusion detection, and which likely apply to other domains of cybersecurity. They cite a dearth of training data, the difficulty of turning anomaly detections into actionable results, significant variabilities in input data, and a discrepancy in standards for evaluating ML success between academic and operational settings. These limitations need to be addressed in order to advance the quality of ML-based intrusion detection systems and other AI-driven cybersecurity tools significantly.

AI's decision-making is limited by context

AI will not automate all activities currently done by cybersecurity analysts. In the foreseeable future, AI-based tools will likely be limited by their inability to contextualise information. ML algorithms are efficient in identifying patterns and anomalies in data. However, context is often required to separate benign anomalous activity from malicious anomalous activity. As a result, machines will likely have trouble identifying which anomalous activities are, in fact, attacks. For example, an ML-based intrusion detection system may identify a series of port scans as abnormal. However, it will be unable to determine whether these scans are an instance of real malicious activity or merely a penetration test (a task that could be easily accomplished by humans). This is not to say that the ML-based systems are not useful, but simply that their output should be reviewed by human analysts, when feasible before any critical actions are taken.

ML tools will also have limited utility in tracking and defending against Advanced Persistent Threats (APTs). APTs are well-resourced and sophisticated cyber-attacks

that are often conducted over long periods (Chen, Desmet and Huygens, 2014). APTs notoriously require an abundance of context and expertise to detect and track, even by human analysts. Moreover, adversaries in advanced threat campaigns often respond to their targets' defensive measures by altering their technique, tactics, and procedures over time. Identifying them is a task ill-fitted for pattern or anomaly detectors alone. Nevertheless, ML-based tools can still be significant facilitators for human analysts in the process of tracking APT groups.

False sense of security

An overreliance on AI could create a false sense of security. As AI-based tools mature, some security teams may be tempted to rely excessively on these solutions. However, firms that rely excessively on a single AI solution may be vulnerable to malicious actors that can evade detection by those tools, especially as adversarial learning techniques are developed to help hackers exploit trained algorithms. This is especially true for black-box AI algorithms that provide little feedback on the recommendations made to security teams. Instead, human analysts should “trust but verify” decisions made by ML tools (Ettinger, 2019) and security teams should use AI-based tools thoughtfully as part of a defence in depth strategy (US-CERT, 2005).

Furthermore, anomaly detection systems that function by baselining a system or network and identifying deviations are only valid under the assumption that the system has not been compromised. This is a significant assumption given that more than a tenth of breaches in 2016 took over a year to discover (Collins, 2017). A system that is deployed to detect anomalies in an already compromised environment would hardly be useful as it treats signs of compromise as normalcy. Consequently, it would give security teams a false sense of safety in a grim situation.

Privacy risk

AI poses a risk to user privacy. This is a concern that limits the data that can be used to train ML algorithms. Much like in other domains, AI algorithms in cybersecurity may pose a privacy risk to individuals whose data is used to train the algorithm. This is especially true as the large quantities of data necessary to improve AI algorithms would exacerbate the damage from data breaches (Ettinger, 2019). Likewise, it has been shown that decisions made by an algorithm can be used to make inferences about the information used to train it (ibid.).

Recently, researchers have applied the use of techniques like federated learning and differential privacy (DP) to protect user's private data. By using federated learning, ML algorithms can be trained at the source of the data while the results of the learning are sent to a central server. This reduces the need to expose users' private data. In differential privacy techniques, noise is added to the training data so that individuals cannot be identified (Yang et al., 2019). By exploring the use of these techniques, companies may be able to train their AI algorithms while minimising the risk to users' data. Just as importantly, however, firms should consider the appropriate contexts to use AI algorithms in order to preserve user privacy.

5.3 Opportunities

Need for human process improvements

Improvements in human processes are needed to maximise the productivity benefits of AI. Several interviewees expressed the need for organisations to improve their human processes in order for AI to maximise their productivity. This draws from the wisdom of “Kasparov’s law” which emphasises the importance of processes in allowing humans to take advantage of automation (Thornhill, 2017). AI can guide human professionals by providing decision support and by optimising their efforts. However, the benefits derived depend on the context in which AI is used. By improving their processes, security organisations can allow the efforts of cyber professionals to be maximised by AI tools.

Universal improvements in cybersecurity

AI could raise the bar for the entire cybersecurity domain. The nature of cybersecurity may be changed significantly if AI applications are rolled out democratically. That is, if they are more accessible to security teams regardless of budget or expertise. In the case that the implementations of AI0base cybersecurity tools are released as open-source projects, sold commercially at low cost, and designed to be easy for firms to adopt regardless of size, the security posture of all firms would rise significantly and require adversaries to deploy advanced tactics to achieve results.

Force multiplier for medium-sized companies

AI-based tools can allow security teams to accomplish more with fewer analysts. Yet for large companies, savings in labour costs may be inconsequential compared to their overall cybersecurity budgets. In contrast, for medium-sized companies with small security teams, freed-up labour can be used in ways that contribute significantly to companies' growth. Still, small security teams can only benefit from AI-based applications if their implementations are simple and cost-effective.

5.4 Threats

Deceptive marketing limits commercial adoption

Exaggerated claims from vendors can discourage firms from adopting AI technologies. Many firms struggle to gauge the quality of commercial cybersecurity tools that purport to use AI. Since there are no standards in the market for what constitutes AI, vendors are left unchecked to make misleading claims about the AI competency of their products. As a result, consumers are often at risk of purchasing products that are not as “intelligent” as advertised, or products that are altogether ineffective.

Teams that opt to purchase AI solutions commercially must identify which vendors and products are reputable and appropriate for their organisations. This can be a significant challenge given the abundance of products on the market, especially for small security teams. Given the noise created by vendors, it takes significant time and expertise for firms to substantiate vendor claims. As a result, security teams must identify reputable cybersecurity vendors or invest their resources in unproven solutions.

