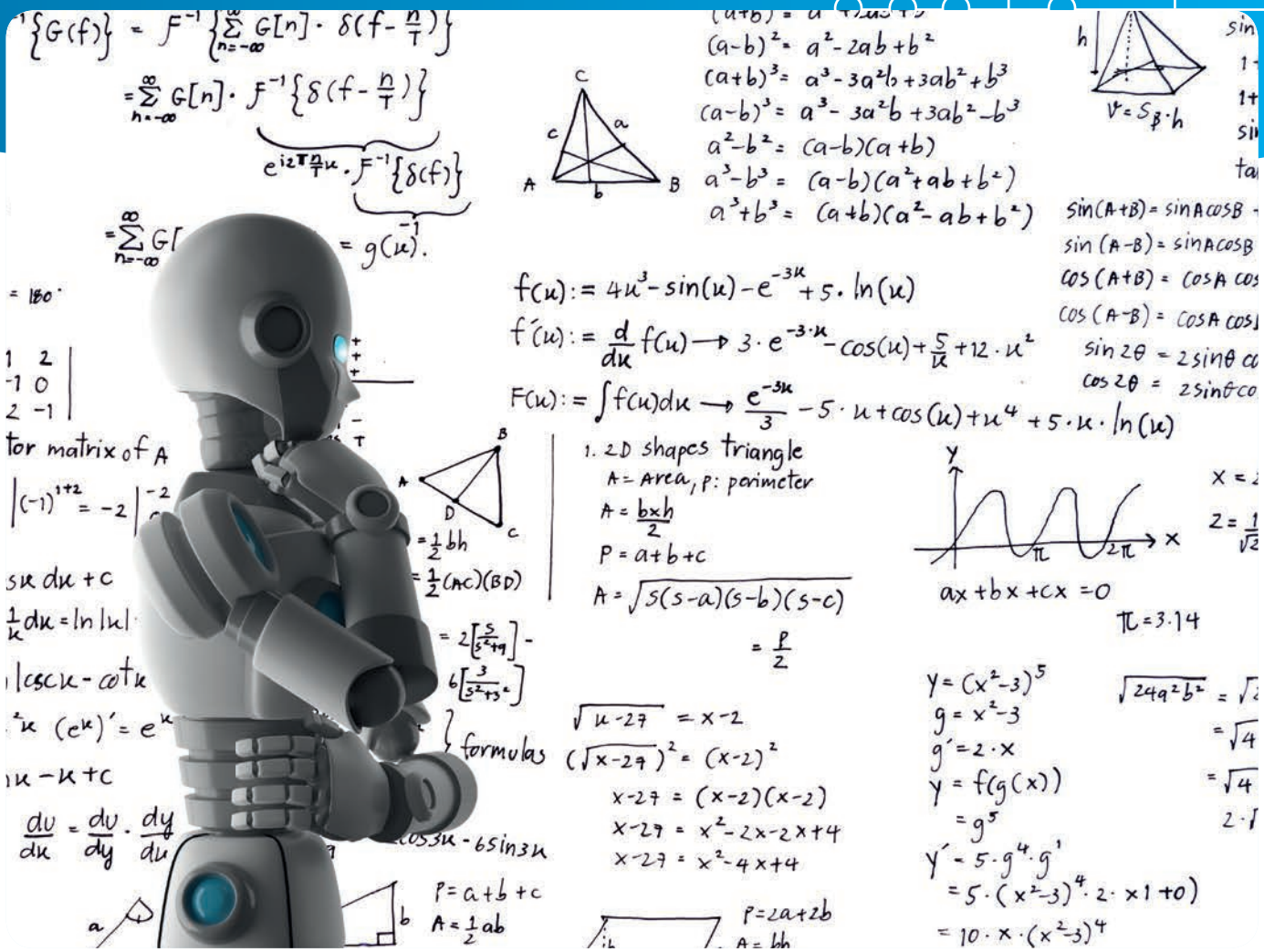
# Understanding the Strategic and Technical Significance of Technology for Security

## *Implications of AI and Machine Learning for Cybersecurity*

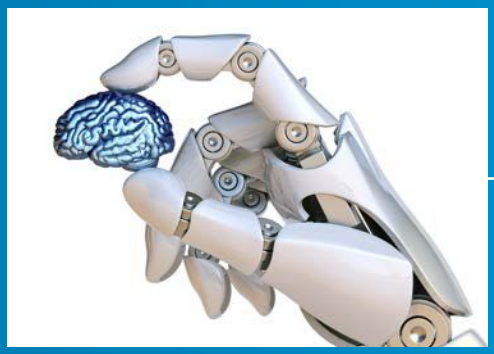


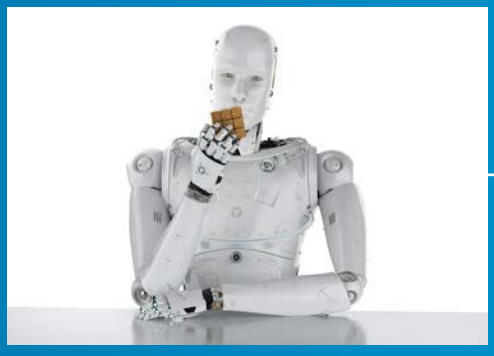
The Hague Security Delta

# Understanding the Strategic and Technical Significance of Technology for Security

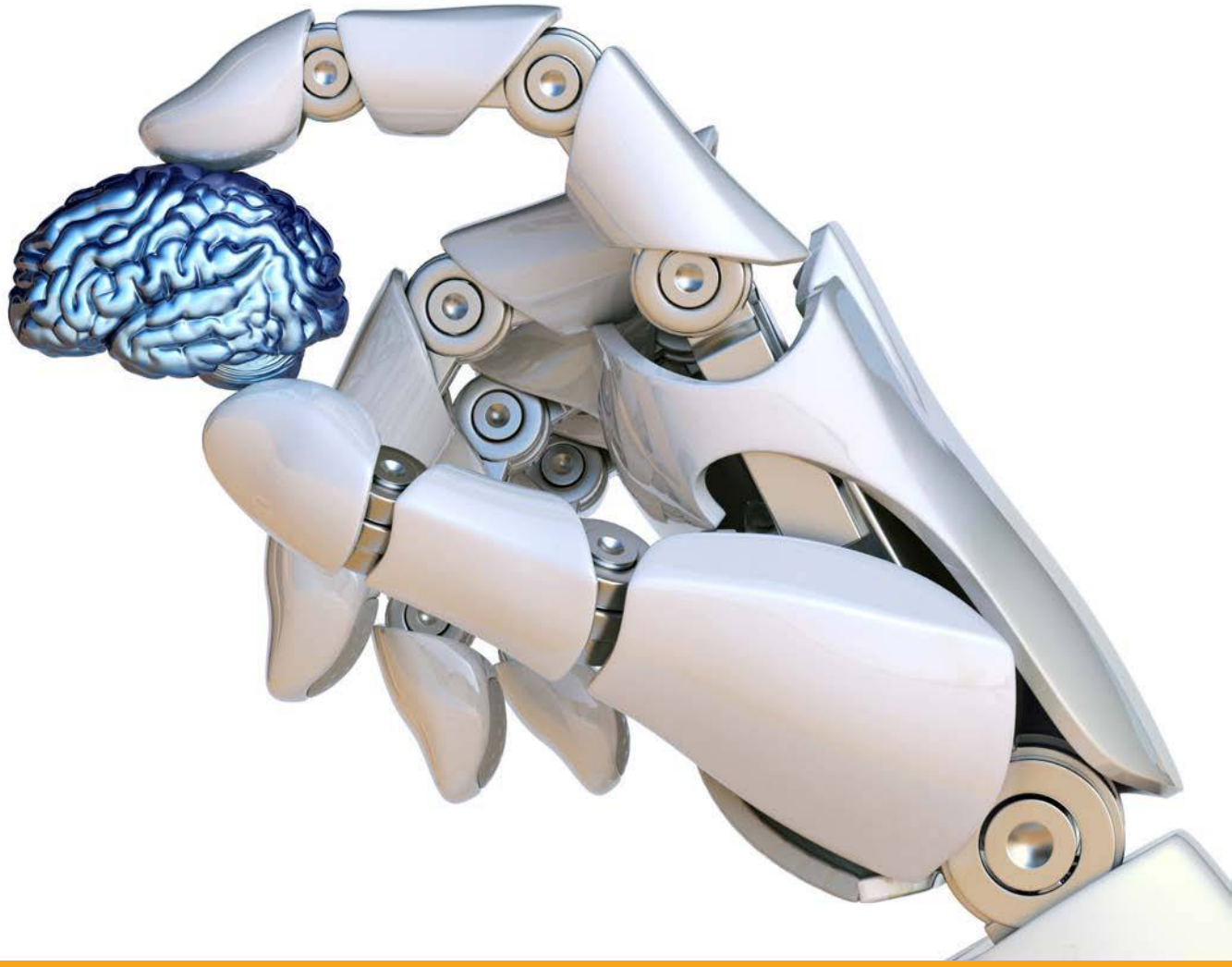
## *Implications of AI and Machine Learning for Cybersecurity*

The Hague **Security Delta**





# Table of Contents

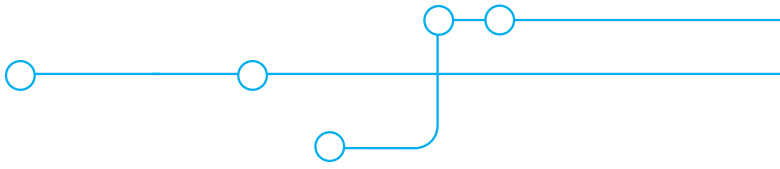


# 1 – Introduction

Artificial Intelligence (AI) has made exponential progress in recent years, especially in terms of Artificial Narrow Intelligence (ANI) and machine learning. As the amount of data breaches and cybersecurity incidents grow, AI is increasingly being hailed for its new way to automatically spot any malware on a network, guide incident response, and detect intrusions before they even occur. The 2018 Ponemon Institute's "Artificial Intelligence (AI) in Cyber-Security" study, for example, shows that AI is able to detect 63% of previously undetectable zero-day exploits.[1] However, despite the potential benefits of AI being touted as a game-changer, estimates on its impact on cybersecurity still vary widely.

Cybersecurity is a field where absolute security is impossible. Instead its objective is to reduce the attack surface to a minimum. The rosy view of what AI can deliver is not entirely wrong, but what next-generation techniques actually do is more muddled and incremental than marketers would want to admit. Fortunately, researchers developing new defense techniques at companies and in academia largely agree on both the potential benefits and challenges.

This study explores how machine learning, in particular unsupervised learning, can play a role in cybersecurity.[2] Chapter 2 introduces the body of AI and the different forms of machine learning. Chapter 3 looks at the possible application and weaknesses of machine learning to improve cybersecurity, while chapter 4 identifies the macro bottlenecks for the technology. Overall, the study uses recent literature on the subject in light of contextual examples, and presents some suggestions and recommendations for Dutch stakeholders seeking to understand how to best profit from the development from a socio-economic context.



# 2 – Introduction to AI and Machine Learning

## 2.1 Artificial Intelligence

The term Artificial intelligence (AI) has become ubiquitous in the media in recent years. However, there is little understanding of what exactly this technology entails. This briefing explains some of the basic AI technologies, and concentrates on a specific technology field known as "unsupervised learning".

Artificial Intelligence is typically classified in three categories: Artificial Narrow Intelligence (ANI),[3] Artificial General Intelligence (AGI), and Artificial Superintelligence (ASI):

| Artificial Narrow Intelligence (ANI) | Artificial General Intelligence (AGI) | Artificial Super Intelligence (ASI) |
|---|---|---|
| Able to *match or exceed* human ability in a *specific task or domain* (e.g. speech recognition) | Able to *match or* exceed human ability in its full range (e.g., AlphaGo that defeated a human in a game of Go) | Able to *exceed* human ability in both *range and ability* |

AI is an extremely broad and nebulous field that encompasses many methodologies, within which machine learning can be seen as a subset or means to achieve AI. Most of the new development and investment within AI is dedicated towards machine learning, with around 60% of all AI investment in 2016 going towards machine learning technology.[4]

## 2.2 Machine Learning

At its core, machine learning is one step or instrument towards achieving all three types of AI. It is generally defined as the usage of algorithms to analyze troves of data in order to train computers to recognize patterns and discern valuable information without human intervention.[5] The fundamental goal of machine learning is to generalize beyond the examples in the data training set.[6] Machine learning itself is not necessarily more effective than human learning. Rather, machines are able to go through data much faster than humans.[7]

Machine learning is chiefly a predictive technology, and many of its tasks fit into several general categories outlined in the table below:[8] For nearly every given cybersecurity application, multiple tasks are used at different stages.

