ESET White Paper

CAN ARTIFICIAL INTELLIGENCE POWER FUTURE MALWARE?

Ondrej Kubovič – ESET Security Awareness Specialist

with contribution of

Peter Košinár – ESET Technical Fellow
Juraj Jánošík – ESET Senior Software Engineer

ENJOY SAFER TECHNOLOGY™

CONTENTS

INTRODUCTION

Artificial intelligence (AI) is almost an omnipresent topic these days. It is the centerpiece of sales pitches, it "powers" various online services and is mentioned in regard to almost any new product seeking investors. While some vendors truly aim to bring the added value of this technology to their customers, others mostly use it as a buzzword but can hardly deliver on their promises.

A simple online search for the term "AI" today returns almost 2.2 billion results, illustrating the scope of the interest experts and the public is taking in the matter. Part of the hype can be attributed to the great new feats achieved thanks to this technology – for example AI helping researchers to see through walls – but it also has darker connotations, mostly predicting that AI could wipe out millions of jobs and render whole industries obsolete.

Machine learning (ML) as a subcategory of a yet-unachievable goal of true and self-sustainable AI has already triggered radical shifts in many sectors – including cybersecurity. Improved scanning engines, increased detection speeds as well as enhanced ability to spot irregularities were all factors that contributed to higher level of protection of businesses, especially against new and emerging threats as well as advanced persistent threats (APTs).

Unfortunately, this technology is not available exclusively to defenders. Black-hats, cybercriminals and other malicious actors are also aware of the benefits of "AI" and will probably try to employ it in their activities, in one form or another. Targeted attacks against businesses, money or data heists can potentially become more difficult to uncover, track and mitigate.

We could even argue that we are looking in the eye of an era in which "AI-powered cyberattacks" will become the norm, dethroning those operated by highly skilled, malicious actors. ESET, as an established security vendor that has been fighting cybercriminals for decades, understands the upcoming challenges and possible future scenarios, elaborating on them in this paper.

To provide a broader view, this white paper also presents the results of a survey ESET commissioned from OnePoll. Attitudes to, and concerns about, the use of AI and ML in cybersecurity contexts were gauged in a survey of almost 1000 US, UK and German IT decision makers, in companies with 50+ employees.

To avoid possible confusion, this white paper also addresses the differences between AI and ML and elaborates on the limits of the latter.

Finally, an overview of "AI-powered attacks" is provided, as is an insight into the design of ESET's machine-learning engine, Augur, and an outline of its enterprise-grade products designed to leverage this technology to counter constantly emerging and evolving cyber-threats.

Artificial intelligence vs. machine learning

The idea of artificial intelligence (AI) has been around for more than 60 years. It represents the yet unachievable ideal of a generally intelligent and self-sustainable machine that can learn independently, based only on inputs from the environment. Of course, all of this with no human interference.

Yet, today "AI" often refers only to a subcategory of this technology – namely machine learning (ML). This field of computer science originated in the 1990s and its real-world applications enable computers to find patterns in vast amounts of data, sort them and act upon the findings. These algorithms are the not-so-secret ingredient in all cybersecurity products that mention AI in their marketing claims.

Supervised, unsupervised or semi-supervised

In cybersecurity contexts, machine-learning algorithms are mainly used to sort and analyze samples, identify similarities and aim to produce a probability value for the processed object – putting it in one of three main categories: malicious, potentially unsafe/unwanted (PUSA/PUA) or clean.

However, to achieve the best possible results, this technology has to be trained on a large training set of correctly labeled clean and malicious samples, allowing it to understand the difference. This training and human oversight is why it is called supervised machine learning. Over the learning process the algorithm is taught how to analyze and identify most of the potential threats to the protected environment and also how to act proactively to mitigate them. Integration of this algorithm into a security solution makes it significantly faster and increases its processing capacity, compared to solutions that would only use human knowledge to protect client systems.

Algorithms without the training on completely and correctly labeled data belong to the category of unsupervised machine learning. These are very well-suited to finding similarities and anomalies in the dataset that could escape the human eye, but they do not necessarily learn how to separate the good from the bad (or more exactly clean from malicious). In cybersecurity, this can be a very useful feature to work with vast sets of labeled samples. Unsupervised learning can be used to organize that data into clusters and help create smaller, yet much more consistent training sets for other algorithms.