Firms are limited by a lack of information

Organisations may find it difficult to adopt AI due to difficulties in justifying spending on any AI-based tools, due to the lack of information on the effectiveness of AI-based cybersecurity products. Depending on the size of the organisation, it might be more economically sensible for firms to invest in proven alternative solutions to improve their security posture. In other cases, security teams are severely underfunded, and may not find it efficient to spend their limited budget on unproven solutions. Additionally, firms with older legacy systems which do not work well with modern

technologies such as AI, might not see the methods and merits which AI-based tools can be adopted.

Shortage in AI talent

Security teams can incorporate AI algorithms into their toolset by either developing them in-house or by purchasing them from security vendors (or through a combination of both). For teams that choose to build their tools manually, the primary barrier lies in obtaining the necessary human capital. Building deployable AI tools in-house requires highly talented data scientists who also have cybersecurity experience. These AI talents are also crucial in validating the effectiveness of AI-enabled commercial products. This is a rare skill-set that often commands high salaries on the employment market.

AI potential for harm

Like any technology, the application of AI in cybersecurity can be harmful to individuals if misused. Security teams should think carefully about the situations in which they deploy AI-based systems, and the power they assign to those systems. In general, AI-based tools are harmless if they are simply used for supporting analysis. However, if they are used to make automated decisions, their risk of harm is dependent on the context in which they are used. Suppose an AI-based intrusion detection system can automatically block user accounts on an internal corporate network due to on anomalous activity. If the system makes a wrong decision, the potential for harm ranges from a user being inconvenienced to a critical business meeting being disturbed. If a hospital network uses this IDS, instead, there could be fatal outcomes if a life support system is automatically disconnected. Ultimately, the potential harm of these applications is not intrinsically dependent on AI itself, but instead on the ability of practitioners to apply these technologies appropriately. This is especially crucial given AI's propensity to make discriminatory decisions when trained with biased data (AI HLEG, 2019).

Restrictive laws may limit benefits

Regulations that restrict the export, import, or use of software used for security would limit the growth of AI for vulnerability scanning. Machine learning may be used to identify vulnerabilities in software. This capability can be used either by

security teams for patching software or by attackers to identify easy targets, allowing them to cause damage at scale. The latter scenario may cause regulators to consider limiting its growth. For example, until recently, the export of “intrusion software” commonly used by security researchers and penetration testers were prohibited by the Wassenaar Arrangement for Export Controls for Arms and Dual-Use Goods and Technologies (Hinck, 2018). However, specific exemptions were recently included for these tools based on the recommendations of security experts (Cross, 2018). Similar prohibitions on the sale or distributions of AI-based vulnerability testing tools may limit the growth of these technologies or prevent organisations from fully benefiting from them.

AI can be used by attackers to exploit IoT devices at scale

The malicious use of AI for vulnerability scanning demands a reevaluation of IoT support services models. The number of IoT devices in use is estimated to be 7 billion today and is expected to grow to 22 billion by 2025 (IoT Analytics, 2018). However, these devices are notoriously insecure and difficult to patch (Bertino et al., 2016). Consequently, in 2016, the Mirai Botnet exploited millions of vulnerable IoT devices (e.g. IP cameras and routers) exposed to the internet to conduct distributed denial of service attacks (Krebs, 2016). Although the Mirai botnet relied on these devices having default credentials, it is conceivable that AI techniques could be used to scan devices for less trivial vulnerabilities in the future, especially if vendors discontinue support services for IoT devices that are still in use.

Currently, the EC is funding a series of projects that would help secure IoT devices in this uncertain future (Loupos, 2018). The EC’s proposed DSM certification framework for ICT products would also improve the security of IoT products (EC, 2017a). In order to prevent similar incidents in the future, it may be helpful for regulators to examine requirements for support services for IoT devices. For instance, regulators could set rules to encourage vendors and consumers to participate in third-party support aftermarkets to ensure that devices can remain patched after they are discontinued by vendors. Similar ideas been floated by the EC, but have not yet been enacted (EC, 2017c).

GDPR may hinder data acquisition

GDPR may hinder the acquisition of data necessary for vendors to train AI algorithms. According to product vendors and developers, data acquisition can be challenging

under the current EU data protection regulation. For instance, to cope with the requirement of opt-in consent in GDPR, which asks organisations to obtain an allowance for collection and processing of personal data from individuals, companies would need to have the capability to explain the logic behind the data used. However, for most businesses, this is difficult to implement based on their current capability. In this sense, stringent data protection regulation set barriers for data collection, which could further impede the datasets required training of AI algorithms.

6. Policy Recommendations

Numerous policies have been implemented by the EC to improve cybersecurity practices, such as the Cybersecurity Act (EC, 2018c). Separately, policies have targeted the development of AI in general, which we believe will have spillover effects on AI applications in cybersecurity. Likewise, policies have been put in place to tackle data-availability concerns (EC, 2018d). Hence, our recommendations avoid issues that will likely be addressed by existing or proposed policies and initiatives, whose effect may have yet to materialise (Table 2).

We analysed the process through which cybersecurity teams adopt AI-based cybersecurity solutions, and found two groups of barriers for AI uptake, namely technical and information. Technical barriers refer to issues or limitations that are inherent to AI in cybersecurity. Information barriers refer to issues that impede firms' adoption due to the lack of knowledge on existing products within the market or on AI applications in cybersecurity.

Based on the issues and barriers that have not been addressed by EU policies and initiatives, we present three policy initiatives that could address the issues, consider their advantages and drawbacks, and indicate a recommended course of action based on our analyses.