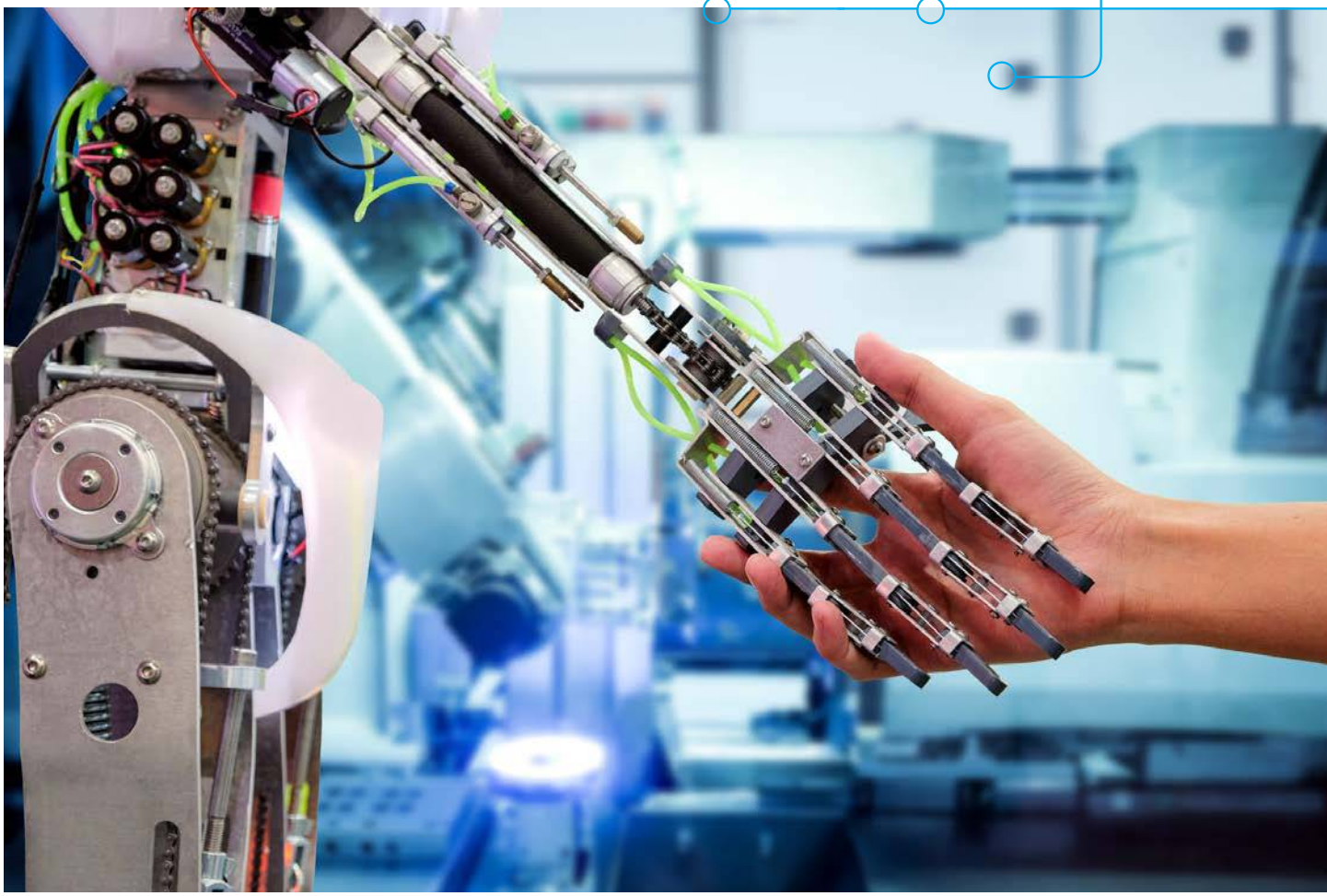
| Function | Cybersecurity application |
|---|---|
| **Classification** tasks associate input data with labels, assigning the data points into certain categories | Spam filters |
| **Regression** tasks attempt to predict the next value based on a set of data that may influence that variable (prediction) | Fraud detection and anomaly detection |
| **Clustering** tasks is the unsupervised learning version of classification. This means that there is no pre-existing knowledge about the classes of the data or whether the data can be classified at all. | Malware and forensic analysis, behavior analytics |
| **Dimensionality reduction** tasks reduce the number of random variables under consideration so that redundant features are eliminated and further data processing is less intensive – similar to classification and clustering but deal with more complex systems, unlabeled data and many potential features. | Face detection solutions (part of two-factor authentication and identifying "deep fakes") and often a supportive task |
| **Association Rule Learning** tasks are used for recommendation systems that have many applications in commerce and entertainment. In the context of cybersecurity, a particular type of response is linked to a particular incident. | Incident response and risk management |
| **Generative models** differ from the rest. Whereas the previous tasks deal with the existing information and associated decisions, generative models simulate the actual data based on the previous decisions | Vulnerability scanning (crawler) |

## 2.3 Supervised versus Unsupervised Learning

Within machine learning two main approaches can be identified – *supervised and unsupervised* learning. These processes are not always mutually exclusive and a solution to a given machine learning task may, at various stages, use both supervised and unsupervised learning techniques. This also goes for cybersecurity. Machine learning is not a one-shot process of building a data set and running a learner[9], but rather an iterative process of running the learner, analyzing the results, modifying the data and/or the learner, and repeating. It often still requires human intervention.

Supervised learning refers to processes of machine learning that require labeled data as input for the artificial agent. Supervised learning is named as such because the data scientist acts like a teacher for the AI, labeling the input data with associated facts and determining the output. The algorithm's possible outputs are therefore known and the data is labeled with 'correct' answers. Typically, this process is applied to classification or regression tasks.

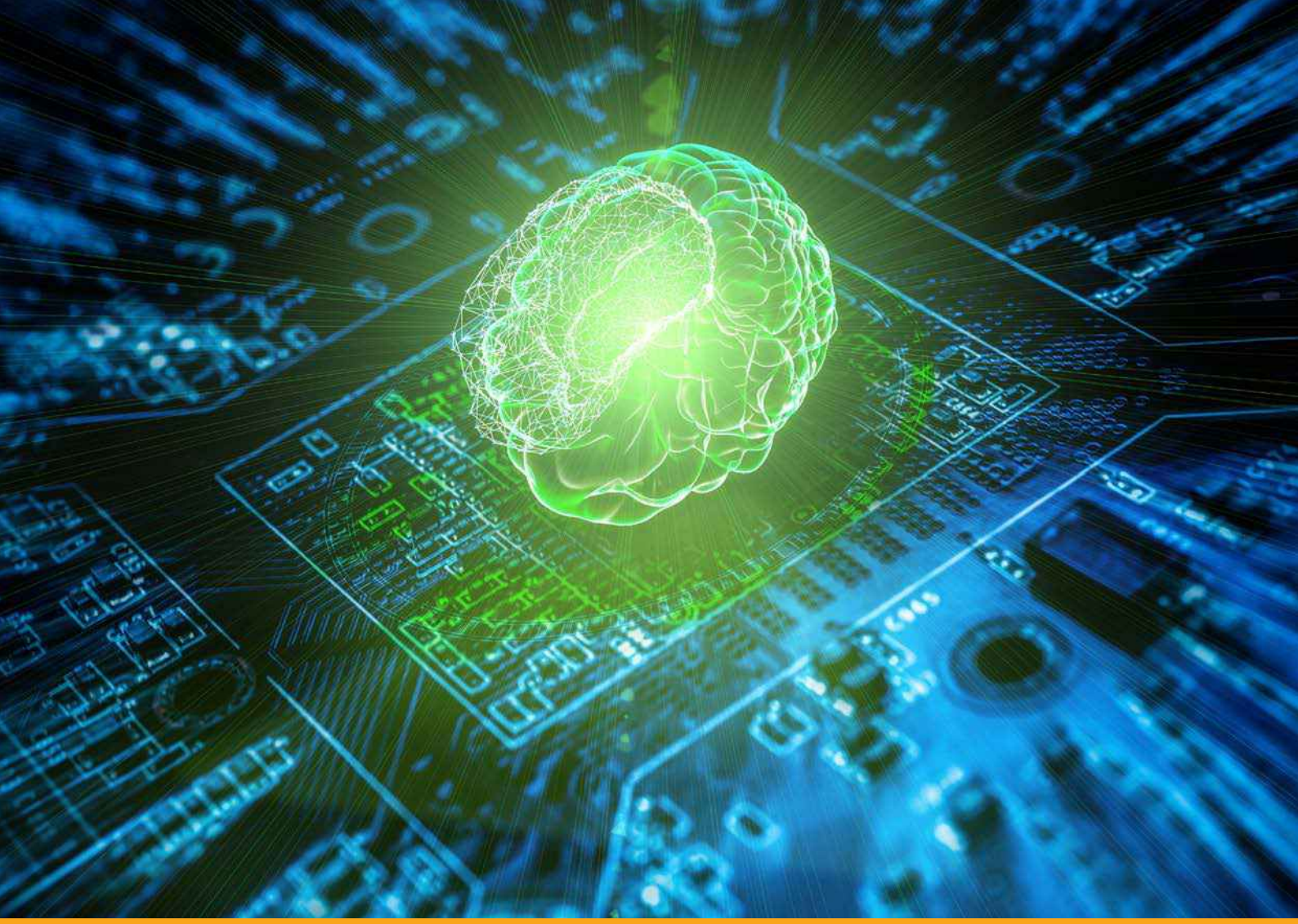
In unsupervised learning, the AI will learn from data sets that have not been classified and in doing so react to either the presence or absence of commonalities in the data.[10] The algorithms infer patterns from data sets without reference to known or labeled outcomes. Sometimes labeled data is very rare, or the task of labeling is very time-consuming, or we may not even know if labels exist. For example, consider the case of network flow data. Attempting to label this huge amount of data would be extremely time-intensive, and it would be very hard for a person to determine labels for each data point. Given how good machines are at finding patterns in large datasets, it is often much easier to have the machine separate data into groups.



The objective of unsupervised learning algorithms is to discover the underlying structure of the data in a given set. Unsupervised learning methods are therefore most suitably applied to tasks such as clustering, anomaly detection, association and dimensionality reduction. To have useful AI that employs unsupervised learning thus requires the actual algorithms, a set of data, and sufficient computational power.

Other approaches include *semi-supervised learning* (analysis on labeled data alongside unlabeled data), and *reinforcement learning* (where an AI agent learns to interact with its environment through trial-and-error).[11]

*Deep learning* is a subset of machine learning that recognizes patterns of patterns – like image recognition – in which tasks are broken down and distributed onto machine learning algorithms that are organized in consecutive layers.[12] "Each layer builds up on the output from the previous layer. Together the layers constitute an artificial neural network that mimics the distributed approach to problem-solving carried out by neurons in a human brain."[13]



# 3 – Machine Learning in Cybersecurity

From a technical perspective, cybersecurity is approached from the CIA-triad - to protect the confidentiality, integrity, and availability of data and networks. As previously mentioned, it is a field where absolute security is impossible. Instead, its objective is to reduce the attack surface. This part of the study parses AI in cybersecurity and offers some examples on the application of machine learning for offensive and defensive purposes.

## 3.1 Automating the Offense

Increasingly, annual reports of cybersecurity companies are warning about the ways attackers will further adopt AI techniques.[14] This includes autonomous malware and offensive cyber capabilities, such as phishing, spam, DDoS, ransomware, spyware, as well as other applications in computational propaganda like deep fakes, which is addressed as a separate issue at the end of this section.[15]

Cyberspace is already defined by the asymmetric power relations between attack and defense. Offensive capabilities are much cheaper and easier to develop and deploy than the necessary defensive measures. The success of an attack is more a reflection of the overall quality of defense rather than the quality of offense. An attacker therefore tends to use the cheapest and easiest tools available, and not necessarily the most advanced.[16]

Nearly all modern devices have some degree of computing, storage, and network capacity that can be appropriated and abused. The proliferation of Internet-connected devices has led to a larger attack surface that can exploited.[17] From an offensive perspective, AI and machine learning can leverage that large attack surface better than before and lead to an increase in the speed, adaptability and agility of an attack and perhaps even its sophistication.[18] The Mirai malware, for example, capitalized on the relative insecurity of newer Internet of Things (IoT) devices and marshaled large populations of these ubiquitous network-enabled devices into coordinated botnets to execute massive DDoS attacks of unprecedented capacity.[19]

Unsupervised learning can be used to conceal various forms of malware within a victim's network, as well as generate credentials to infiltrate IoT devices by automatically cycling through password and username options at a speed faster than a human could test, and in cases where existing "rainbow tables" (massive dictionaries of password hashes) are insufficient.[20] Attack systems will self-learn how and when to attack their target system, be able to change behavior when under counterattack and seek out new targets when the original attack vector is mitigated. Current malware already exposes such traits, albeit in a limited and a pre-programmed manner. For example, it will attempt to cycle through various attack vectors once it determines that the ones under attack are no longer responding, such as attacking different servers after failing to successfully assault one. However, current malware cannot yet independently find new avenues of attack without pre-programmed instructions.[21]

Adaptivity will become a core characteristic of both attack and defense systems. Through supervised and unsupervised learning, attack systems will try to avoid detection and evade defensive or responsive measures through self-initiated changes in signature, behavior and goal-planning. As attack systems become more dynamic in behavior, more voluminous and target a far larger attack plane, the human capacity to identify and mitigate attacks will fall short. This will warrant the development of new defensive systems that use machine learning to support the security analyst.[22] We are already seeing the first wave of commercial products for autonomous anomaly and intrusion detection, threat identification, and mitigation.[23]

As diverse and integrated data becomes available, increased accessibility can lead to stronger and more nuanced threats. Many public data sets are appealing due to their lack of manual labor in creating training data sets but they may include biases and are also available to the creators of malware.[24] Criminals and state actors can also influence such systems by 'poisoning' datasets with false data, or data that disrupts the functioning of the system. If attackers can figure out how an algorithm is set up, or

where it draws its training data from, they can figure out ways to introduce misleading data that builds a counter-narrative about what content or traffic is legitimate versus malicious. For example, attackers may run campaigns on thousands of accounts to mark malicious messages or comments as "Not Spam" in an attempt to skew an algorithm's perspective.[25] Researchers built a machine learning-based phishing attack generator that trained on more than 100 million particularly effective historic attacks to optimize and automatically generate effective scam links and emails. The attack generator was trained on open-source data that would be available to potential attackers and shows how increased access to data can have serious ramifications for applications such as AI-based detection systems.[26] In a world that depends on systems that self-learn from self-acquired or pre-built datasets, intentional manipulation of data becomes a serious attack vector, and one that is almost impossible to detect.

Researchers from Endgame released an open source threat data training set called EMBER,[27] with the hope that they can set an example – even among competing companies – to focus on collaboration in machine learning, particularly on defensive measures.[28] Collaboration between defenders and researchers may be necessary to stay ahead of attackers using machine learning techniques themselves.

## 3.2 Antivirus Defense

The most obvious first defensive cybersecurity application of machine learning can be found in anti-virus defense and malware scanning. Traditionally, anti-virus defense has been signature-based. This means solutions scan specific malicious programs, extract a unique fingerprint, and monitor customer devices to ensure that none of those signatures re-appear. Malware scanning using machine learning works in a similar fashion: the algorithms train how to identify malware from vast catalogues of viruses, but instead of looking at specific signatures the machine learning tool has learned to look for characteristics of specific malware families and is therefore more flexible: “Where attackers could stymie traditional anti-virus by making just slight alterations to their malicious tools that would throw off the signature, machine learning-based scanners, offered by pretty much all the big names in security at this point, are more versatile. They still need regular updates with new training data, but their more holistic view makes a hacker's job harder.”[29]

## 3.3 Spam, Phishing and Communication Filtering

Similarly, machine learning is already indispensable in countering spam and phishing as well as online forms of fraud.[30] Consider an email spam detection algorithm: original spam filters would simply blacklist certain addresses and allow other mail through. Machine learning improved this by comparing messages classified as spam emails with messages classified as legitimate email and identifying "features" that were present more frequently in one or the other. For example, intentionally misspelled words (e.g. V!AGRA), the presence of hyperlinks to known malicious websites, and virus-laden attachments are likely features indicative of spam rather than legitimate email. This process of automatically inferring a label (i.e., "spam" vs "legitimate") is called classification – one of the major applications of machine learning techniques.[31] Most machine learning having to do with spam filtering thus relies upon sets of training data.[32] Machines learn from classification problems, and process both non-spam and spam emails to then learn to distinguish between the two. Not only is this relevant for spam filtering, but also more generally for the financial industry and the detection of fraud. Unsupervised learning systems will continually detect outlier behavior, whether or not that data was in the training sample. Similarly, the same techniques can be employed by large signal intelligence organizations to help them identify objects of interest, but also minimize objects that they are not allowed to examine. For instance, it is known that the US National Security Agency applies so-called “minimization procedures” to automatically scan communication - mostly email - to determine if the sender or receiver can be labeled as a “US person”, and thus largely protected from surveillance. These minimization procedures are employed in unsupervised machine learning as one of the techniques to sort through the mass of communication they automatically collect.[33]

As offensive strategies evolve and phishing schemes become more pernicious by setting up fake but convincing links or tampering a spam filter's idea of which messages are malicious, service providers need to adapt to hackers who know how to evolve, and they use machine learning to keep up. As a result, companies like Google have found applications for machine learning in almost all of its services, especially through deep learning, which allows algorithms to do more independent adjustments and self-regulation as they train and evolve. In an interview with Wired magazine, Elie Bursztein, who leads the anti-abuse research team at Google, claims that

"Before we were in a world where the more data you had the more problems you had. Now with deep learning, the more data the better. We are preventing violent images, scanning comments, detecting phishing and malware in the Play Store. We use it to detect fraudulent payments, we use it for protecting our cloud, and detecting compromised computers. It's everywhere."[34]

### 3.4 Fuzzing and Crawling for Zero Days

The 2018 Ponemon Institute's "Artificial Intelligence (AI) in Cyber-Security" study, indicates that AIs are able to detect 63 percent of previously undetectable zero-day exploits.[35] Zero days is the shorthand for undiscovered vulnerabilities in code that represent one of the surest ways to hack a system. They are essential for advanced cyberattacks, and they have been described as being the "bullets of cyberwar." Finding zero days is therefore an essential task for the offense.