Semi-supervised machine learning falls in between the categories of supervised and unsupervised learning. Only partially labeled data is used for the learning process of the algorithm and the results are supervised and tweaked by human experts until a desired level of accuracy is achieved. The reason behind this approach is that creating a fully labelled training set is often laborious, time-consuming and costly. Also, for some problems, completely and correctly labeled data is currently non-existent, leaving semi-supervised learning as the only option to produce a working algorithm. ESET's machine-learning engine, Augur, works on a similar basis. It is used to classify items that were not part of its training set and were not previously labeled.

IS "AI" JUST MORE HYPE?

Apart from the original scientific term, *artificial intelligence*[1] is also a buzzword. Yet, how big is the hype? Thanks to significant advances in the field of machine learning and its broader application to real-world problems, the interest in AI grew over the last few years, reaching peaks in 2017 and 2018 not seen since the last decade.

This is documented by the search trend of the terms "machine learning" and "artificial intelligence".

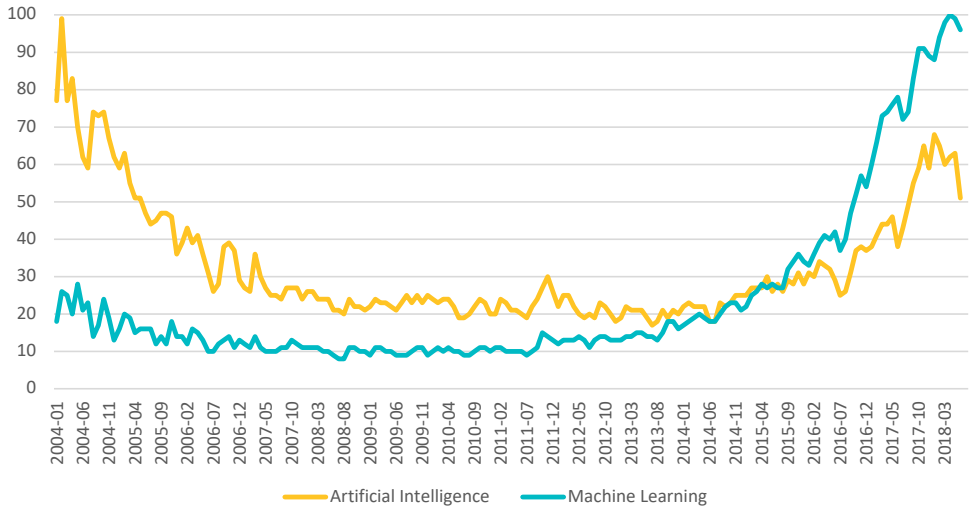


Figure 1 // Search trend of the terms "Artificial Intelligence" and "Machine Learning" 2004-2018 Source: Google Trends

This has also translated into business environments where machine learning (or AI) appears to be widely implemented, as observed in OnePoll's survey conducted on ESET's behalf.

According to the results 82% of IT decision makers in US, UK and German businesses with 50+ employees believe that their organization has already implemented a cybersecurity product utilizing machine learning. Of the rest, 53% declared their organization is planning to implement such a solution in the next 3-5 years, with 23% stating the opposite.