6.1 Focus Funding on Use-inspired Basic Research with Possible Benefits to AI in Cybersecurity

Policy context

Since the EU member states jointly agreed to cooperate on AI during the Digital Declaration Day 2018 (EC, 2018e), there has been a constant stream of revenue in support of basic research on AI. The first of such initiatives are the investments in

#	Weakness/Threats Identified	Barrier	Currently addressed?
1	<ul style="list-style-type: none"> — Regulations restriction for AI in cybersecurity — Barriers to data necessary 	Technical	Yes, under EU's initiative for AI in 'Digital Single Market' - 'Building a
2	<ul style="list-style-type: none"> — Lack of AI talent in cybersecurity 	Technical	Yes - 'Enablers for Skills and Education' group
3	<ul style="list-style-type: none"> — An overreliance on AI could create a false sense of security 	Information	No. (Policy recommendation 2)
4	<ul style="list-style-type: none"> — Exaggerated claims from vendors can disincentivise firms from adopting AI technologies 	Information + Information	No. (Policy recommendation 3)
5	<ul style="list-style-type: none"> — Many applications of AI in cybersecurity are still immature — The deployment of AI in cybersecurity is more challenging than other domains — AI's decision-making is limited by its ability to understand context 	Technical	No. (Policy recommendation 1)
6	<ul style="list-style-type: none"> — AI poses a potential risk to user privacy — The application of AI in cybersecurity can be harmful if misused — The malicious use of AI for vulnerability scanning demands a reevaluation of IoT support services model 	Technical	Yes, under EU's initiative for AI in 'Digital Single Market' with study of GDPR implications

Table 6.1: Identification of Adoption Barriers from SWOT

AI by the EC under Horizon 2020. Under this programme, the EC will increase its annual investments in AI to 1.5 billion euros from 2018 - 2020 (EC, 2018a). This investment will be three-pronged, as it will focus on: 1) connecting and strengthening AI research excellence centres across Europe; 2) bolstering the development of an "AI-on-demand platform" such as the AI4EU project to facilitate collective work in AI research; 3) enabling the usage of AI in key sectors (ibid.). Furthermore, the EU will continue to invest at least 7 billion euros from 2021 to 2027 in innovation and research programmes such as Horizon Europe, which seeks to maintain Europe's position as being one of the global leaders in research (EC, 2018f). Yet, while there exists a plethora of funding for general AI basic research, no specific focus is placed on AI and its applications in cybersecurity (EC, 2014).

Policy specifics

Since the use AI is still relatively unexplored in cybersecurity, the areas in which basic research fundings are directed to support this field can be missed by other areas of general AI research. In particular, the funding should be directed towards use-inspired basic research, which focuses on gaining scientific knowledge about AI, with an orientation towards developing means to exploit the findings for cybersecurity applications (Stokes, 1997).

For example, as seen from section 4.2, to limit the privacy risks to users, basic research can be directed into the use of novel distributed ML techniques, such as federated learning. Additionally, since AI applications need to be very precise in cybersecurity, funding should be directed at basic research to improve the resistance of AI against adversarial learning attacks, as underscored by a recent research paper from the MIT-IBM Watson AI Lab (Weng et al., 2018). At the same time, AI's uses in other domains, including healthcare, would benefit from improvements in privacy and resilience to attacks. The emphasis on use-inspired basic research would drive the maturity of AI and cybersecurity applications as seen from section 4.3 while also benefiting the application of AI in other domains.

Policy feasibility

One of the most convenient policy window to implement this policy would be the upcoming Horizon Europe programme from 2021 - 2027. The importance of AI in cybersecurity can be observed in the EU strategic foresight study BOHEMIA, in support of the Commission's proposal for Horizon Europe (EC, 2018b), where the

importance of understanding ICT security has been emphasised. Additionally, AI is one a key objective for the programme given that it is one of the programme's main lines of research activities. Horizon Europe is, therefore, a possible platform to achieve this funding.

6.2 Create a Centralised Online Platform and Implementation Guides for ML-based Cybersecurity Applications

Policy context

From the interviews, we found that there are clear-cut applications of ML that can offer benefits to firms in the short run (also indicated in the upper right quadrant of the matrix from Figure 2). We contend that many firms would benefit from implementing these applications in-house if they all have access to the necessary information. Thus, existing resources on simple implementations of ML in cybersecurity should be pooled onto one website or digital platform. A centralised knowledge base would decrease the knowledge barriers restricting firms from the basic uses of ML. This digital platform can be created by ENISA or by the recently proposed European Cybersecurity Industrial, Technology and Research Competence Centre (EC, 2018g).

Policy specifics

The proposed organisation would be responsible for the creation of an electronic platform to the ease adoption of ML-based cybersecurity applications. The platform should contain links to existing materials, such as open-source projects and relevant literature. In addition, the proposed organisation should develop custom guidelines for incorporating rudimentary ML algorithms into a cybersecurity workflow. These guidelines, also accessible on the platform, should detail why, when, and how companies can incorporate AI-based tools into their cybersecurity processes. Furthermore, the guides should include code repositories, step-by-step walkthroughs, and other material necessary to ease the implementation process. They should be iteratively modified until they are easy to follow by organisations with low ML expertise. As an end-goal, they should enable security teams to incorporate basic ML algorithms into their workflow.

Moreover, the guides should focus on high impact and high maturity applications of ML in cybersecurity (as identified in Figure 2). Annual reassessments of the impact and maturity of each of these applications should also be carried out by the proposed organisation through the framework presented in Figure 2, and should be published annually. This is to ensure that knowledge about the applications and their suitability can be disseminated within the industry to allow companies to make more informed decisions regarding the adoption of AI-based cybersecurity tools.

Policy feasibility

Currently, ENISA acts as a body of expertise for cybersecurity-related issues within the EU (ENISA, 2019b). As one of its three core activities, ENISA collaborates with member states and private sector organisations to deliver ‘hands-on’ advice and solutions (ENISA, 2019a). The development of a knowledge base would fit within the bounds of its current responsibilities. As opposed to ENISA, the Cybersecurity Competence Centre could better leverage the resources and expertise available in member states’ National Coordination Centres and in the Cybersecurity Competence Community to develop a more comprehensive knowledge base. This would contribute to its intended goal of “contribut[ing] to a high level of network and information security (NIS) within the Union” (ENISA, 2019a, p. 1).