Fuzzing is a method in which random data inputs are fed into the system until one of the permutations reveals a software vulnerability that can be hacked.[36] It's an old but common process that allows attackers to find and exploit vulnerabilities, and allows defenders to find and fix them first. It has become an essential tool in the zero day arms race. The process of fuzzing has already been automated to a large extent, and applications of machine learning are already improving fuzzing techniques, its locations and the strategies and parameters used.[37] Microsoft, Google, Baidu and many other tech companies are using resources to refine the fuzzing process using machine learning, and companies like Peach Fuzzer and Codenomicon have established their business model around this process.[38]

A crawler is a program designed to traverse websites to retrieve HTML documents.[39] Once the focused crawler targeting a specific topic or virtual community has been manually configured, cyber-artifacts of interest can be collected from multiple virtual communities in an automated fashion with little supervision.[40] Already, machine learning plays an important role in the automation of the crawler. Furthermore, experts from the Arizona State University have developed an AI crawler - an operational system for cyber threat intelligence-gathering that uses machine learning models. It identifies emerging threats by collecting information from discussions on hacker forums and marketplaces offering products and services for malicious hacking. These threat warnings include information on newly developed


malware and zero day exploits that have not yet been deployed. This provides a significant service to cyber-defenders. The system is augmented through the use of various data mining and machine learning techniques. With the use of machine learning models, they claim to be able to recall 92% of products in marketplaces and 80% of discussions on forums relating to malicious hacking with high precision. It offers insights to cybersecurity experts in terms of what vendors and users have a presence in multiple deep web markets/forums, what zero day exploits are being developed, and what vulnerabilities the latest exploits target.[41]

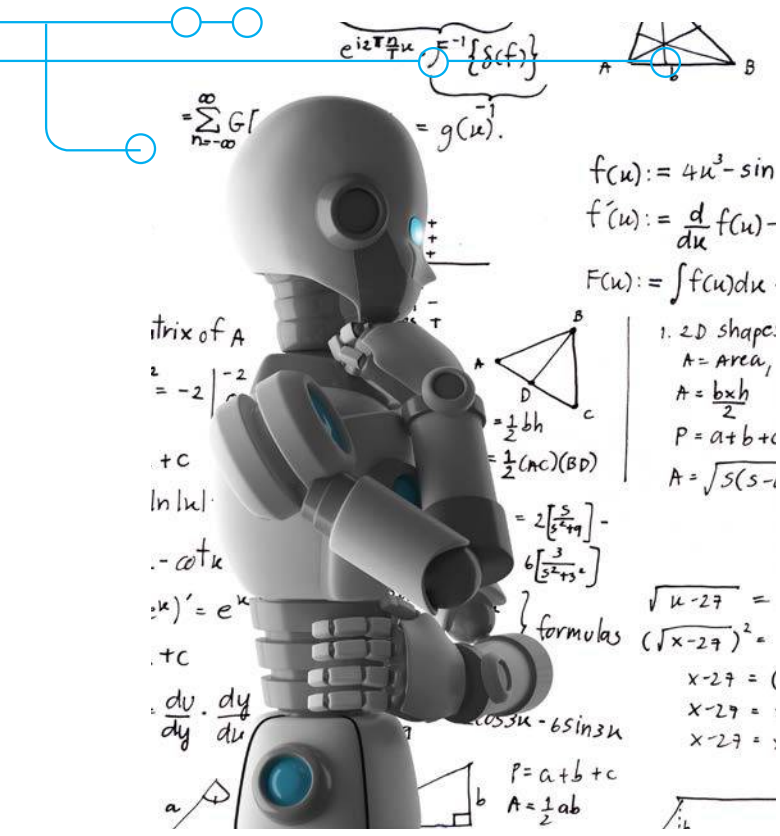
### 3.5 Setting the Baseline

At its core, machine learning's biggest strength in cybersecurity is to help understand what is "baseline" or "normal" behavior for a system, and then flagging anything unusual for human review: i.e. anomaly detection. Machine learning essentially speeds up this process for the security analyst. This is particularly important since cybersecurity has moved from perimeter defense to also include network scanning for unusual behavior or anomalies that may constitute a breach. This concept applies to all sorts of machine learning-assisted threat detection, but researchers say that the machine learning-human interplay is the crucial strength of the techniques.[42]

From a defensive point of view, there is a constant push to respond in real time and to mitigate attacks. Automating network scanning and anomaly detection allows suspicious behavior to be pinpointed much faster, mitigate a potential breach, minimize its impact, and thereby enhance the resilience of a system. Companies like Symantec employ targeted attack analytics using advanced algorithms and machine learning for precise

detection of suspicious activity on cloud-based platforms.[43] Other examples of time-saving machine learning systems that scan log files for suspicious activity come from academia.[44] In fact, sixty-nine percent of respondents from the Ponemon Institute's Artificial Intelligence Study state that the biggest benefit of AI in cybersecurity will be an increase in the speed of analyzing threats and the containment of infected devices, saving 2.5 million USD in operating costs.[45] Another way to reduce response time and human intervention is coordinated behavior among intelligent AI systems (e.g. multi-agent system coordination), which allows network protection systems to assess attack strategies among themselves.[46]

This is a promising development when one takes into consideration the cybersecurity information overload (200,000 security events per day) and skill shortage (1.5 million unfilled security jobs by 2020) in the field. The application of machine learning in cybersecurity therefore addresses the acute problem of scarce and expensive cybersecurity expertise through resource optimization or increases in staff productivity.[47] It can therefore lead to more informed decisions by security experts at an unprecedented speed and scale. A reduction in false positive rates would also positively impact cybersecurity operations and machine learning is effective in achieving this goal. Because machine learning is able to help keep track of what historically have been false positives and therefore can adjust the AI heuristics, it is far less likely to flag false positives than a human would. This means systems will be able to focus their processing power on protecting against actual attacks. In other words, false positives are costly and too many of them will overwhelm security systems, meaning fewer time to deal with actual threats.[48]



AI and machine learning offers training opportunities for a number of different fields. NATO's Cooperative Cyber Defence Centre of Excellence runs a yearly exercise called 'Locked Shields' which puts teams to the test in defending a fictional country against a severe cyberattack. To this end, it offers a course that uses machine learning as a tool for solving practical monitoring challenges.[49]

## 3.6 Patching

Machine learning is no cure-all, most notably because it won't help in the distribution of security patches. What it can do is help identify security holes – if current techniques are perfected.[50] Today, the average number of days for companies to patch a software vulnerability varies wildly. According to Bruce Schneier, a leading cybersecurity expert, the industry rule of thumb is that "a quarter of us install patches on the day they're issued, a quarter within the month, a quarter within the year, and a quarter never do." The patch rate is even lower for military, industrial, and healthcare systems because of how specialized the software is.[51]

In the early 2010s, experts at the US Defense Advanced Research Projects Agency (DARPA) observed advances in critical areas of computer science necessary to automate cybersecurity and patching of software using machine learning.[52] In 2016, DARPA hosted the AI Cyber Grand Challenge: "a competition to create automatic defensive systems capable of reasoning about flaws, formulating patches and deploying them on a network in real time."[53] Prizes of $2 million, $1 million, and $750 thousand were awarded to the top three finishers. Eventually, the tech behind the winner is now used by the Pentagon.[54] The company claims it has started adapting its machine learning technology to be able to automatically find and patch flaws in certain kinds of commercial software, including that of internet devices such as routers.

This competition is a good example how government (and other stakeholders) can function as a responsible accelerator for local talent and startups through which it can ultimately benefit as well.

## 3.7 Computational Propaganda

AI and ML technologies are not only changing the way computational propaganda is produced and disseminated, but also have the potential to play a central role in tackling disinformation[55] in the long run. Computational propaganda is a term that encompasses recent manipulative disinformation efforts which accompanied the 2016 US Presidential Election, for example. It is best defined as "the use of algorithms, automation, and human curation to purposefully distribute misleading information over social media networks."[56]

Computational propaganda is considered to constitute one of the most powerful new tools against democracy.[57] In the Russian context, Moscow is believed to pursue a 4D offensive: 1) to dismiss an opponent's claims or allegations, 2) to distort events to serve political purposes, 3) to distract from one's own activities, and 4) to dismay those who might otherwise oppose one's goals.[58] Until now, disinformation operations have relied on human operatives to produce content, and on a combination of human 'trolls' and basic automated algorithms to spread that content.[59] Online platforms have the capacity to defend against most of the tools employed by information adversaries, as predictable 'bot-like' patterns make it possible to identify most automated algorithms.[60] For example, in August this year, Facebook successfully removed 652 fake accounts and pages with ties to Russia and Iran that attempted to influence political outcomes in the US, UK, Middle East and Latin America.[61] Twitter has also taken steps to remove fake accounts and shut down bot networks.[62]

Advances in AI and machine learning systems may be the key to solving the spread of computational propaganda. Existing tracking systems are based on manual fact checking, which is neither effective nor efficient to evaluate every piece of information that appears online. Taking a cue from articles that were flagged as inaccurate by users and fact checkers in the past, AI can be leveraged to find (patterns of) words that can identify false stories. Through pattern recognition, it is possible to identify harmful online behavior. As this technology is still under development, there is a risk of false positives – identifying stories as fake when they are not.[63] As such, the focus has been on using a combination of machine learning to do the most repetitive work (i.e. making basic judgments on large amounts of content), and people for making more nuanced and complex decisions that require deeper expertise.[64]

Google, Facebook and Twitter already rely on machine learning to stamp out trolls, remove fake bot accounts, and to proactively identify dubious content. In addition to its image search algorithm, Google introduced a browser extension, which allows users to flag misleading information as well as to see news flagged by others.[65] Facebook already applies machine learning to detect sensitive content – videos and photos containing violence, sexual and terrorist content – and is now moving to use similar technologies to detect false stories, as well as to spot duplicates of stories that have already been debunked.[66] The platform recently acquired a London-based start-up Bloomsbury AI to help improve language processing, and to better tackle disinformation spread on the platform.[67] For the time being, Facebook continues to rely on a combination of technology and human review.

While advances in machine learning technologies will unarguably benefit those who defend against computational propaganda, they are also likely to allow adversaries to magnify the effectiveness and scale of their operations in the short term.[68] As argued by Rosenbach and Mansted, "breakthroughs are likely to spread quickly and widely, equipping both state and non-state adversaries with a technological edge".[69]

Three threats stand out in particular. First, with advances in machine learning, adversaries will progressively be able to identify and target those audiences that are most susceptible to influence. Known data inputs (for example, indicators of race or gender), combined with machine learning tools that predict with growing accuracy people's political and religious beliefs, personality and preferences, real-time emotions, or identity characteristics such as sexual preference, adversaries will be able to deliver highly-personalised content with maximum effectiveness.[70] Second, by creating tools that are more capable of understanding human language and of processing it in the right context, the next wave of AI research may put *bots in the driver's seat* with little to no human supervision at the tactical level. In this scenario, humans will be pushed out of certain tasks that would have previously never been considered to be handed over to machines. The third looming threat are the so-called *'deep fakes'* – digitally manipulated audio or visual material that is highly-realistic, cheaper and virtually indistinguishable from real material.[71]

# 4 – The Bottlenecks of Machine Learning

The use of machine learning technologies in cybersecurity bring along bottlenecks and systemic vulnerabilities that can be exploited. Traditional machine learning risks, such as input data bias[72] and overfitting,[73] remain obstacles but are too technical in nature for the scope of this study. Instead this chapter focuses on three core macro components that drive AI and machine learning technologies: **(i) software algorithms, (ii) data, and (iii) computational power.**[74] This study does not presuppose a hierarchical order where one component is more important than the other. Rather it deals with them as individual legs of a tripod and addresses them from a Dutch socio-economic approach to identify points for further research or development. This ought to be placed behind an important international context. Media reports about humongous Chinese investments[75] indicate there is increasing belief that the development of the AI industry and clusters is not only possible, but an essential part of government policy both for local economic development but also national security.