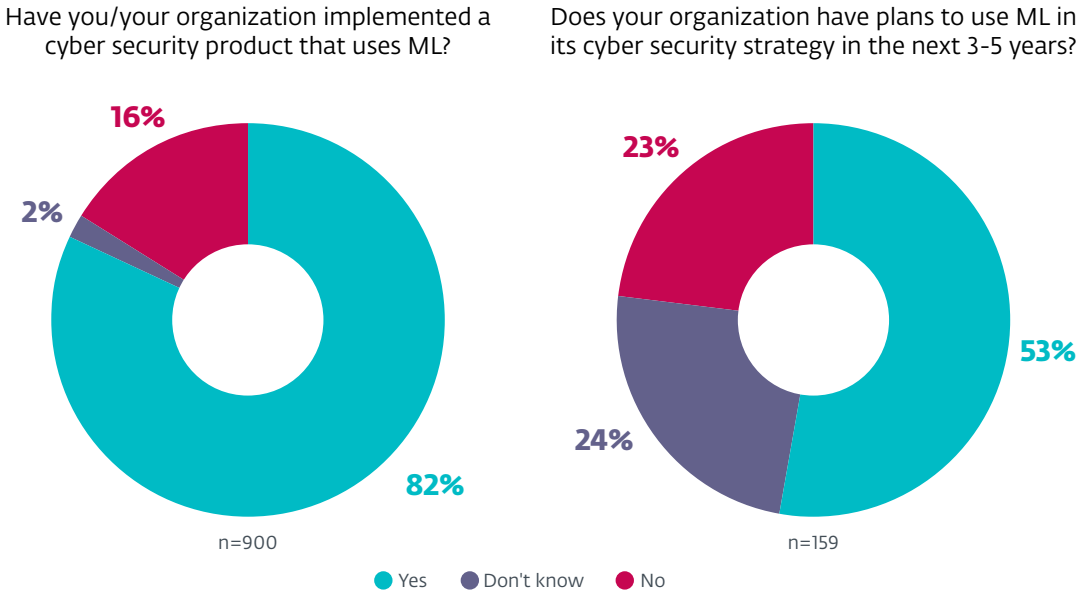


Figure 2 Figure 3

Eighty percent of respondents also believed that AI and ML will or does help their organization detect and respond to threats faster. IT decision makers also hope that these technologies will help them solve cybersecurity skills shortages in their workplace, with 76 percent somewhat or strongly agreeing with such statement.

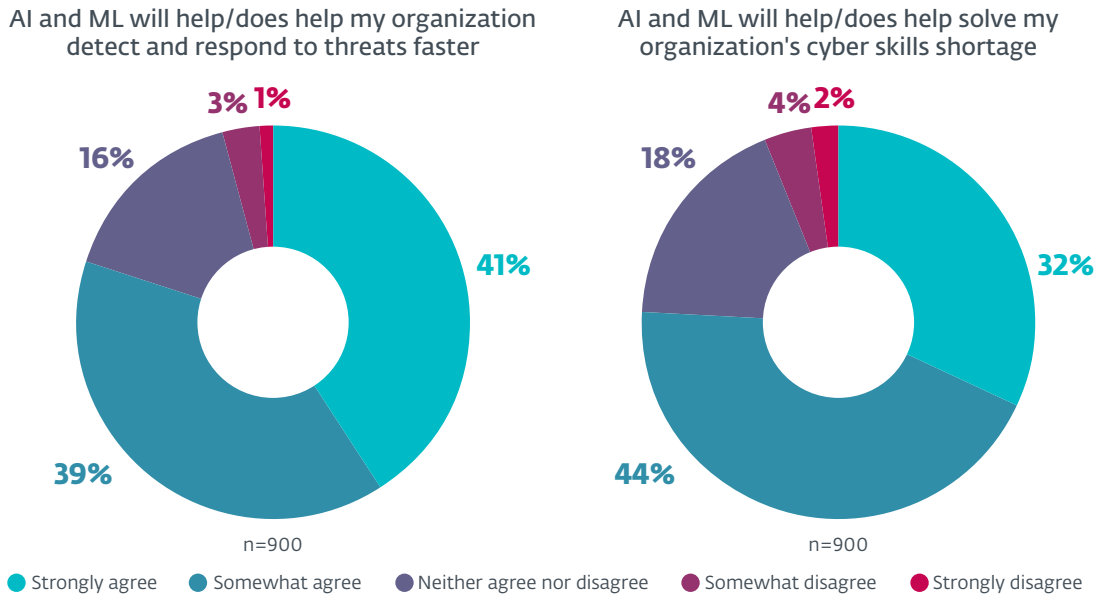


Figure 4 Figure 5

With the amount of marketing around AI and ML, many of the respondents tended to think that these technologies could be the key to solving their cybersecurity challenges, yet the majority also agreed that the discussions about the implementation of AI/ML in the defensive infrastructure are hyped.