If the proposed organisation is ENISA, the need to develop a platform could overextend its resources. Many crucial cybersecurity-related issues require ENISA’s attention, such as the protection of critical information infrastructure, and the need to increase the overall cybersecurity preparedness amongst EU member states (ECA, 2019; ENISA, 2019c). Such issues may be more critical than the adoption of AI, which is seen as an add-on benefit for cybersecurity rather than a universal solution for all cybersecurity-related problems. The need to create AI-specific materials may put a strain on resources required for more pressing problems. However, there has been an increased focus on strengthening cybersecurity within the EU, and efforts have been made to reinforce ENISA’s mandate through the Cybersecurity act (EC, 2018c). This can also further cement ENISA’s role as a body of cybersecurity expertise, and strengthen cybersecurity within the EU through the possible increase in adoption of AI-based tools, thus reinforcing the Cybersecurity Act.

6.3 Address Information Asymmetries of the AI Cybersecurity Product Market Through a Labelling Scheme

Policy context

Currently, cybersecurity managers choose products based on word-of-mouth, consumer reviews on software rating websites, and through the reviews of independent research companies like Gartner. In the last two channels, however, reviews are often indicative of overall performance, as opposed to AI-competency. Many consumers, especially small and medium-sized businesses, lack the expertise to judge a product's use of AI. Moreover, vendors are likely secretive of their algorithms, preventing any such scrutiny from occurring. Thus reviews are based only on the overall quality of the product as opposed to its use of AI. Though this process is generally useful, it is not beneficial to AI uptake. Since many products on the market claim to apply AI irrespective of actually the actual use of AI, these claims may lose value, causing consumers to ignore them altogether.

This is a classic asymmetric market failure in which both buyers and sellers lose out because of incomplete information on the buyer side (Barbaroux, 2014). It can be solved through “signalling,” in which the seller acquires a credible credential to indicate higher product quality. In the digital age, online reviews have emerged as an alternate solution to information asymmetry (Ögüt and Taş, 2012). Should the EC wish to take an aggressive approach to AI uptake, an AI-specific labelling scheme for cybersecurity products would encourage AI uptake and enable firms to select better products.

Policy specifics

The EC could develop an AI-specific labelling scheme as part of the proposed certification framework for ICT cybersecurity products (2017). A voluntary labelling scheme would standardise AI terminology in the cybersecurity domain and inform consumers of the extent to which products substantially employ AI algorithms. This additional information could help cybersecurity managers make better-informed decisions when selecting products and avoid falsely advertised products. It would also reward vendors for their contributions to AI-based cybersecurity product innovation.

Policy feasibility

A labelling system would create a channel for vendors to showcase the legitimate uses of AI as a competitive advantage. Moreover, it may incentivise other firms to invest in AI research and allow innovative AI-based startups to differentiate themselves in the market. Labels would also help consumers build trust in the use of AI within the cybersecurity product market, increasing their propensity to invest time and resources in emerging products.

However, “AI competency” labels also may create undesired incentives in the market unless they are indicated alongside products’ overall performance. After all, the unique aspect of this problem is that consumers should not choose products solely based on the quality of AI, but based on performance in general. That is, a cybersecurity manager should probably not aim to have the best AI products, but rather seek to improve rates of detection or increase productivity. Moreover, it is possible for a cybersecurity product to be labelled inaccurately as AI, yet have better rates of detection than a product that truly uses AI. Providing “AI competency” labels would drive AI uptake but perhaps at the cost of promoting inferior products.

Depending on the difficulty of acquisition, an AI labelling scheme may act as a barrier against new entrants to the market. According to Cihon et al. (2018), firms with low revenue are less likely to seek labels due to monetary costs and required time investment. Therefore, though it may increase the uptake of incumbent firms’ AI products, and AI labelling scheme may discourage growth and innovation from startups, which are critical contributors to this new market.

Furthermore, questions may be raised about implications for vendor liability when a breach occurs involving a labelled AI cybersecurity product. It should be made clear that the use of AI does not inherently contribute to a product’s overall quality and, therefore, should not reduce liability in disputes. In other words, the label does not certify the quality of the product but rather informs consumers of the product’s use of AI.

An AI-specific labelling scheme for cybersecurity products would provide critical information to cybersecurity teams that would raise interest and adoption of AI-based cybersecurity tools. It would also encourage vendors to increase research investments and pursue real innovation as opposed to false advertisement. Admittedly, this a liberal recommendation that should only be adopted to increase AI uptake, and would only be beneficial as long as labelling costs are low, and the labels do not overshadow comprehensive product reviews. In any other case, the EC should maintain the

status quo.

7. References

- Agerholm, Harriet (2016). *The terrifying reason the internet was brought to its knees on Friday*. The Independent. URL: <http://www.independent.co.uk/life-style/gadgets-and-tech/news/cyber-attack-hack-twitter-paypal-netflix-major-internet-outages-caused-by-everyday-devices-latest-a7374971.html> (visited on 05/20/2019).
- Analytics, IoT (2018). *State of the IoT 2018: Number of IoT devices now at 7B – Market accelerating*. URL: <https://iot-analytics.com/state-of-the-iot-update-q1-q2-2018-number-of-iot-devices-now-7b/> (visited on 05/18/2019).
- Apruzzese, Giovanni et al. (2018). “On the effectiveness of machine and deep learning for cyber security”. In: *2018 10th International Conference on Cyber Conflict (CyCon)*. 2018 10th International Conference on Cyber Conflict (CyCon). Tallinn: IEEE, pp. 371–390. ISBN: 978-9949-9904-2-9 978-9949-9904-3-6. DOI: [10.23919/CYCON.2018.8405026](https://doi.org/10.23919/CYCON.2018.8405026). URL: <https://ieeexplore.ieee.org/document/8405026/> (visited on 04/27/2019).
- Baker, Pam (2018). *The Attack Surface Is Growing and Changing: What That Means for Channel Partners*. URL: <https://www.channelpartneronline.com/files/2018/10/S191018.pdf> (visited on 05/19/2019).
- Barbaroux, Pierre (2014). “From market failures to market opportunities: Managing information under asymmetric information”. In: *Journal of Innovation and Entrepreneurship* 3. DOI: [10.1186/2192-5372-3-5](https://doi.org/10.1186/2192-5372-3-5).
- Barrinha, André and Helena Farrand-Carrapico (2018). *How coherent is EU cybersecurity policy?* EUROPP. URL: <https://blogs.lse.ac.uk/euoppblog/2018/01/16/how-coherent-is-eu-cybersecurity-policy/> (visited on 05/30/2019).
- Bertino, Elisa et al. (2016). “Internet of things (iot): Smart and secure service delivery”. In: *ACM Transactions on Internet Technology (TOIT)* 16.4, p. 22.
- Blackledge, Jonathan, Sergei Bezobrazov, and Paul Tobin (2015). “Cryptography using artificial intelligence”. In: pp. 1–6. DOI: [10.1109/IJCNN.2015.7280536](https://doi.org/10.1109/IJCNN.2015.7280536).
- Brook, Chris (2018). *What is User and Entity Behavior Analytics? A Definition of UEBA, Benefits, How It Works, and More*. Digital Guardian. URL:

- <https://digitalguardian.com/blog/what-user-and-entity-behavior-analytics-definition-ueba-benefits-how-it-works-and-more> (visited on 04/29/2019).
- Brown, Tom B. et al. (2017). “Adversarial Patch”. In: *arXiv:1712.09665 [cs]*. arXiv: [1712.09665](https://arxiv.org/abs/1712.09665). URL: <http://arxiv.org/abs/1712.09665> (visited on 04/30/2019).
- Brundage, Miles et al. (2018). “The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation”. In: *arXiv:1802.07228 [cs]*. arXiv: [1802.07228](https://arxiv.org/abs/1802.07228). URL: <http://arxiv.org/abs/1802.07228> (visited on 05/21/2019).
- Burt, Tom (2019). *New steps to protect Europe from continued cyber threats*. Microsoft EU Policy Blog. URL: <https://blogs.microsoft.com/eupolicy/2019/02/20/accountguard-expands-to-europe/> (visited on 05/20/2019).
- CBIR (2016). *Deals To Cybersecurity Are Increasingly Global With Israel In The Lead*. CB Insights Research. URL: <https://www.cbinsights.com/research/cybersecurity-funding-geographic-trends/> (visited on 05/10/2019).
- Cihon, Peter et al. (2018). “Why Certify? Increasing adoption of the proposed EU Cybersecurity Certification Framework”. In: p. 62.
- Cisco (2015). *Mitigating the Cybersecurity Skills Shortage - Top Insights and Actions from Cisco Security Advisory Services*. Cisco.
- (2018). *Cisco 2018 Annual Cybersecurity Report*. Cisco. URL: https://www.cisco.com/c/dam/m/hu_hu/campaigns/security-hub/pdf/acr-2018.pdf (visited on 05/21/2019).
- CrowdStrike (2018). *The Rise of Machine Learning in Cybersecurity*. CrowdStrike. URL: <https://go.crowdstrike.com/rs/281-OBQ-266/images/WhitepaperMachineLearning.pdf>.
- Cylance (2017). *AI Based Malware Prevention and Detection*. URL: https://threatvector.cylance.com/en_us/home/ai-based-malware-prevention-and-detection.html (visited on 04/25/2019).
- Dignan, Larry (2017). *IBM defends Watson, cognitive computing, AI efforts amid analyst questions*. ZDNet. URL: <https://www.zdnet.com/article/ibm-defends-watson-cognitive-computing-ai-efforts-amid-analyst-questions/> (visited on 04/30/2019).
- Directive (EU) 2016/1148 of the European Parliament and of the Council of 6 July 2016 concerning measures for a high common level of security of network and information systems across the Union* (2016). In collab. with European Parliament and Council of the European Union. URL: <http://data.europa.eu/eli/dir/2016/1148/oj/eng> (visited on 05/12/2019).
- EC (2013a). *Cybersecurity*. Digital Single Market - European Commission. URL: <https://ec.europa.eu/digital-single-market/en/cyber-security> (visited on 05/12/2019).

- EC (2013b). *Cybersecurity Strategy of the European Union: An Open, Safe and Secure Cyberspace*. URL: https://eeas.europa.eu/archives/docs/policies/eu-cyber-security/cybsec_comm_en.pdf (visited on 05/18/2019).
- (2014). *EU-funded projects on Digital security*. Digital Single Market - European Commission. URL: <https://ec.europa.eu/digital-single-market/en/programme-and-projects/project-factsheets-digital-security> (visited on 05/19/2019).
 - (2017a). *Cybersecurity factsheet (European Commission)*. URL: <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52017JC0450&from=EN> (visited on 05/17/2019).
 - (2017b). *Cybersecurity- EU agency and certification framework*. URL: [http://www.europarl.europa.eu/RegData/docs_autres_institutions/commission_europeenne/com/2018/0630/COM_COM\(2018\)0630_EN.pdf](http://www.europarl.europa.eu/RegData/docs_autres_institutions/commission_europeenne/com/2018/0630/COM_COM(2018)0630_EN.pdf).
 - (2017c). *Resilience, Deterrence and Defence: Building strong cybersecurity in Europe*. URL: <https://www.consilium.europa.eu/media/21480/cybersecurityfactsheet.pdf> (visited on 05/20/2019).
 - (2017d). *The EU cybersecurity certification framework*. Digital Single Market - European Commission. URL: <https://ec.europa.eu/digital-single-market/en/eu-cybersecurity-certification-framework> (visited on 05/23/2019).
 - (2018a). *Artificial Intelligence*. Digital Single Market - European Commission. URL: <https://ec.europa.eu/digital-single-market/en/artificial-intelligence> (visited on 05/20/2019).
 - (2018b). *Beyond the Horizon: foresight in support of future EU research and innovation policy (BOHEMIA)*. European Commission - European Commission. URL: https://ec.europa.eu/info/research-and-innovation/strategy/support-policy-making/support-eu-research-and-innovation-policy-making/foresight/activities/current/bohemia_en (visited on 05/24/2019).
 - (2018c). *Cybersecurity Act*. European Commission - European Commission. URL: https://ec.europa.eu/commission/news/cybersecurity-act-2018-dec-11_en (visited on 05/23/2019).
 - (2018d). *Data in the EU: Commission steps up efforts to increase availability and boost healthcare data sharing*. URL: http://europa.eu/rapid/press-release_IP-18-3364_en.htm (visited on 05/21/2019).
 - (2018e). *European Commission - PRESS RELEASES - Press release - Digital Day 2018: EU countries to commit to doing more together on the digital front*. URL: http://europa.eu/rapid/press-release_IP-18-2902_en.htm (visited on 05/24/2019).
 - (2018f). *European Commission - PRESS RELEASES - Press release - EU budget: Commission proposes most ambitious Research and Innovation programme yet*.