## 4.1 From Software Algorithms to Talent

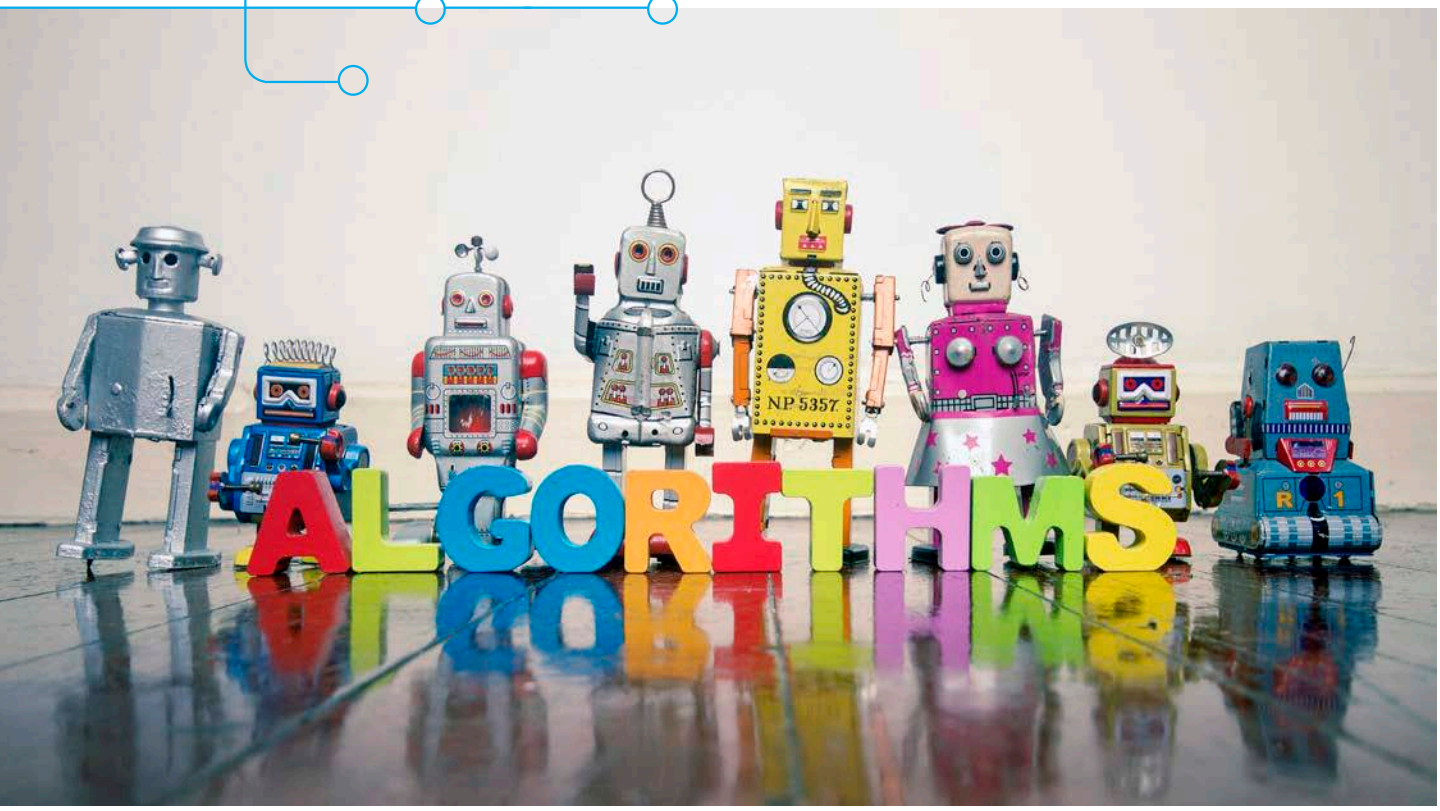
One leg of the tripod is the development of algorithms and the creation of better environments in which these algorithms can thrive. Algorithms have four logical building blocks: the procedure, the sequence, alternatives, and iteration. In other words, to build an algorithm a function must be selected, it must operate in a pattern, possess a series of 'if then' statements of alternative solutions to the problem, followed by an iteration to build loops in order to solve problems.[76] In essence, algorithms are created as the mathematical solution to problems expressed in code.[77] One of the main software challenges to building complex algorithms is known as the black box problem. This refers to the situation in machine learning where an AI has determined the solution to a problem using algorithms and data, but the scientist does not know precisely how it determined the answer.[78] They can see the input and the output but not the process. This makes further advancement in the field difficult.

On the basis of the interviews, it became clear that there is cooperation between Dutch companies (that lead in machine learning implementation) and algorithm experts[79] from Dutch technical universities. For now this is largely done on a case-by-case or project basis to address a specific problem. Cooperation is not happening on a structural basis. This means that much of the algorithmic knowledge is not being transferred, which adds to the black box problem, especially when there are no evaluation mechanisms. There is a need for large Dutch companies that are already using these techniques to leverage algorithmic power in a more structured way than is currently happening.

We now turn to the components that could potentially lead to a more conducive environment for software algorithms to be developed, in particular for the Netherlands. A starting point is to identify the best practices of algorithm teams. To this end, the Google Brain Team, the research team in charge of machine learning algorithms and techniques, serves as an example. In addition to its headquarters in California, it decided to set up satellite teams in Boston, London, Montreal, New York, San Francisco, Toronto, Zurich, and more recently in its office in Amsterdam. The motivation behind the expansion to Amsterdam offers some insight into the components that are important for its algorithm teams:

> "In general, the primary driver for our choice of AI research office locations is the **availability of talent** who are either already living in the area or would be **attracted to move there**. Aside from that it is the usual factors, such as **strength of local academic institutions, business friendly environment and good transport infrastructure.** We think that with world-class universities and a thriving developer community, Europe is well positioned to play a leading role in AI research and application."[80]

If we take a closer look at the primary driver, availability of talent, which is closely connected to the strength of its local academic institutions, the number of Dutch students who want to study computer science (and especially AI) at university is increasing.[81] However, Dutch Universities are having difficulties accommodating this increase in interest. A number of universities offering full-time bachelor or master degrees in these fields are considering or have already introduced a numerus fixus.[82] To close the gap, the capacity of Dutch universities needs




to increase, i.e. more experts and scientists are required to teach students.[83] Other ways to increase capacity are found in the increase of training places through cooperation between universities of applied sciences and industry. One way to cater to the short term need for more expertise in AI and cybersecurity could come from retraining and in-service training of specialists in either field. Here lies not only an opportunity for academic institutions, but also for initiatives and consortia from public and/or private organizations including summer schools, executive courses, master classes.

Dutch researchers and their work in computer science and AI are of a high quality, however they are being surpassed by other countries, leading to a decrease in its global share of academic publications.[84] Here lies not only a task for universities but also for government in reducing the gap in its research capacity compared to the AI powers. One of the components of the French AI strategy, for example, plans to make careers in AI research more attractive and lucrative to both local and foreign talents by increasing salaries of researchers, creating interdisciplinary AI institutes and strengthening partnerships between academia and industry.[85]

Other factors that may attract machine learning talent is the opportunity to collaborate with external partners, creating positions and residency programs as well as developing opportunities for AI scientists to publish their work, whether it be in academic journals or through their own platforms, such as Distill.pub.[86] There are already many Dutch public-private partnerships on the way such as TNO, Bitdefender, and Volto Labs as well as TU Delft, RoboValley and YES!Delft.[87]

Particularly as the amount of students looking to study computer science and AI is increasing,[88] there is a large opportunity for the Hague and the Netherlands to create those jobs and positions for students looking to remain after their studies. Companies will take note of the wide range of talent on offer and potentially consider moving offices and branches to the Netherlands.

One strategy that seems to be effective is holding workshops with a number of stakeholders, including local companies, academia, and larger multinational tech giants that can act as a capital accelerator. By creating an environment in which to stimulate collaboration between all of these different actors, it not only provides the stepping stone to build more Netherlands-based AI knowledge hubs, but can also encourage the participation of students and potential employees in activities and initiatives related to AI research and development. Similar to Google's Workshop on Algorithms and Optimization

in Zurich, a workshop in The Hague could instead focus on AI in cybersecurity.[89] These workshops could build on initiatives already existing in The Hague, such as Cybersecurity Summer Schools.[90]

## 4.2 Data

Machine learning analyzes troves of data to discern valuable information without human intervention.[91] Resulting conclusions are then used as input for AI applications. This methodology encourages and rewards the acquisition of large amounts of data, and the approaches taken in national AI strategies place divergent emphasis on the character and legal status of data. Regardless of the approach, data is a fundamental resource for AI development. The issues addressed here deal with the availability of data, data quality[92] (as opposed to quantity), and the security of data. These factors affect the tangible benefits that result from machine learning applications to real-world products and services. They are discussed in light of the regulatory approach,[93] as innovative approaches in the Dutch context for encouraging further research or development in AI will be particularly affected by this approach.

An abundance of data that is available for use in machine learning techniques is crucial for AI. Although data creation occurs through various means and in a wide-range of industries, that data remains confined to the ecosystem in which it originates. In order for data to be collected by others and used in AI and machine learning, it must first be made accessible. There are commercial but also regulatory incentives for entities that create or collect data to store vast amounts of it in private silos. This exacerbates the risk that control over potentially important or useful data is concentrated among a few major players.[94] In today's platform economy, value emerges not only from the ownership of intellectual property rights inherent to the data, but from the ability to make licensed use of large amounts of data for other purposes than its original intent. This reshuffling of owners and users demands a corresponding readjustment in intuitions about fair use.[95] The most obvious obstacle to egalitarian machine learning is that the highest quality datasets are inaccessible not because of copyright law, but because of secrecy.

Where AI applications are based on user-data, service providers will want to have exclusive control over data sets. The more user data a company can collect, the more it can improve data-driven services like machine learning. This will attract more users and thus more data. This positive feedback loop enables so-called "super-platforms" to consolidate market power. Even where applications are not dependent on user data there are issues in creating an open data environment – some data creators are unaware of the types of data existing within their domain, the way in which these can or should be managed, or the potential that this data holds. For example, data analytics in the oil and gas industry relies on many different sources of data;[96] by some estimates, internal data generated by O&G companies exceeds 1.5 terabytes a day. But this is not translating into direct economic benefits, partly because the rising need to expand the scope of data is being restricted by companies' weak data management capabilities.[97] If such data is not made available for effective use, a lack of available data will lead to heavier reliance on publicly available data sets, where data might be incomplete or inaccurate.

In the EU's Communication 'Artificial Intelligence for Europe', one of the primary focuses is the creation of guidelines and standards for data sets. This will mean that cybersecurity firms have more opportunities to create Artificial Narrow Intelligence (ANI) that have more experience working through clean and operable data sets.

The need to be circumspect with the security and use of data and datasets was increased with recent EU legislation. Legislation relating to data within the EU covers several aspects,[98] although data protection primarily addresses the imbalance that has long existed between those that collect data, and the data subject on the other hand. The General Data Protection Regulation (GDPR) sets a high threshold for the collection and use of data, but it also encourages organizations to have a clear framework in place for collecting, securely storing, accessing and controlling that data. Correctly labelling data points at the point of collection is therefore strongly encouraged. This should assist in reducing the amount of effort that goes into cleaning data sets and pre-processing. Inaccurate output can have deleterious effects for AI,[99] and for a well-functioning cybersecurity solution a machine needs an ordered and comprehensive threat data set. In order for this to work, the data must be clean.[100] 'Cleaning' data sets is a necessary task, and one that takes up a significant amount of time and effort;[101] figuring out how to undertake such tasks in more productive ways[102] could be key to making the development and deployment of AI more efficient whilst improving the quality of data used for machine learning.

Emphasizing a responsible approach to data collection and use also shifts the burden of protection onto those collecting data, rather than end users.[103] This incentivizes the collection and use of data in accordance with a data management architecture,[104] promoting accountability and transparency.

Much criticism has been leveled at the effects of regulation of data management, in particular on the supposed negative implications for AI research, development and implementation.[105] Such views however neglect to consider the potential benefits that such regulation can have for development of the AI industry more generally. They also fail to take into account some of the problems that follow from other divergent strategies.[106][107] A more proactive approach to advancing data regulation can set the EU apart and drive innovation in AI to compete with larger players in the market. Whilst data is continuously created and shared, ever-larger pools of data are not always better, especially when the quality or structure of the data varies greatly. More researchers are looking into what can be done with “small data” – i.e. using fewer data to train algorithms – particularly in manufacturing and the internet of things. This is where Europe, home to many industrial firms, could have an advantage.[108] Europeans are also more focused on protecting the privacy of the user. Because emergent technologies arise in response to social demand, this culture-bound constraint on data collection, in turn, would reorient the development of AI in a more social instead of consumer-marketing direction, which has been the main focus of both China and Silicon Valley.[109] The future will see AI deployed in a range of settings, from autonomous systems such as cars and cargo ships to clinical support systems, raising concerns about implementing self-learning systems where there is little control of humans in the process or knowledge of how algorithms reach decisions. By identifying strong guiding principles, and stimulating the market for AI applications, organizations are encouraged to innovate in a race to the top, potentially reducing the safety, ethical and privacy implications of increasing amounts and access to data.[110]

There are also strong incentives for ensuring the application of data regulation works in conjunction with, and not contrary to AI development. For example, the EU is looking to transport, healthcare[111] and manufacturing sectors to lead the way in AI adoption. Operators in these industries, whilst offering the most promising social and economic returns from the application of AI, are also increasingly deploying business models that depend on the availability and accessibility of real-time, high-quality data.[112] Beyond consumer-facing applications that utilize personal data, regulation also needs to account for data sharing initiatives between organizations. The idea has been floated for institutions (such as the EU, or national Member States) to act as brokers in data sharing agreements, opening up the data market to more intervention and use[113] than is currently possible. Although a market for training data in machine learning has already developed – for which personally identifiable information is immensely valuable[114] – data protections have accelerated a paradigm shift, where both public and private groups need to rethink the way in which they exchange data. High standards of protection for personal data and the concept of data portability[115] can be leveraged to incubate a platform for businesses willing to share the data they collect.[116] A wealth of data remains untapped with numerous benefits that could have an impact on AI,[117] expanding the market for training data in machine learning. The EU has made significant efforts over the past 15 years to open up public sector information and publicly funded research results for re-use, such as data generated by the EU's space programs. Other Member State initiatives, such as those of Germany,[118] Finland[119] and France[120] build on these efforts.

A healthy AI and machine learning ecosystem requires services from cloud-based platforms to store the vast amounts of data. In the interviews it became clear that there is no apparent Dutch alternative to foreign third-party cloud-based services platforms, such as IBM’s Watson. This reliance means that much of the Dutch data is leaving the Netherlands and even Europe. Some respondents believed there is an added value of EU data protection policies such as GDPR to keep Dutch data from going across the EU border, while others applaud the spirit of the regulation but find the actualization constraining. Instead, they would prefer to opt for Dutch or European services cloud-based platforms. Unfortunately, they are not finding these local or European alternatives.