So without diminishing the value of AI and ML as tools in the fight against cybercrime, there are limitations that need to be taken into account – such as the fact that relying on a single technology is a risk that can possibly lead to damaging consequences. Especially if an attacker has motivation, financial backing and time to find a way around the protective ML algorithm. A safer and more balanced approach to enterprise cybersecurity is thus to deploy a multi-layered solution that can leverage the power and potential of AI/ML, but backs it up with other detection and prevention technologies.

IS AI THE FUEL FOR FUTURE CYBERATTACKS?

Technological advances of machine learning have an enormous transformative potential for cybersecurity defenders. Unfortunately, not only for them, as cybercriminals too are aware of the new prospects. According to the OnePoll survey, managers and IT staff responsible for company security find this concerning:

Two-thirds (66%) of almost 1000 US, UK and German IT decision makers in the survey strongly or somewhat agreed that new applications of AI will increase the number of attacks on their organization. Even more respondents thought that AI technologies will make the threats more complex and harder to detect (69% and 70% respectively).

AI will/would increase the number of attacks my organization will have to detect and respond to

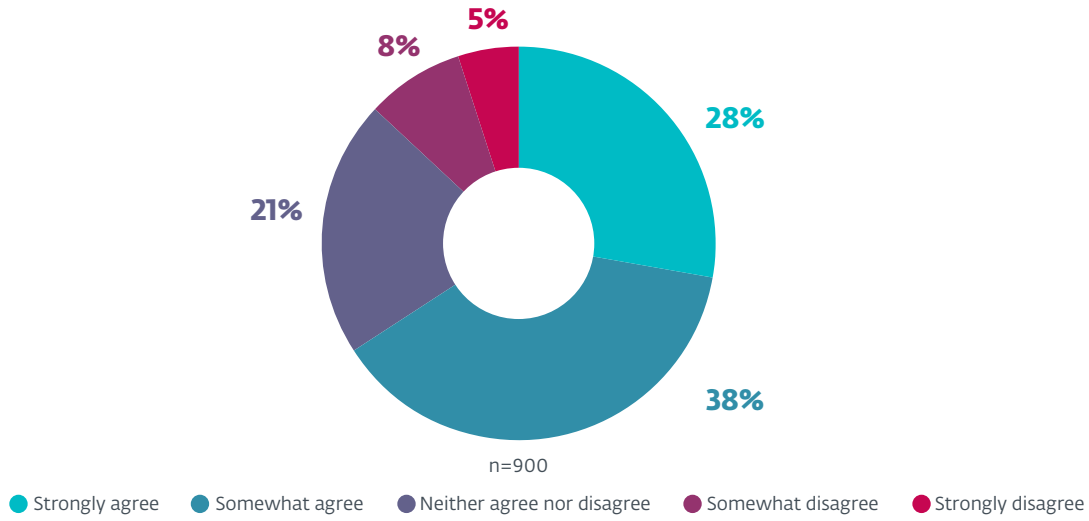


Figure 6

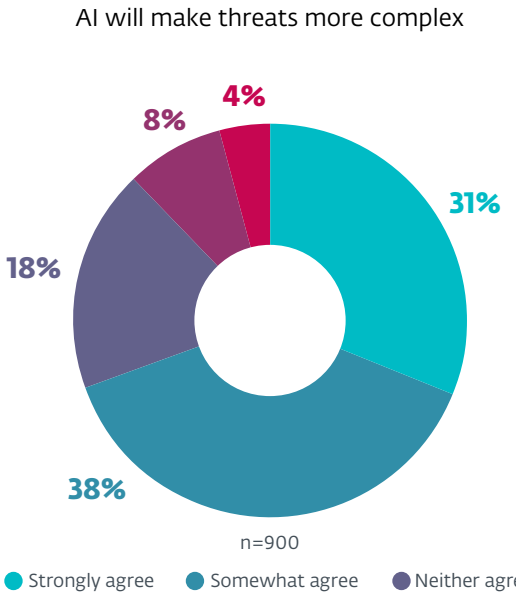


Figure 7

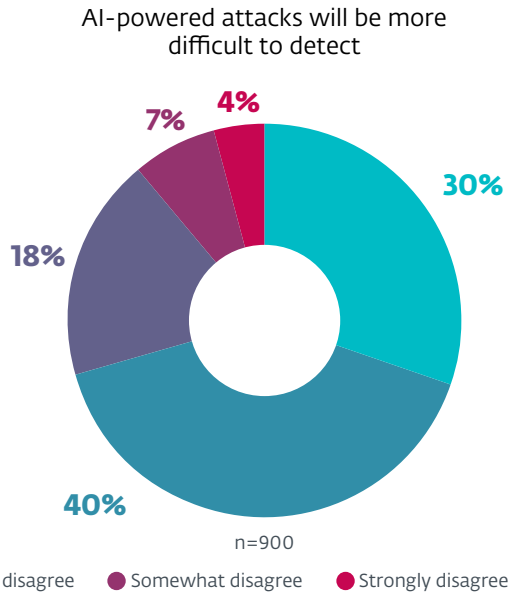
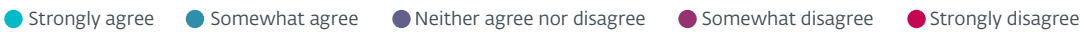


Figure 8



If and how these concerns materialize is yet to be seen. However, it surely won't be the first time the attackers have used technology to extend the reach of their malicious efforts. Already in 2003, the *Swizzor Trojan horse*[2] used automation to repack the malware once-a-minute. As a result, each victim was served a polymorphically modified variant of the malware, complicating detection and enabling its wider spread.

This approach would not be as effective against modern anti-malware solutions – such as ESET endpoint products – that detect malware's "DNA" and are able to identify it also via its network detections. However, by using advanced machine-learning algorithms, attackers could take a mechanism similar to Swizzor's and attempt to vastly improve this strategy. Without machine learning being a part of the defensive measure, an attacker's algorithm could learn the limits of the protective solution and alter the malicious code just enough to fly under the radar.

Automated variations of malware are far from the only possible malicious application of machine-learning algorithms. We looked at some of the areas where the use of this technology could give the attackers an advantage (and added some illustrative examples).

AI as a tool for the attackers