- URL: http://europa.eu/rapid/press-release_IP-18-4041_en.htm (visited on 05/24/2019).
- EC (2018g). *Proposal for a European Cybersecurity Competence Network and Centre*. Digital Single Market - European Commission. URL: <https://ec.europa.eu/digital-single-market/en/proposal-european-cybersecurity-competence-network-and-centre> (visited on 05/30/2019).
- (2019). *A definition of Artificial Intelligence: Main Capabilities and Disciplines*. Text. European Commission, p. 9. URL: <https://ec.europa.eu/digital-single-market/en/news/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines> (visited on 05/11/2019).
- ECA (2019). *Challenges to effective EU cybersecurity policy*. European Court of Auditors, p. 74.
- Emet Gürel and Tat Merba (2017). “SWOT Analysis: A Theoretical Review”. In: *Journal of International Social Research* 10.51, pp. 994–1006. ISSN: 1307-9581. DOI: [10.17719/jisr.2017.1832](https://doi.org/10.17719/jisr.2017.1832). URL: http://sosyalarastirmalar.com/cilt10/sayi51_pdf/6iksisat_kamu_isletme/gurel_emet.pdf (visited on 05/20/2019).
- ENISA (2019a). *About ENISA*. URL: <https://www.enisa.europa.eu/about-enisa/about-enisa> (visited on 05/20/2019).
- (2019b). *General FAQ’s on ENISA*. URL: <https://www.enisa.europa.eu/faq-on-enisa/general-faqs-on-enisa> (visited on 05/23/2019).
- (2019c). *National Cyber Security Strategies*. URL: <https://www.enisa.europa.eu/topics/national-cyber-security-strategies> (visited on 05/23/2019).
- (2019d). *Towards a framework for policy development in cybersecurity - Security and privacy considerations in autonomous agents*. URL: <https://www.enisa.europa.eu/publications/considerations-in-autonomous-agents> (visited on 05/21/2019).
- Ettinger, Jared (2019). *Cyber Intelligence Tradecraft Report: The State of Cyber Intelligence Practices in the United States (Study Report Only)*. Software Engineering Institute. URL: <https://resources.sei.cmu.edu/library/asset-view.cfm?assetid=546686> (visited on 05/30/2019).
- Eykholt, Kevin et al. (2017). “Robust Physical-World Attacks on Deep Learning Models”. In: *arXiv:1707.08945 [cs]*. arXiv: [1707.08945](https://arxiv.org/abs/1707.08945). URL: <http://arxiv.org/abs/1707.08945> (visited on 04/30/2019).
- Feiman, Joseph (2019). *Council Post: AI And The Cybersecurity Workforce: A Whole New World*. URL: <https://www.forbes.com/sites/forbestechcouncil/2019/04/26/ai-and-the-cybersecurity-workforce-a-whole-new-world/#3e7f59361c6a> (visited on 04/26/2019).
- GOfS (2017). *The Futures Toolkit*. Government Office of Science.

- Golden, Deborah and Ted Johnson (2017). *AI-augmented cybersecurity: How cognitive technologies can address the cyber workforce shortage*. Deloitte Insights. URL: <https://www2.deloitte.com/insights/us/en/industry/public-sector/addressing-cybersecurity-talent-shortage.html> (visited on 04/24/2019).
- GOV.UK (2014). *Guidance on Technology Readiness Levels*. GOV.UK. URL: <https://www.gov.uk/government/news/guidance-on-technology-readiness-levels> (visited on 05/30/2019).
- HLEG, AI (2019). *Ethics Guidelines for Trustworthy Artificial Intelligence*. Text. High Level Expert Group on Artificial Intelligence. URL: <https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines> (visited on 05/21/2019).
- Howarth, Fran (2018). *User and Entity Behavior Analytics*. Bloor Research. URL: <https://www.bloorresearch.com/technology/user-and-entity-behavior-analytics/> (visited on 04/29/2019).
- Infosec (n.d.). *Phishing As An Attack Vector*. Infosec Resources. URL: <https://resources.infosecinstitute.com/category/enterprise/phishing/phishing-as-an-attack-vector/> (visited on 05/10/2019).
- (2018). *Top 10 Endpoint Protection Software Solutions*. Infosec Resources. URL: <https://resources.infosecinstitute.com/top-10-endpoint-protection-software-solutions/> (visited on 05/10/2019).
- Institute, Ponemon (2015). *The Cost of Malware Containment*. Ponemon Institute. URL: <https://www.ponemon.org/local/upload/file/Damballa%20Malware%20Containment%20FINAL%203.pdf>.
- Ismail, Nick (2019). *AI in cybersecurity: is this a new tool at the hackers' disposal?* Raconteur. URL: <https://www.raconteur.net/technology/ai-cybersecurity> (visited on 04/30/2019).
- Jowitt, Tom (2018a). *Artificial Intelligence Poses Risk From Malicious Actors | Silicon UK Tech News*. Silicon UK. URL: <https://www.silicon.co.uk/e-innovation/artificial-intelligence/artificial-intelligence-risk-malicious-228559?description=Bolster%20your%20AI%20defences.%20Experts%20warn%20of%20serious%20risks%20within%20a%20five%20year%20period%20due%20to%20recent%20advances%20in%20AI> (visited on 04/30/2019).
- (2018b). *IBM DeepLocker Turns AI Into Hacking Weapon | Silicon UK Tech News*. Silicon UK. URL: <https://www.silicon.co.uk/e-innovation/artificial-intelligence/ibm-deeplocker-ai-hacking-weapon-235783?description=Artificial%20intelligence%20can%20power%20a%20new%20generation%20of%20malware%20that%20can%20bypass%20top%20tier%20cyber%20defences> (visited on 04/30/2019).
- Kobie, Nicole (2018). “To cripple AI, hackers are turning data against itself”. In: *Wired UK*. ISSN: 1357-0978. URL: <https://www.wired.co.uk/article/>