Looking beyond government-facilitated or compulsory data sharing, private companies are increasingly willing to share valuable datasets.[121] Another good example of pooling and sharing data in the private sector is Google’s federated learning. In February of 2017, Google announced “a completely new, lightweight, machine learning architecture” that enables Android wearable devices to generate predictive text using users’ locally stored data, without having to copy those data to cloud servers. Federated learning allows for these local models

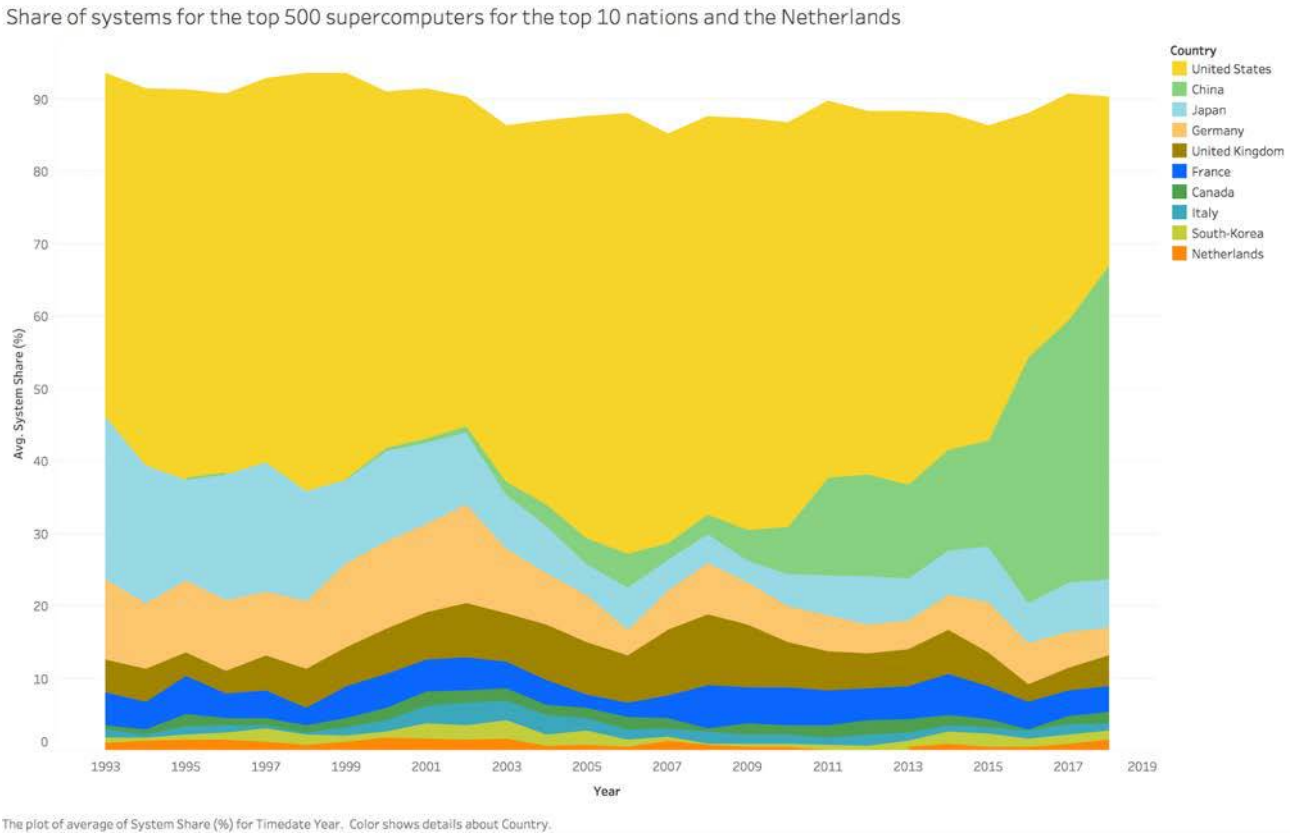
to train and be updated by a shared model stored in the cloud. Under this model, all training data are stored on users' devices and only small updates are transmitted to the cloud. This, performed across many devices, updates a shared model that can in turn be downloaded back onto users' devices.[122] Federated learning removes the need to construct a centralized database of user data for certain applications of machine learning; Google primarily touts this as a boon to user privacy and data security.

## 4.3 Computational Power

To achieve any kind of large-scale data processing, a great amount of computational power is needed. In that vein of thought, the development and subsequent race of high-performance computing (including supercomputers) is one that factors into the development of machine learning.

There is a widely perceived supercomputer race currently underway between the United States, China, Japan, and Europe. In the top ten list of supercomputers, the top two are in the United States, with Summit taking first place with 200 petaflops in power and Sierra in second place with 125 petaflops.[123] Of the top ten, five of the supercomputers originate in the United States, two from China, and one each from Germany, Switzerland and Japan.[124] While many saw the supercomputer race as purely between China, Japan and the United States, the European Union is taking steps to address its apparent lack of competitiveness by investing € 1,200,000,000 (1.2 billion)[125] in the European High Performance Computing initiative in 2018.[126] In the longer term (2021-2027), the Commission proposed to invest € 2.7 billion in the Joint Undertaking to strengthen supercomputing and data processing in Europe as part of the Digital Europe Programme.[127]

In total, the Netherlands has six supercomputers listed in the top 500 list, which together amount to 11.9 petaflops.[128] In April 2018, supercomputer producer Maxeler established an office in The Netherlands at Tech incubator YES!Delft. This branch in Delft will focus on projects that require a lot of processing power such as developing AI for creating cybersecurity solutions for self-driving cars and other solutions involving the medical field.[129] Furthermore, the Netherlands has its own supercomputer, Cartesius, which is available for large data set projects.[130] More specifically, the Netherland's Institute for Radio Astronomy has also developed a new supercomputer for the Square Kilometre Array, the most sensitive radio telescope worldwide.[131] However, it is not



just large supercomputers that take up the field. Leiden University in collaboration with IBM developed a very small supercomputer in 2017, which fits on a carrier bicycle.[132]

To power these supercomputers and many other aspects of AI, a progression in computational efficiency and hardware is needed, which is mainly driven by semiconductors. Because these pieces of technology are so essential to the operation of an AI, many countries such as China are aiming to become independent from international suppliers. In the last six months, two Chinese companies, Alibaba and Baidu, have announced plans to develop semiconductors.[133] Other companies are attempting to branch out in other ways to support the vast amount of computing power needed for AI – for example, Lightelligence is creating optical chips, which would allow computers to run faster matrix multiplications, a function key to Deep Learning in AI.[134] Japan, on the other hand, is developing semiconductors specialised for AI applications, such as learning and inference, where the goal is to improve processing speed more than accuracy. The Japanese strategy stresses the importance of reducing power consumption; miniaturization of high-performance computers and development of new architecture such as neuromorphic and quantum architecture.[135]

In the Netherlands, semiconductors are a common export product and several Dutch companies compete in a large global market. Companies like NXP Semiconductors N.V., ASML, and Nexperia are Netherlands-based but have a global reach. Other producers of semiconductors include Ampleon, BE Precision Technology, and over sixty other companies based in the Netherlands.[136] Despite this wide range of companies, the Netherlands is not leading the world in semiconductor production. Rather the industry is led by South Korea, the United States, China, Japan, Singapore and Taiwan.[137] The rise of Field-Programmable Gate Array (FPGA) chips however will increase the difference between "fabless" semiconductor companies who design chips, and the manufacturers who actually create them. The outsourcing of actual chip manufacturing to low cost producers will therefore continue to be a rising security concern.

### A Lack of Normative Leadership

In the last two years alone, more than a dozen countries have devised national strategies on AI.[138] The Netherlands still visibly lacks such an official government strategy or roadmap, although it will most likely be launched in March 2019. A call for action by the AINED raises important socio-economic, technical, policy, educational components for such a national strategy.[139] In particular, it notes that the Dutch competitive advantage resides in its experience in public-private partnerships.

In a recent analysis of 12 AI strategies, the Observer Research Foundation observed a void in normative leadership: "who is scripting the governing principles behind AI? With each passing year, deep learning technologies become more and more accessible to nations around the world. In five years it is likely that the competing nations will have comparable technical capacities. The battle then will not be technical, it will be in creating the dominant governing principles."[140]

Countries such as South Korea and Canada and organizations such as the EU have this international normative ambition in their strategies, albeit from different national angles. The Netherlands, and The Hague in particular, is well-positioned to be a credible leading actor: it has a strong reputation in general, and in the "rule of law" in particular, whilst it has a traditionally decentralized governance model that can provide the basis for a competitive advantage in the multi-stakeholder AI environment if it is met with the right level of government leadership and coordination with other stakeholders. It requires early concerted diplomatic efforts, much in the same way as the Netherlands has shown itself as a leader in establishing norms, principles and rules of the road in a similar domain: cyberspace. It should likewise consider doing this for AI as well and address this ambition in its upcoming strategy.

## Industry Leadership

Spearheaded by the high-tech, telecom and financial sector, AI is permeating through all sectors. Gartner estimates that by 2020, AI technologies will be in almost every new software product.[141]Over seventy percent of mid-market businesses are already reliant on AI and machine learning, while in companies with more than 1000 employees, over seventy-five percent are reliant on AI and machine learning.[142]

The United States remains the frontrunner because of companies like Google, Amazon and Facebook. China's industries and government are ambitious: it wants to become the world leader in AI by 2030.[143] In China alone, AI has birthed 14 unicorns – companies valued at over US$ 1 billion.[144] The Netherlands has some leading companies, such as TomTom, Booking.com and Euvision and Scyfer, but the latter three are now in American hands, which bears consequences on the ownership of data and the development of the technology. There are more than 200 AI startups in the Netherlands, but only 0.4% manages to grow into an international company, which is mainly attributed to the lack of capital injections.

The partnerships that Chinese and US-based companies may form and the national security implications of this raises some questions.

In October 2018, Chinese search giant Baidu joined the US-based Partnership on AI.[145] American companies such as Qualcomm Inc, Nvidia Corp, and IBM also have a variety of activities in China, ranging from research labs to training initiatives. However, the US is considering tightening restrictions on these partnerships.[146]

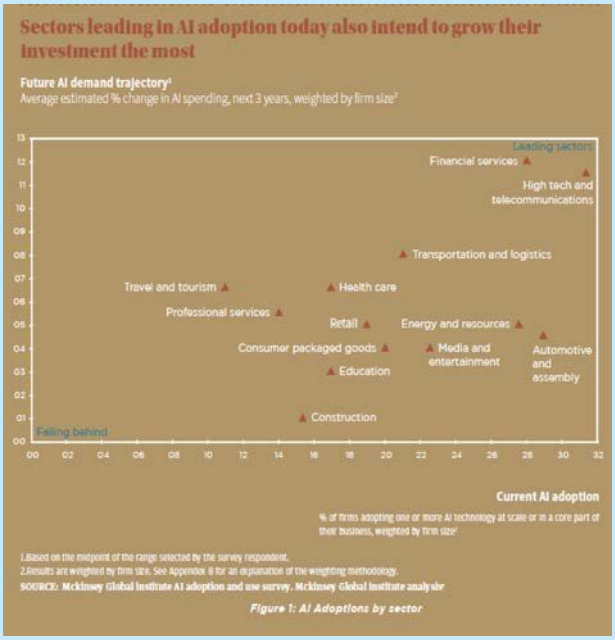


*Figure 1: AI Adoptions by sector*

# 5 – Conclusions

The word 'AI' is often equated with the murderous computer HAL from 2001: A Space Odyssey. The reality is much different, namely that "true" AI (i.e. Artificial Superintelligence or ASI) is still far off. Similarly, despite the proliferation of useful machine learning (ML) capabilities, it is important to temper expectations as to how this will impact cybersecurity.

This study has shown that a solution to a given machine learning task may, at various stages, use both supervised and unsupervised learning approaches. This also goes for cybersecurity. Undoubtedly, machine learning has already changed and will continue to change the landscape of cybersecurity. However, as much as it aids in the creation of more comprehensive cybersecurity solutions, it also becomes a tool for the offense: this includes autonomous malware and offensive cyber capabilities, such as phishing, spam, DDoS, ransomware, and spyware, as well as other applications in computational propaganda like deep fakes.

Artificial Narrow Intelligence (the capacity of an AI technology to match or exceed human ability in a specific task or domain, versus Artificial Superintelligence, where AI exceeds humans in both range and ability) will certainly automate processes that are now manual, but ANI also opens the door to new vulnerabilities. It will be important to create counter-ANI abilities to target and protect against these vulnerabilities. This new stage in cybersecurity will require analysts to also consider new kinds of ANI-enabled economic and military sabotage. The uptake of ANI has the potential to further upset the already precarious balance between state and non-state actors. Software can be reproduced at a low cost, meaning that it will become easier for smaller actors to obtain and take advantage of ANIs for malign purposes.

Machine learning is already being implemented in communication filtering, anti-virus, vulnerability scanning, malware and forensic analysis, spam-filters, phishing defense, and most notably behavior analytics and anomaly detection. It is also used to tackle the spread of computational propaganda. Commercial avenues for addressing these issues should be supported, in particular assessing and testing products according to transparent guidelines before their release on the market. Currently, machine learning's most common role is additive. It acts as a sentry, rather than a cure-all. Overall, much of the routine or repetitive security tasks are now being automated using machine learning to relieve the overburdened and understaffed security analysts, rather than replace them. This combination of combining technology and human review also applies in the countering of computational propaganda as it may take another five to ten years for an AI to proactively vet linguistically nuanced content such as hate speech with minimal to no human input. According to the Ponemon survey, the average cost of not using an AI to address cyberattacks is estimated at more than $3 million versus just over $800,000 USD if an AI is used.[147] In other words, while there remain major cybersecurity implications that machine learning cannot fully address, such as the weak social layer (i.e. human factor) and patching, it makes more sense to use it rather than go without, as it is more cost effective and addresses the acute problem of scarce and expensive cybersecurity expertise through resource optimization.

As attack systems become more dynamic in behavior, more voluminous and target a far larger attack plane, the human capacity to identify and mitigate attacks will fall short. This will warrant the development of new defensive systems that use machine learning to support the security analyst. Ultimately, adaptivity will become a core characteristic for both the offense and defense

Despite the overwhelming hype, there is real promise behind machine learning in cybersecurity. It will become the new state of art. The challenge is keeping expectations in check. Pursuing a cluster of government, private sector, and civil society (including the technical, academic, and advocacy community) actors to engage in this topic is likely to be a promising strategy for local economic development in the field. This holds true for the Netherlands in particular because of its combination of the (i) concentration of advanced Dutch and European talent and research, (ii) the regulatory focus, and (iii) associated industries.

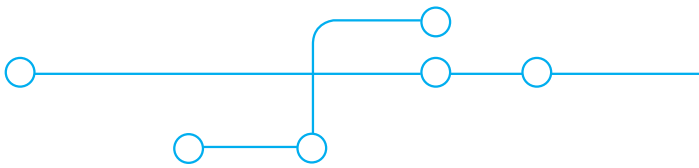
**Each component has its own set of observations and recommendations:**

### (i) Concentration of advanced Dutch and European talent and research

- **Prioritize long-term cultivation of ML talent and expertise**. While talent appears available for the moment (for example, see Google in Amsterdam, section 4.1), the global share of Dutch academic publications is shrinking and universities struggle to cope with the increasing demand of students that want to study computer science and AI. To strengthen this, i) attract (foreign and domestic) experts, teachers and scientists through financial and professional incentives, ii) create training places through cooperation with industry and iii) foster interdisciplinary education and research by creatively fusing methods from computer science, machine learning and artificial intelligence, for example by establishing a dedicated faculty department within one of the Dutch universities.
- **Cultivate and train talent in the short-term.** There is a short-term need for more expertise, especially with respect to teachers. **Retraining and in-service training** of specialists can support the lack of supply. An assessment should be made as to the level of re-training required, which skills are necessary for future employment and how these objectives can best be reached. They should be undertaken in cooperation with those stakeholders best placed to provide input, such as educational institutions and employment bodies.
- **Develop an AI research agenda**. Without such an agenda, research and commercial activities tend to be fragmented and not effective. The agenda should identify and define relevant societal challenges that AI can help tackle, take into consideration the needs of relevant sectors, such as cybersecurity, prioritize the most pressing AI issues, and aim to defragment ongoing efforts.
- **Create a community and knowledge hubs to address the lack of critical mass.** An environment in which collaboration between all of these different actors is stimulated, provides the stepping stone to build more Netherlands-based AI knowledge hubs and encourages the participation of students and potential employees in activities and initiatives related to AI research and development. The DARPA AI Cyber Grand Challenge is a good example of how government (and other stakeholders) can function as an accelerator for local technical talent and startups, from which it ultimately benefits as well. This also leads to an increased awareness and capacity of academics working on AI and machine learning to apply their expertise to cybersecurity. This can be done by by workshops or conferences with relevant stakeholders, including local industry, academia, large multinational tech giants and government.
- **Develop applied programs with accompanying budgets.** Through triple helix collaboration and across different sectors, specific AI/Cybersecurity applications should be understood, developed and valorized. (e.g., AI implications of unmanned SOCs at interorganizational levels)

### (ii) Regulatory focus potentially addressed by platforms such as The Hague Legal Delta

- **Encourage the sharing of decentralized data** on a Dutch and European scale. This could encompass the creation of platforms or networks that work towards supporting data sharing agreements. It might also include encouraging the collection and maintenance of large-scale, high-quality data sets that are subject to use by specified third parties. Such agreements should incorporate the principles of accountability and transparency.
- **Assess how untapped information could be of better use.** Many industries maintain huge data collections without approaches to make proper use of it, but the exact scope of underuse remains unclear. Research could determine which specific industries and business models are more likely to collect and consolidate large amounts of data, making assessments on whether data should be more open and how to support this.
- **Provide legal guidance, expertise and research in the field of AI**. Where there is currently a lack of national and international regulatory frameworks for AI technology, long-term research and development, particularly with a focus on ANI will make it possible to create more comprehensive regulation. A clear advisory role could be assumed by The Hague Legal Delta. An example of this role would be the Article 29 Working Party, which provided input to the development of the GDPR.



- **Develop normative leadership in AI and cyber.** Similar to international cyberpolicy. The Hague is well positioned to be a credible international actor that can easily ramp its engagement in negotiations to address the apparent lack of normative leadership. This requires the right level of government leadership and active coordination with other stakeholders. The upcoming official national AI strategy should address this ambition.

### (iii) Associated industries

- **Keep momentum in high-performance computing:** The Netherlands should keep momentum in current Dutch and European high-performance computing to not fall further behind the US and China.
- **Use wide range of domestic semiconductor companies to address security concerns.** Despite its wide range of semiconductor companies, the Netherlands is not leading in global production. The rise of Field-Programmable Gate Array chips, however, will increase the difference between "fabless" semiconductor companies who design chips, and the manufacturers who actually create them. The outsourcing of chip manufacturing to low cost producers will therefore continue to be a rising security concern, which can create opportunities for Dutch industry.
- **Use the international tech companies.** These companies have significant satellite offices in the Netherlands that increasingly work together with local companies and academics (see also the first set of recommendations).
- **Provide Dutch or European alternatives:** Large national companies are interested to find Dutch or European alternatives to foreign cloud-based services platforms, such as IBM's Watson, but find it difficult to locate them. PPP platforms such as The Hague Security Delta are well positioned to conduct a stakeholder and market analysis, linking these companies together.

These factors indicate that there are a lot of potential synergies that could entice many companies to invest in or initiate projects. It, however, requires much-needed leadership, which in these fields will not only determine who captures value in the supply chain but also help establish their "autonomy" in the AI race.

# Endnotes

## 1 – Introduction

1 "The Value of Artificial Intelligence in Cybersecurity." Ponemon Institute. July 2018, 16 https://public.dhe.ibm.com/common/ssi/ecm/41/en/41017541usen/ibm-ai-report-final-1_41017541USEN.pdf.

2 For a perspective of developments in the broader AI domain, see also HSD report "Notitie Risico-Analyse in Onzekerheid Artificial Intelligence (Kansen en Bedreigingen)", 2017, The Hague Security Delta.

## 2 – Introduction to AI and Machine Learning

3 In the field of Artificial Narrow Intelligence (ANI) - specifically in unsupervised learning, these strides are becoming much larger. One example of unsupervised learning of ANI to equal or exceed human intelligence is DeepMind's AI 'AlphaGo' that played the game 'Go' without direct help from humans, beating 18-time world champion Lee Se-dol in 2016. IBM has also created two ANIs, namely Watson, which plays Jeopardy and Deep Blue, which plays chess.

4 Bughin, Jaques et. al "Artificial Intelligence The Next Digital Frontier?" McKinsey Global Institute, June 2017, 8

5 Fischer, Sophie-Charlotte. "Artificial Intelligence: China's High-Tech Ambitions." CSS Analyses in Security Policy, 2018.

6 Domingos, P. "A few Things to Know about Machine Learning." Department of Computer Science and Engineering, University of Washington, 2012.

7 While a human may learn a lot about the game of chess on the first run through, a machine will iterate through 500 games to learn the same amount about the game and its rules. In that way, machines are actually quite inefficient.

8 Sobel, Benjamin. "Artificial Intelligence's Fair Use Crisis." Columbia Journal of Law & Arts. 41:1, 2017, 58 https://lawandarts.org/wp-content/uploads/sites/14/2017/12/41.1_Sobel-FINAL.pdf; "Machine Learning: The Power and Promise of Computers that Learn by Example." Royal Society. 2017; Polyakov, Alexander. "Machine Learning for Cybersecurity 101." Towards Data Science. October 4, 2018. https://towardsdatascience.com/machine-learning-for-cybersecurity-101-7822b802790b.

9 A learner is a program that builds a decision tree from data

10 Castle, Nikki. "Supervised vs. Unsupervised Learning." Data Science, July 2017. https://www.datascience.com/blog/supervised-and-unsupervised-machine-learning-algorithms

11 "Machine Learning: The Power and Promise of Computers that Learn by Example." Royal Society. 2017

12 Polyakov, Alexander. "Machine Learning for Cybersecurity 101." Towards Data Science. October 4, 2018. https://towardsdatascience.com/machine-learning-for-cybersecurity-101-7822b802790b.

13 World Wide Web Foundation. "Artificial Intelligence: The Road Ahead in Low and Middle-Income Countries." June, 2017. http://webfoundation.org/docs/2017/07/AI_Report_WF.pdf.

## 3 – Machine Learning in Cybersecurity

14 See "Attackers Will Exploit Artificial Intelligence (AI) Systems and Use AI to Aid Assaults" in Thompson, H. & Trilling, S. "Cyber Security Predictions: 2019 and Beyond" Symantec, November 28, 2018 https://www.symantec.com/blogs/feature-stories/cyber-security-predictions-2019-and-beyond; See also Osoba, S.A. & Welser, W. "The Risk of Artificial Intelligence to Security and the Future of Work." RAND Corporation, 2017 https://www.rand.org/content/dam/rand/pubs/perspectives/PE200/PE237/RAND_PE237.pdf. The Kenna Security's Remediation Gap report found that automated attacks are on the rise: There were over 1.2 billion successful exploits witnessed in 2015, compared to 220 million successful exploits witnessed in 2013 and 2014 combined – an increase of 445 percent. Unlike more widely publicized advanced persistent threats, these non-targeted attacks pose a different challenge for security organizations. Rather than targeting a specific company, attackers attempt to exfiltrate valuable data from as many companies as possible, relying on automated tools and techniques to scale their attacks and exploit commonly found vulnerabilities.The recent discovery of the Heartbleed vulnerability in the OpenSSL brought this to the forefront as a threat that exploited multiple targets at once. See "The Remediation Gap: Why Companies Are Losing the Battle Against Non-Targeted Attacks" Kenna, September 2015 http://pages.kennasecurity.com/rs/958-PRK-049/images/Kenna-NonTargetedAttacksReport.pdf

15 Allen, G. & Chan, T. "Artificial Intelligence and National Security." Belfer Centre Study July 2017, 33

16 Klimburg, Alexander "The Darkening Web: The War for Cyberspace" Penguin Books, 2017

17 Looking more closely at the ICT sector that shapes the digital environment of people and companies alike, there's a worrisome incentive structure that fuels risk exposure when

commercial incentives to rush to the market override the incentive to create secure products.

18 "A key function of artificial agents (both informational and cyberphysical artificial agents) is the efficient manipulation of information. Thus, artificial agents may be particularly suited to information warfare and cybersecurity applications. Augmenting Internet of Things (IoT)–targeting malware like Mirai (Newman, 2017) with intelligence can vastly improve the strategic potential of malware." Supra (Note 10), Osoba, S.A. & Welser, W., 5-6

19 The Mirai botnet was responsible for the massive network outage on October 21, 2016, that disrupted large-scale operations like Twitter, GitHub, and Netflix. See Newman, Lily H., "The Botnet that Broke the Internet isn't going Away." Wired, 2016. https://www.wired.com/2016/12/botnet-broke-internet-isnt-going-away/

20 Kubovič, O., Košinár, P., Jánošík, J. "Can Artificial Intelligence Power Future Malware?" ESET White Paper, 2018 https://www.welivesecurity.com/wp-content/uploads/2018/08/Can_AI_Power_Future_Malware.pdf

21 Ibid.

22 One example of this is Data Visor's UML Engine, which extracts behaviors and activities from various forms of analysis and then clusters results in patterns from large-scale data sources. By extracting more forms of behaviors and activities, AI systems will become more creative as they will have more functions to choose from. Furthermore, with more large-scale data sets, AI systems will have the ability to learn from more innovative forms of analysis. Peng, T. & Sarazen, M. "DataVisor Uses Unsupervised Learning to Combat Online Fraud." Synced, 2018 https://syncedreview.com/2018/03/01/datavisor-uses-unsupervised-learning-to-combat-online-fraud/

23 UK-based Darktrace offers AI-powered autonomous responses to cyber attacks, visualizes, clusters, and mitigates threats. Darktrace noticed the spread of WannaCry ransomware, as well as a variety of other use-cases such as automated credential threats, targeted biometric attacks and compromised video conferencing systems. "Detecting Ransomware: How Darktrace identifies ransomware before it spreads." DarkTrace, 2018. https://www.darktrace.com/en/technology/

24 Beaugnon, A. & Chifflier, P. "Machine Learning for Computer Security Detection Systems: Practical Feedback and Solutions." French National Cybersecurity Agency (ANSSI). https://www.ssi.gouv.fr/uploads/2018/11/machine-learning-for-computer-security-abeaugnon-pchifflier-anssi-.pdf

25 Andrea, P. et al. "Detection of Adversarial Training Examples in Poisoning Attacks through Anomaly Detection" Cornell University Library. February 8, 2018. https://arxiv.org/pdf/1802.03041.pdf

26 An average phishing attacker will bypass an AI-based detection system 0.3 percent of the time, but by using AI this 'attacker' is able to bypass the system more than 15 percent of the time. See Bahnsen, A.C. et al. "DeepPhish: Simulating Malicious A.I" Cyxtera Technologies. 2018. https://albahnsen.com/wp-content/uploads/2018/05/deepphish-simulating-malicious-ai_submitted.pdf

27 Roth, Phil. "Introducing Ember: An Open Source Classifier And Dataset". Endgame. April 16, 2018. https://www.endgame.com/blog/technical-blog/introducing-ember-open-source-classifier-and-dataset

28 "There are good reasons that the security industry doesn't have as many open data sets," Endgame's Roth says. "These kinds of data might have personally identifying information or might give attackers information about what a company's network architecture looks like. It took a lot of work to sanitize the EMBER dataset, but my hope is to spur more research and get defenders to work together." Newman, Lily H. "AI can help cybersecurity - if it can fight through the hype". Wired. 29 April 2018. https://www.wired.com/story/ai-machine-learning-cybersecurity/

29 Newman, Lily H. "AI Can Help Cybersecurity-If It Can Fight Through the Hype." Wired, 2018. https://www.wired.com/story/ai-machine-learning-cybersecurity/

30 Madrid-based Openbank, for example, utilizes unsupervised machine learning algorithms to detect fraud and money laundering. Supra (Note 18), Peng, T. & Sarazen https://syncedreview.com/2018/03/01/datavisor-uses-unsupervised-learning-to-combat-online-fraud/

31 Kanal, E. "Machine Learning in Cybersecurity." Software Engineering Institute of Carnegie Mellon University, 2017 https://insights.sei.cmu.edu/sei_blog/2017/06/machine-learning-in-cybersecurity.html

32 Tretyakov, Konstantin. "Machine Learning Techniques in Spam Filtering." Institute of Computer Science, University of Tartu, 2004. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1.3483&rep=rep1&type=pdf

33 Klimburg, Alexander. "The Darkening Web: The War for Cyberspace" Penguin Books, 2017

34 Newman, Lily H. "AI can help cybersecurity - if it can fight through the hype" Wired, April 29, 2018 https://www.wired.com/story/ai-machine-learning-cybersecurity/

35 "The Value of Artificial Intelligence in Cybersecurity." Ponemon Institute. July 2018, 16 https://public.dhe.ibm.com/common/ssi/ecm/41/en/41017541usen/ibm-ai-report-final-1_41017541USEN.pdf.

36 Greenberg, Andy. "Hacker Lexicon: What is Fuzzing?" Wired, 2016. https://www.wired.com/2016/06/hacker-lexicon-fuzzing/

37 Johnson, Ann. "Application fuzzing in the era of Machine Learning and AI." Microsoft Secure, 2018. https://cloudblogs.microsoft.com/microsoftsecure/2018/01/03/

application-fuzzing-in-the-era-of-machine-learning-and-ai/

38 Metz, Cade “Google's Training Its AI To Be Android's Security Guard” Wired, February 6, 201 https://www.wired.com/2016/06/googles-android-security-team-turns-machine-learning/. Baidu uses deep neural networks to identify malware. So do security startups such as Deep Instinct, Cylance, Darktrace, Jask and Harvest.ai. Just as a neural net can identify the particular characteristics of a photo, it can recognize a malicious software application – or a bit of flawed operating system code that exposes your phone to malicious hackers.