- [artificial-intelligence-hacking-machine-learning-adversarial](#) (visited on 04/30/2019).
- Koslowski, Thomas and Martin Felle (n.d.). *Cognitive computing for cyber security*. Deloitte Switzerland. URL: <https://www2.deloitte.com/ch/en/pages/risk/articles/cognitive-computing-for-cyber-security.html> (visited on 04/30/2019).
- Kostas, Kahraman (2018). “Anomaly Detection in Networks Using Machine Learning”. PhD thesis.
- Krebs, Brian (2016). *Who Makes the IoT Things Under Attack? — Krebs on Security*. URL: <https://krebsonsecurity.com/2016/10/who-makes-the-iot-things-under-attack/> (visited on 05/18/2019).
- Lewis, James (2018). *Economic Impact of Cybercrime Report | McAfee*. McAfee. URL: https://www.mcafee.com/enterprise/en-us/solutions/lp/economics-cybercrime.html?cid=SA&eid=18TL_ECGLQ1_ML_WW&elqCampaignId=23174&tag=ec (visited on 04/30/2019).
- Loucks, Jeff, David Schatsky, and Tom Davenport (2018). *State of AI in the Enterprise, 2nd Edition*. Deloitte. URL: <https://www2.deloitte.com/insights/us/en/focus/cognitive-technologies/state-of-ai-and-intelligent-automation-in-business-survey.html> (visited on 04/25/2019).
- Loupos, Konstantinos (2018). *Fighting for cybersecurity: eight new EU funded projects for a more secure IoT*. Digital Single Market - European Commission. URL: <https://ec.europa.eu/digital-single-market/en/blogposts/fighting-cybersecurity-eight-new-eu-funded-projects-more-secure-iot> (visited on 05/24/2019).
- Mackinnon, Marc (2018). *Smart cyber: How AI can help manage cyber risk*. Deloitte, p. 16.
- Marr, Bernard (2018). *What Is Deep Learning AI? A Simple Guide With 8 Practical Examples*. Forbes. URL: <https://www.forbes.com/sites/bernardmarr/2018/10/01/what-is-deep-learning-ai-a-simple-guide-with-8-practical-examples/> (visited on 04/25/2019).
- Menn, Joseph (2018). “New genre of artificial intelligence programs take computer hacking...” In: *Reuters*. URL: <https://uk.reuters.com/article/us-cyber-conference-ai-idUKKBN1KT120> (visited on 04/30/2019).
- Mittu, Ranjeev and William Frere Lawless (2015). “Human factors in cybersecurity and the role for ai”. In: *2015 AAAI Spring Symposium Series*.
- Morestin, F (2012). “A Framework for Analyzing Public Policies: Practical Guide”. In:

- Newman, Lily Hay (2018). “AI Can Help Cybersecurity—If It Can Fight Through the Hype”. In: *Wired*. ISSN: 1059-1028. URL: <https://www.wired.com/story/ai-machine-learning-cybersecurity/> (visited on 04/27/2019).
- Pham, Thu (2018). *Security Report Finds Phishing, Not Zero-Days, Is the Top Malware Infection Vector*. Duo Security. URL: <https://duo.com/blog/security-report-finds-phishing-not-zero-days-is-the-top-malware-infection-vector> (visited on 05/10/2019).
- Polyakov, Alexander (2018). *How AI-Driven Systems Can Be Hacked*. Forbes. URL: <https://www.forbes.com/sites/forbestechcouncil/2018/02/20/how-ai-driven-systems-can-be-hacked/> (visited on 04/30/2019).
- Rozenblum, Danny (2001). “Understanding Intrusion Detection Systems”. In: p. 9.
- Sadowski, Gorka et al. (2018). “Market Guide for User and Entity Behavior Analytics”. In: p. 22.
- Schneier, Bruce (2018). *Machine Learning Will Transform How We Detect Software Vulnerabilities*. Security Intelligence. URL: <https://securityintelligence.com/machine-learning-will-transform-how-we-detect-software-vulnerabilities/> (visited on 04/30/2019).
- Shravan (2018). *A guide to Machine Learning for Application Security*. Templarbit Inc. URL: <https://www.templarbit.com/blog/2018/08/13/a-guide-to-machine-learning-for-application-security/> (visited on 05/10/2019).
- Sommer, R. and V. Paxson (2010). “Outside the Closed World: On Using Machine Learning for Network Intrusion Detection”. In: *2010 IEEE Symposium on Security and Privacy*. 2010 IEEE Symposium on Security and Privacy, pp. 305–316. DOI: [10.1109/SP.2010.25](https://doi.org/10.1109/SP.2010.25).
- Stoecklin, Marc (2018). *DeepLocker: How AI Can Power a Stealthy New Breed of Malware*. Security Intelligence. URL: <https://securityintelligence.com/deeplocker-how-ai-can-power-a-stealthy-new-breed-of-malware/> (visited on 05/21/2019).
- Stokes, Donald E. (1997). *Pasteur’s Quadrant: Basic Science and Technological Innovation*. Google-Books-ID: 9tnCDAEACAAJ. Brookings Institution Press. 180 pp. ISBN: 978-0-8157-8177-6.
- Thornhill, John (2017). *Man versus machine: when Garry Kasparov met his match*. Financial Times. URL: <https://www.ft.com/content/19a2c21a-33e6-11e7-99bd-13beb0903fa3> (visited on 05/21/2019).
- Tsai, Chih-Fong et al. (2009). “Intrusion detection by machine learning: A review”. In: *Expert Systems with Applications* 36.10, pp. 11994–12000. ISSN: 0957-4174. DOI: [10.1016/j.eswa.2009.05.029](https://doi.org/10.1016/j.eswa.2009.05.029). URL: <http://www.sciencedirect.com/science/article/pii/S0957417409004801> (visited on 04/26/2019).