39 Robertson, J. et. al. “Darkweb Cyber Threat Intelligence Mining” Cambridge University Press, 2017.

40 Chen, Hsinchun. “Dark Web: Exploring and Data Mining the Dark Side of the Web”, Springer, 2012, 326

41 Nunes, Eric. et. al. “Darknet and Deepnet Mining for Proactive Cybersecurity Threat Intelligence” Arizona State University, 2016. https://arxiv.org/pdf/1607.08583.pdf

42 Newman, Lily H. “AI can help cybersecurity - if it can fight through the hype” Wired, April 29, 2018 https://www.wired.com/story/ai-machine-learning-cybersecurity/

43 “Symantec Endpoint Detection and Response: On-premises and Cloud-based EDR Solution.” Symantec, 2018 https://www.symantec.com/content/dam/symantec/docs/data-sheets/endpoint-detection-and-response-atp-endpoint-en.pdf

44 MIT’s Computer Science and Artificial intelligence Laboratory has worked with startup PatternEx to construct a ML system that can review more than 3.6 billion lines of log files each day for suspicious activity, autonomously learning from and countering cyber attacks as they evolve in real time. “Artificial Intelligence in Defence and Security Industry.” AI.Business, 2016 http://ai.business/2016/06/21/artificial-intelligence-in-defence-and-security-industry/

45 “The Value of Artificial Intelligence in Cybersecurity.” Ponemon Institute, 2018, 2

46 For example, many intelligent systems may be able to work together and delegate tasks amongst themselves, making themselves more productive than if a human or several humans were coordinating them. Particularly the question of response time will be drastically lessened because the AI systems will together be able to learn from their mistakes and change their approach much faster than a human would be able to reprogram an entire system.

47 Nearly all large ICT companies have implemented Machine Learning for this reason. IBM dubbed their machine learning sidekick Watson on which they rely for these "knowledge consolidation" tasks and other areas of threat detection. Watson is a cloud-based version of the company’s cognitive technology trained on the language of cybersecurity and large corpus of incidents to find and discover patterns for cyberattacks and threats that could otherwise be missed by human being. Mina, George. “AI is the future of cybersecurity – How Watson helps detect threats faster and better protect your organization.” IBM, 2017 https://www.ibm.com/blogs/watson/2017/08/ai-is-the-future-of-cybersecurity-how-watson-helps-detect-threats-faster-and-better-protect-your-organization/)

48 Beaugnon, A. & Chifflie, P. “Machine Learning for Computer Security Detection Systems: Practical Feedback and Solutions.” French National Cybersecurity Agency, 2018. https://www.ssi.gouv.fr/uploads/2018/11/machine-learning-for-computer-security-abeaugnon-pchifflier-anssi-.pdf

49 “Training Catalogue 2018”. NATO Cooperative Cyber Defence Centre of Excellence, 2017. https://ccdcoe.org/sites/default/files/documents/CCDCOE_Training_Catalogue_2018.pdf

50 Cade Metz. “Google’s Training Its AI to Be Android’s Security Guard.” Wired. 2016. https://www.wired.com/2016/06/googles-android-security-team-turns-machine-learning/

51 Three 2017 examples illustrate the problem. Equifax was hacked because it didn’t install a patch for its Apache web server that had been available two months previously. The WannaCry malware was a worldwide scourge, but it only affected unpatched Windows systems. The Amnesia IoT botnet made use of a vulnerability in digital video recorders that had been disclosed and fixed a year earlier, but existing machines couldn’t be patched.
Daemmrich, Arthur. “AI and the Challenge of Cybersecurity.” *Lemelson Center for the Study of Invention and Innovation*, 2017. http://invention.si.edu/ai-and-challenge-cybersecurity;
Schneier, Bruce . “Click Here to Kill Everybody: Security and Survival in a Hyper-connected World” *W. W. Norton & Company*, 2018.

52 In 2014, DARPA announced a two-year competition to design and build computer systems that would block attacks or find and isolate malicious code. Specifically, DARPA sought to underwrite the development of an “architecture” for an artificial intelligence machine, which they termed a “cyber reasoning system.” Systems had to integrate autonomous analysis, autonomous patching, autonomous vulnerability scanning, autonomous service resiliency, and autonomous network defense. To speed development, DARPA announced that qualified systems would go head-to-head in a series of rounds, and finalists would meet to compete for $4 million in prizes, with the top system winning $2 million. http://archive.darpa.mil/cybergrandchallenge_competitorsite/Files/CGC_Technical_Paper_Guidelines.pdf

53 For more information about the DARPA Cyber Grand Challenge visit https://www.darpa.mil/program/cyber-grand-challenge.

54 Bing, Chris. “The Tech behind the DARPA Grand Challenge winner will now be used by the Pentagon.” Cyberscoop. August 2017. https://www.cyberscoop.com/mayhem-darpa-cyber-grand-challenge-dod-voltron/

55 Pursuant to the findings of the independent High level group on fake news and online disinformation, by disinformation we refer to “all forms of false, inaccurate, or misleading information designed, presented and promoted to intentionally cause public harm or for profit”. See “A multi-dimensional approach to disinformation.” European Commission, March 2018, 35

56 Wooley, Samuel C., and Philip N. Howard. “Computational Propaganda Worldwide: Executive Summary.” Computational Propaganda Research Project. University of Oxford, 2017. http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/89/2017/06/Casestudies-ExecutiveSummary.pdf

57 Wooley, Samuel C., and Philip N. Howard. “Computational Propaganda Worldwide: Executive Summary.” Computational Propaganda Research Project. University of Oxford, 2017. http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/89/2017/06/Casestudies-ExecutiveSummary.pdf.

58 Nimmo, Ben. “Anatomy of an Info-War: How Russia’s Propaganda Machine Works, and How to Counter It.” StopFake.org, May 19, 2015. https://www.stopfake.org/en/anatomy-of-an-info-war-how-russia-s-propaganda-machine-works-and-how-to-counter-it/.

59 Rosenbach & Mansted. “Can Democracy Survive in the Information Age?” 2018. Harvard Belfer Centre, 12.

60 Ibid.

61 Solon, Olivia. “Facebook removes 652 fake accounts and pages meant to influence world politics.” The Guardian, 2018. https://www.theguardian.com/technology/2018/aug/21/facebook-pages-accounts-removed-russia-iran

62 Gadde, Vijaya. “Confidence in Follower Counts.” Twitter, July 11, 2018. https://blog.twitter.com/official/en_us/topics/company/2018/Confidence-in-Follower-Counts.html

63 “Strengthening Network & Information Security & Protecting against Online Disinformation.” ENISA, April 2018.

64 Zuckerberg, Mark. “A Blueprint for Content Governance and Enforcement.” Facebook, November 15, 2018. https://www.facebook.com/notes/mark-zuckerberg/a-blueprint-for-content-governance-and-enforcement/

65 “Fake News.” Fake News Detector. https://fakenewsdetector.org/en

66 Murison, Malek. “Facebook using machine learning to fight fake news.” Internet of Business. 2018. https://internetofbusiness.com/facebook-machine-learning-fake-news/

67 Tett, Gillan. “Facebook’s fight against fake news.” Financial Times. 2018. https://www.ft.com/content/4feb7268-7f1c-11e8-bc55-50daf11b720d

68 Supra (Note 45), Rosenbach & Mansted, 12.

69 Ibid,14.

70 Ibid.

71 Chesney, R. & Citron, D. “Disinformation on Steroids: The Threat of Deep Fakes” Council on Foreign Relations, 2018 https://www.cfr.org/report/deep-fake-disinformation-steroids

## 4 – The Bottlenecks of Machine Learning

72 Input data bias: The decision-making capabilities of a machine learning algorithm depend on its training data. Bias is a learner’s tendency to consistently learn the same wrong thing. Biased data could lead an algorithm to perpetuate and reinforce those biases. Easy access to less biased training data could mitigate these problems. In fact, the relative inaccessibility of low-bias training data is a problem. (Levendowski, supra note 98, at 6, 19)

73 Machine learning models sometimes reconstruct idiosyncrasies of input data instead of reflecting underlying trends about those data. In technical terms, these models are “overfitted”. They are undesirable in a predictive context because they capture noise rather than signals. Domingos, P. “A Few Useful Things to Know about Machine Learning” University of Washington. https://homes.cs.washington.edu/~pedrod/papers/cacm12.pdf

74 Overall, these three components reoccur in many AI strategies. See for example the translation of the Chinese strategy: “Focusing on the urgent need to raise China's international competitiveness in AI, next-generation AI key general technology R&D and deployment should make algorithms the core; data and hardware the foundation.” Webster, G. et. al. “Full Translation: China's 'New Generation Artificial Intelligence Development Plan.'” New America, August 1, 2017. https://www.newamerica.org/cybersecurity-initiative/digichina/blog/full-translation-chinas-new-generation-artificial-intelligence-development-plan-2017/

Not all AI strategies place the same emphasis on solely algorithms, data and computational power. The US National AI R&D Plan, for example, places emphasis on those three, but also notes that the government must also consider the ethical, safety, and societal implications of AI and create benchmarks. United States National Science and Technology Council “The National Artificial Intelligence Research and Development Strategic Plan.” *Networking and Information Technology and Research and Development Subcommittee*. October 2016. https://www.nitrd.gov/PUBS/national_ai_rd_strategic_plan.pdf

75 China’s 2030 plan envisions building a $1 trillion AI industry, and investors poured $4.5 billion into more than 200

Chinese AI companies between 2012 and 2017. Barhat V. “China is determined to steal A.I. crown from US and nothing, not even a trade war, will stop it” CNBC. May 4, 2018. https://www.cnbc.com/2018/05/04/china-aims-to-steal-us-a-i-crown-and-not-even-trade-war-will-stop-it.html

76 “Foundations of Computer Science/Algorithm design.” WikiBooks. December 20, 2017. https://en.wikibooks.org/wiki/Foundations_of_Computer_Science/Algorithm_Design

77 BBC. “What is an algorithm?” BBC Bitesize. 2018. https://www.bbc.com/bitesize/articles/z3whpv4

78 https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/

79 Algorithmic teams are collaborative groups that work on the implementation of algorithms and their application in problem-solving, in particular algorithm development, mathematical modelling and rapid prototyping

80 Shead, Sam. “Google Brain Has Expanded To Amsterdam”. Forbes. July 10, 2018. https://www.forbes.com/sites/samshead/2018/07/10/google-brain-has-expanded-to-amsterdam/#69ff11d426f4

81 The number of freshmen in computer science and AI increased over 10% in the past five years. The total number of freshmen over that period increased with 4% per year. “AI voor Nederland: vergroten, versnellen en verbinden.” Dutch Digital Delta 2018. https://dutchdigitaldelta.nl/uploads/pdf/Rapport-AI-voor-Nederland_181106_105304.pdf

82 Kamphuis, Bart. “Universiteiten kunnen belangstelling voor kunstmatige intelligentie niet aan. NOS. 16 July, 2018. https://nos.nl/artikel/2241732-universiteiten-kunnen-belangstelling-voor-kunstmatige-intelligentie-niet-aan.html

83 UK has invested in 1000 new PhD positions, and MIT invested 1 bn in a new AI university

84 Saran, S., Natarajan, N. & Srikumar, M. “In Pursuit of Autonomy: AI and National Strategies”. Observer Research Foundation; https://www.orfonline.org/wp-content/uploads/2018/11/Ai_Book.pdf and “AI voor Nederland: vergroten, versnellen en verbinden.” 2018. https://www.vno-ncw.nl/sites/default/files/aivnl_20181106_0.pdf on p.18.

Sack, Jörg-Rüdiger. “To: Members of the Canadian academic computer science research community.” *NSERC Computer Science Liaison Committee*, April 5, 2013. http://nwo.h5mag.com/nwo/edit-jaaroverzicht_2016_preview01/wiskunde_van_de_woestijn_en_nl_ict_wereldtop/7438/Canada_studie.pdf // https://www.elsevier.com/__data/assets/pdf_file/0010/823654/ACAD_RL_RE_AI-Report_WEB.pdf

85 “For A Meaningful Artificial Intelligence: Towards A French And European Strategy.” Mission assigned by the Prime Minister Édouard Philippe. March 29, 2018. https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ ENG-VF.pdf

86 Dean, Jeff. “The Google Brain Team’s Approach to Research.” Google AI Blog. September 13, 2017. https://ai.googleblog.com/2017/09/the-google-brain-teams-approach-to.html

87 Silicon Canals Editorial Team. “Ahold Delhaize and TU Delft join forces: 3 ways they plan to transform Dutch retail industry through robotics.” Silicon Canals. November 22, 2018. https://siliconcanals.nl/news/startups/ahold-delhaize-and-tu-delft-join-forces-3-ways-they-plan-to-transform-dutch-retail-industry-through-robotics/

88 “AI voor Nederland: vergroten, versnellen en verbinden.” Dutch Digital Delta, 2018. https://dutchdigitaldelta.nl/uploads/pdf/Rapport-AI-voor-Nederland_181106_105304.pdf

89 Lattanzi, Silvio. “A Summary of the Google Zürich Algorithms & Optimization Workshop”. Google AI Blog. February 23, 2018. https://ai.googleblog.com/2018/02/a-summary-of-google-zurich-algorithms.html

90 The Hague Security Delta, Ïnternational Cyber Security Summer School https://www.thehaguesecuritydelta.com/talent/international-cyber-security-summer-school; National Cyber Security Summer School https://www.ncs3.nl/

91 Murphy, Kevin. “Machine Learning a Probabilistic Perspective” MIT Press Cambridge, Massachusetts. https://www.cs.ubc.ca/~murphyk/MLbook/pml-intro-22may12.pdf

92 ‘Data quality’ can encompass several elements, such as accuracy/correctness, completeness, relevance, reliability and useability. See Lee, Y & Strong, D. “Knowing-Why About Data Processes and Data Quality” Journal of Management Information Systems Vol. 20, No. 3, 2004, 13-39

93 The regulatory approach refers to the increased desire to regulate and set guidelines for the collection of use of data within the European Union. This is contrary to others approaches taken by countries such as the United States and China, where the use of data for AI purposes is less constricted by legislation. For the EU approach to the data economy, see “Delivering an area of freedom, security and justice for Europe's citizens
Action Plan Implementing the Stockholm Programme” European Commission, April 20, 2010 https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2010:0171:FIN:en:PDF; “A Digital Agenda for Europe” European Commission, May 5, 2010 https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52010DC0245&from=en; On strategic competition in AI, see Horowitz, M. et. al. “Strategic Competition in an Era of Artificial Intelligence” Center for New American Security, July 25, 2018 https://www.cnas.org/publications/reports/strategic-competition-in-an-era-of-artificial-intelligence

94 We have already began to see some push-back and

responses to this. See Finley, K. “Tim Berners-lee, Inventor Of The Web, Plots A Radical Overhaul Of His Creation” Wired, April 4, 2017. https://www.wired.com/2017/04/tim-berners-lee-inventor-web-plots-radical-overhaul-creation/. See also Solid, https://solid.mit.edu/

95 Auerswald, P. “The Code Economy: A Forty-thousand-year History” Oxford University Press. 2017. 198. See also Ezrachi, A. & Stucke, M. “Virtual Competition - The Promise and Perils of an Algorithm-Driven Economy” Harvard University Press. 2016

96 These data sources can include: Data from sensors during drilling, exploration production or transportation; Traditional enterprise data from operational systems; Social media; Web browsing patterns (on informational websites); Demographic data; and historical oil and gas exploration, delivery and pricing data. See Oracle Enterprise Architecture White Paper, “Improving Oil and Gas Performance with Big Data: Architect’s Guide and Reference Architecture Introduction” Oracle 2015 http://www.oracle.com/us/technologies/big-data/big-data-oil-gas-2515144.pdf

97 Companies are struggling to alleviate these bottlenecks, in large part due to a lack of open standards that is limiting the flow of data at the aggregate stage and thus analysis. Slaughter, A., Bean, G. & Mittal, A. “Connected Barrels – Transforming Oil and Gas Strategies with the Internet of Things” Deloitte University Press 2015, 7

98 Emerging digital technologies must meet the essential health and safety requirements laid down in the applicable EU safety legislation, such as Directive (EC) 2006/42/EC on machinery (which is the relevant safety legislation for robots) and Directive 2014/53/EU on radio equipment. Emerging digital technologies are also being incorporated in other products, therefore other EU legislative instruments also apply. The European Standardization Organizations are also working on standards for "combined" products, i.e. where several pieces of EU safety legislation apply. See “Staff Working Document on Liability for Emerging Digital Technologies” EU Commission. 137 Final. 2018, 5.

99 Associated problems in this regard have to do with input data bias (Levendowski, supra note 98, at 6, 19), overfitting & dimensionality (Domingos, P. “A Few Useful Things to know About Machine Learning.” Communications of the ACM Volume 55, Issue 10. October 2012 https://homes.cs.washington.edu/~pedrod/papers/cacm12.pdf) and the fact that correlation does not equal causation (Pearl, J. “Causality: Models, Reasoning, and Inference”. Cambridge University Press, 2000)

100 For instance, a machine cannot learn to differentiate between a spam email and a regular email unless it has good examples of what email is what.

101 Pre-processing is more laborious than testing the training sets and applying the algorithms to generate an actionable output. Some estimates claim that eighty percent of a data scientist’s time is spent cleaning data. See Oliver, J. “Is Big Data Enough for Machine Learning in Cybersecurity?” Trend Micro. July 19, 2018. https://www.trendmicro.com/vinfo/us/security/news/security-technology/is-big-data-big-enough-for-machine-learning-in-cybersecurity

102 Because data will need to be specific to the application, it makes sense to only clean data where necessary and on a case-by-case basis. Further study is needed to improve the efficiency of data cleaning techniques, to create methods for discovering inconsistencies and anomalies in the data, and to develop approaches for incorporating human feedback. Researchers need to explore new methods to enable data and associated metadata to be mined simultaneously. See: the U.S. National Artificial Intelligence Research and Development Strategic Plan, 17 https://www.nitrd.gov/PUBS/national_ai_rd_strategic_plan.pdf

103 See Articles 5, 6 and 15-22, “Regulation (EU) 2016/679 of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data” (GDPR)

104 “Enterprise Data Sovereignty: If You Love Your Data, Set It Free.” Deloitte. December 5, 2017. www2.deloitte.com/insights/us/en/focus/tech-trends/2018/data-sovereignty-management.html

105 The EU is home to very few multinational technology companies, whilst it also suffers from a structural disadvantage - a lack of scale. Benefiting from huge, homogeneous home markets, America’s and China’s tech giants have a surfeit of the most vital resource for AI: data. See: “Can the EU become another AI superpower?” The Economist. September 20, 2018 https://www.economist.com/business/2018/09/20/can-the-eu-become-another-ai-superpower

106 Technology companies in the United States have lined up to praise certain elements of the GDPR, calling for the U.S. to develop its own approach in this area. See: Faitelson, Yaki. “Data Privacy Disruption in the U.S.” Forbes. December 12, 2018. https://www.forbes.com/sites/forbestechcouncil/2018/12/12/data-privacy-disruption-in-the-u-s/

107 Simply having access to large amounts of data, or gaining subsidies, are not the only solutions to innovating in AI. Data has to be of sufficient quality, and without any oversight there may be a lack of control over input bias. Companies that benefit from large amounts of funding are not always competitive in other areas. It remains difficult for developers and researchers to “get access to this big wealth of data sitting in government agencies.”

Chinese AI researchers face not only lack of access to domestic data but also government "firewalls" that prevent them from tapping into international data sets. See: Leopold, G. "China Ahead on AI Strategy, Behind on Data Access." Datanami. July 30, 2018. https://www.datanami.com/2018/07/30/china-ahead-on-ai-strategy-behind-on-data-access/

108 "Can the EU become another AI superpower?" The Economist. September 20, 2018. https://www.economist.com/business/2018/09/20/can-the-eu-become-another-ai-superpower

109 Berggruen, N. & Gardels, N. "A wakeup call for Europe" The Washington Post. September 27, 2018 https://www.washingtonpost.com/news/theworldpost/wp/2018/09/27/europe/?noredirect=on&utm_term=.6d0fea47e1ff

110 Morrison, Gordon. "To realise the full potential of AI, we must regulate it differently." World Economic Forum. 17 September 2018. https://www.weforum.org/agenda/2018/09/the-potential-and-pitfalls-of-ai-artificial-intelligence/

111 Sample, Ian. "It's going to create a revolution: How AI is transforming the NHS" The Guardian. July 4, 2018. https://www.theguardian.com/technology/2018/jul/04/its-going-create-revolution-how-ai-transforming-nhs

112 Behr, Alyson. "More than an auto-pilot, AI charts its course in aviation" Arstechnica, December 5, 2018 https://arstechnica.com/information-technology/2018/12/unite-day1-1/

113 Intervention could involve the creation of platforms or institutions that assist in making data more readily available for use by others. Currently the term "data brokers" refers to commercial entities that collect, maintain or assemble information in order to sell it on to third parties. In this context, the EU or its Member States (or other public or certified institutions) would be brought in to act as intermediaries, encouraging organizations to share data more freely and providing guarantees to those that share data.

114 Duhigg, C. "How Companies Learn Your Secrets." New York Times. February 16, 2012 https://perma.cc/P29W-UD85; "The World's Most Valuable Resource Is No Longer Oil, But Data." The Economist. May 6, 2017 https://perma.cc/JZN3-YXHJ.

115 The right to data portability, Article 20 GDPR

116 "In Pursuit of Autonomy: AI and National Strategies". Observer Research Foundation. 18 https://www.orfonline.org/wp-content/uploads/2018/11/Ai_Book.pdf

117 Metadata in this context refers to any additional information that helps data consumers better understand the meaning of data, its structure, and to clarify other issues (such as who generated the data, the data quality, and data access methods). See W3C, Data on the Web Best Practices, W3C Recommendation, 31 January 2017 https://www.w3.org/TR/dwbp/#metadata
Personally identifiable information is not the only available data – there are other types of data that might be useful across the board, and metadata, machine data, industrial data and research data are forms of information that are not yet fully utilized (meaning they have immense value, although the applications in which they might be used are not necessarily known)

118 Germany brings a business model-led approach to data sharing where new businesses operate on cooperative networks with other businesses. This data sharing capability is expected to traverse businesses of varying sizes across different sectors. It aims to increase the quantity of data available for analytics without compromising on rights of data holders. It aims to open up data held by the public sector and academic communities for both commercial and non-commercial use. To enable this, Germany will analyse to what extent existing rules on access to data need to be revised. See Germany's strategy here: https://www.plattform-lernende-systeme.de/home-en.html

119 See Finland's strategy here: https://www.tekoalyaika.fi/

120 France has outlined a clear strategy, building on the Villani report (Note 62.) The data sharing platform or exchange would be established with the support of French data protection authority and the Direction Généralé des Entreprises who can guide the industry on best practices and standard contracts.

121 Baidu recently released a gigantic dataset linked to autonomous vehicles. See http://apolloscape.auto/. Dar, Pranav "Baidu has Released a Gigantic Self-Driving Dataset named ApolloScape" Analytics Vidhya, March 19, 2018 https://www.analyticsvidhya.com/blog/2018/03/baidu-apollo-released-gigantic-self-driving-dataset/

122 Ravi, S. "On-Device Machine Intelligence." Google Research Blog. February 9, 2017. https://perma.cc/WQ8L-WS5D; McMahan, B. & Ramage, D. "Federated Learning: Collaborative machine learning without Centralized Training Data" Google Research Blog. April 6, 2017 https://perma.cc/XVA2-J96J. For a more technical discussion of federated learning, see McMahan, B. et al., "Communication-Efficient Learning of Deep Networks from Decentralized Data." ARXIV. February 28, 2017 https://perma.cc/P7GX-LXB9.

123 A petaflop is used to measure the processing speed of a computer. 1 Petaflop is a quadrillion (thousand trillion) floating point operations per second (FLOPS) or a thousand teraflops. In June 2008, IBM created the first supercomputer to break what was called "the petaflop barrier."

124 “November 2018.” Top 500. https://www.top500.org/lists/2018/11/

125 We use billion here to mean a thousand million (1.000.000.000)

126 “Council backs Commission's plans to invest €1 billion in world-class European supercomputers.” European Commission. 28 September, 2018. http://europa.eu/rapid/press-release_IP-18-5864_en.htm

127 Ibid. and “EU budget: Commission proposes €9.2 billion investment in first ever digital programme.” European Commission. 6 June, 2018 http://europa.eu/rapid/press-release_IP-18-4043_en.htm

128 For a complete overview, visit: https://www.top500.org/statistics/details/country/NL

129 HSD Foundation. “Supercomputer Manufacturer Maxler Chooses the Netherlands.” The Hague Security Delta. April 10, 2018. https://www.thehaguesecuritydelta.com/news/newsitem/1062-supercomputer-manufacturer-maxeler-chooses-the-netherlands

130 “Dutch National Supercomputer.” Surf. August 27, 2018. https://www.surf.nl/en/services-and-products/dutch-national-supercomputer/index.html

131 Nijman, I. “Astron develops heart of new supercomputer.” Astron. June 4, 2018. http://www.astron.nl/astron-develops-heart-new-supercomputer

132 “Netherlands Smallest Supercomputer” University of Leiden, April 5, 2017 https://www.universiteitleiden.nl/en/news/2017/04/dutch-smallest-supercomputer

133 Triolo, Paul & Webster, Graham. “China's Efforts to Build the Semiconductors at AI's Core.” New America. December 7, 2018. https://www.newamerica.org/cybersecurity-initiative/digichina/blog/chinas-efforts-to-build-the-semiconductors-at-ais-core/

134 Knight, Will. “Making AI algorithms crazy fast using chips powered by light.” Technology Review. November 29, 2018. https://www.technologyreview.com/s/612449/making-ai-algorithms-crazy-fast-using-chips-powered-by-light/

135 “Artificial Intelligence Technology Strategy - Report of Strategic Council for AI Technology”, Strategic Council for AI Technology. March 31, 2017. http://www.nedo.go.jp/content/100865202.pdf; Observer Research Foundation, AI Report, 18 https://www.orfonline.org/wp-content/uploads/2018/11/Ai_Book.pdf

136 https://bcsemi.nl/network-members/#168-14

137 Export of discrete semiconductors as of 2016, by United Nations Harmonized Commodity Description and Coding Systems 4

138 For an analysis of 12 recent AI strategies, see “In Pursuit of Autonomy: AI and National Strategies”. Observer Research Foundation. 18 https://www.orfonline.org/wp-content/uploads/2018/11/Ai_Book.pdf

139 AINED, “AI voor Nederland - Vergroten, Versnellen en Verbinden”. October 2018. https://dutchdigitaldelta.nl/nieuws/bedrijven-en-wetenschappers-waaronder-topteam-ict-publiceren-1e-aanzet-voor-een-nationale-ai-strategie

140 Supra Note 142

141 Moore, Susan “Gartner Says AI Technologies Will Be in Almost Every New Software Product by 2020.” Gartner. July 18, 2017. https://www.gartner.com/en/newsroom/press-releases/2017-07-18-gartner-says-ai-technologies-will-be-in-almost-every-new-software-product-by-2020

142 Cybersecurity Special Report “Small and Mighty: How small and Midmarket Businesses Can Fortify Their Defenses Against today’s Threats” Cisco, 2018, 7 https://www.cisco.com/c/dam/en/us/products/collateral/security/small-mighty-threat.pdf

143 China’s current strategy is less a moonshot, but rather a wishlist of desiderata and objectives with little insights on how these are achieved other than by investing money in it. See: Webster, G. et. al. “Full Translation: China's 'New Generation Artificial Intelligence Development Plan.'” New America, August 1, 2017. https://www.newamerica.org/cybersecurity-initiative/digichina/blog/full-translation-chinas-new-generation-artificial-intelligence-development-plan-2017/

144 “Nina Xiang, “China’s AI Industry Has Given Birth To 14 Unicorns: Is It A Bubble Waiting To Burst?,” Forbes, October 5, 2018, https://www.forbes.com/sites/ninaxiang/2018/10/05/chinas-ai-industry-has-given-birth-to-14-unicorns-is-it-abubble-waiting-to-pop/#7f3cfd8b46c3

145 Gershgorn, Dave. “Chinese search giant Baidu says it cares about what AI does to society.” Quartz. October 17, 2018. https://qz.com/1427340/baidu-is-joining-an-american-artificial-intelligence-partnership/

146 Qing, Koh Gui. “Exclusive: U.S. considers tightening grip on China ties to corporate America.” Reuters. April 27, 2018. https://www.reuters.com/article/us-usa-trade-china-regulations-exclusive/exclusive-us-considers-tightening-grip-on-china-ties-to-corporate-america-idUSKBN1HY0FJ

## 5 – Conclusions

147 “The Value of Artificial Intelligence in Cybersecurity” Ponemon Institute. 2018, 6

Understanding the Strategic and Technical Significance of Technology for Security @2019, The Hague Centre for Strategic Studies (HCSS) and The Hague Security Delta

A publication from
The Hague Security Delta (HSD)
Wilhelmina van Pruijsenweg 104
2595 AN Den Haag
T + 31 (0)70 204 5180
Info@thehaguesecuritydelta.com
www.thehaguesecuritydelta.com
@HSD_NL

Authors
Louk Faesen
Erik Frinking
Gabriella Gricius
Elliot Mayhew

Contributors
Alexander Klimburg
Katarina Kertysova
Martijn Neef (TNO)

Reviewer(s)
Emmy Koning (HSD)
Peter Zinn (HSD)

Design
Studio Maartje de Sonnaville by the design of
Studio Koelewijn Brügenwirth

Print
Drukkerij Edauw + Johannissen

# Together we Secure the Future

www.thehaguesecuritydelta.com