- Vadapalli, SR, G Hsieh, and KS Nauer (2018). *TwitterOSINT: Automated Cybersecurity Threat Intelligence Collection and Analysis using Twitter Data*. United States of America: CSREA Press. 310 pp. ISBN: 978-1-60132-488-7.
- Meulen, Rob van der (2017). *Gartner Says 8.4 Billion Connected "Things" Will Be in Use in 2017, Up 31 Percent From 2016*. Gartner. URL: <https://www.gartner.com/en/newsroom/press-releases/2017-02-07-gartner-says-8-billion-connected-things-will-be-in-use-in-2017-up-31-percent-from-2016> (visited on 05/10/2019).
- Vincent, James (2019). *Gmail is now blocking 100 million extra spam messages every day with AI*. The Verge. URL: <https://www.theverge.com/2019/2/6/18213453/gmail-tensorflow-machine-learning-spam-100-million> (visited on 05/09/2019).
- Webroot (2017). *Game Changers: AI and Machine Learning in Cybersecurity; a U.S./Japan Comparison*. URL: https://www-cdn.webroot.com/6015/4999/4566/Webroot_AI_ML_Survey_US-2019.pdf (visited on 05/10/2019).
- Weng, Tsui-Wei et al. (2018). "Evaluating the Robustness of Neural Networks: An Extreme Value Theory Approach". In: *arXiv:1801.10578 [cs, stat]*. arXiv: 1801.10578. URL: <http://arxiv.org/abs/1801.10578> (visited on 05/24/2019).
- West, Darrell M. (2018). *What is artificial intelligence?* Brookings. URL: <https://www.brookings.edu/research/what-is-artificial-intelligence/> (visited on 04/25/2019).
- Yeboah-Boateng, Ezer Osei and Francis Edmund Boaten (2016). "Bring-Your-Own-Device (BYOD): An Evaluation of Associated Risks to Corporate Information Security". In: *arXiv:1609.01821 [cs]*. arXiv: 1609.01821. URL: <http://arxiv.org/abs/1609.01821> (visited on 05/27/2019).

8. Appendix

8.1 Appendix A: Interview Questions

The following lists show the series questions answered by over 40 interviewees from across 10 different countries. The inquests were categorized according to the interviewee's area of expertise. The classified domains were:

1. Academia,
2. Cybersecurity vendors,
3. Firms employing cybersecurity products/services
4. Governmental bodies and policy think-tanks.

8.1.1 Questions for academia

- How has AI been applied in cybersecurity so far?
- How will the deployment of AI in cybersecurity benefit various sectors and industries?
- What does a successful implementation of AI for cybersecurity look like?
- In what cases would using AI for cybersecurity be unfavourable? What factors would you worry about?
- What are some limitations related to the use of AI in cybersecurity?
- What are some risks posed by the use of such cognitive technologies in cyber-crime? How do you think those risks can be mitigated?
- Are there any risks or harm posed to users by AI implemented in security?

8.1.2 Questions for cybersecurity vendors

- How does your AI-enabled cybersecurity products provide value for consumers?
- Can you (broadly) discuss examples in which your products have helped customers defend themselves against attacks?
- What are the main roadblocks/challenges to the development of your product?
- What challenges do consumers face in adopting AI based security products?

8.1.3 Questions for firms employing cybersecurity services

- What role do AI/Machine learning tools play in your defensive toolset?
- What are the main benefits that AI-based tools provide your organization?
- Where do most of your AI-based tools come from?
- To what extent do you trust the decision made by your AI tools? Why?
- To what extent are you concerned about adversaries using AI tools in their attacks?

8.1.4 Questions for governmental bodies and policy think-tanks

- How has government progressed in cybersecurity and AI in cybersecurity thus far?
- Does GDPR affect the adoption of AI in cybersecurity?
- What do you reckon about the future of AI in cybersecurity, and what kind of policies and regulations do we need in the future?
- To what extent do you think government will support the use of AI in cybersecurity, and in what ways?

8.1.5 General Questions (other)

- What do you see as a primary barrier for the adoption of AI in cybersecurity (AIC)? To what extent do you think the GDPR poses a problem for this adoption? *

- Technical barriers: maturity of AIC, projected effectiveness of AIC, availability of data
- Financial barriers: quantifying costs and benefits, how much to invest
- Privacy barriers: Data protection (from providing data)
- Regulatory barriers: policy support or lack thereof
- What can governments do to make that adoption easier? *
 - Are there any regulatory measures they should take?
 - How should they invest resources?
- Moving forward, what are some developments in AI that could aid in cybersecurity in the next 5-10 years?
- What are some changes or developments necessary for enhancing current applications or supporting future applications? Do think those changes will happen in the next 5-10 years?
- Is there anything government can do to help accelerate the development/adoption of AI-enabled cybersecurity products?
- What is the general process through data for your AI tools are acquired? How do you ensure training data for you tools are of good quality?
- When is it important to have humans involved, or not involved (human in the loop)? How much is it appropriate to delegate decision-making to algorithms?
- What do you make of the tension between explainability and security for cybersecurity AI algorithms?

8.2 Appendix B: Potential Impact Data

Application	Potential Savings in Dollars	Log of Savings
Malware Analysis	\$331,224	5.52
Phishing and Spam	\$481,067	5.68
Intrusion Detection	\$475,594	5.68
User and Event Behavior Analysis	\$284,234	5.45
Application Security	\$229,961	5.36
Vulnerability Management	\$461,709	5.66
Threat Intelligence	\$479,071	5.68
Decision Support and Workflow Automation	\$375,201	5.57

Table 8.1: Potential Savings in Dollars per Application. Based on Ponemon report - Value of AI in Cybersecurity (2018) and Accenture Cost of Cybercrime Study (2017)