Attackers could utilize AI/ML to:

Protect their infrastructure, for example by**:**

- Detecting intruders, i.e., researchers, defenders, threat hunters in their systems
- Detecting inactive thus suspicious nodes in their network

Generate and distribute new content, such as:

- Phishing emails
  Fully or partially crafted and adjusted by the algorithm
- High-quality spam
  Thanks to machine learning the creation of new high-quality spam would be possible also for less prevalent languages, based on the amount of training material
- Disinformation
  Automatically combining legitimate information with disinformation and learning what works best and is most shared by the victims

Identify recurring patterns, oddities or mistakes in the generated content and help the attackers remove them.

Identify possible red flags that defenders are likely to look for.

Create false flags to divert attention to other actors/groups.

Choose the best target for their attack or to divide various tasks between infected machines according to their role in the network, without the need for outbound communication.

Misuse the AI model from **a defender's solution as a black box**

Attackers can install victim's protective solution on their device using the same configuration and use it to identify what traffic/content will pass through the defenses.

Find the most effective attack technique

Attack techniques can be abstracted and combined to identify the most effective approaches. These can be prioritized for future exploitation. In case defenders render one of the vectors ineffective, the attacker only needs to restart the algorithm and based on this new input, the technology will follow a different learning path.

Find new vulnerabilities

By combining the previous approach with fuzzing – i.e. providing the algorithm with invalid, unexpected, or random data as inputs – AI could learn a routine for finding new vulnerabilities.

AI in malware

Malware developers could utilize AI to:

Generate new, hard-to-detect malware variants

As already described in this paper, some older malware families (such as Swizzor) were using automation to generate new variants of themselves every minute. This technique could be reinvented and improved by using machine-learning algorithm(s) that would learn which of the newly created variants are the least likely to be detected and produce new strains with similar characteristics.

Conceal their malware in the victim's network

Malware can monitor behavior of nodes/endpoints in the targeted network and build patterns resembling legitimate network traffic.

Combine various attack techniques to find the most effective options that cannot be easily detected and prioritize them over less successful alternatives.

Adjust features/focus of the malware based on the environment.

If attackers want to targets for example browsers, instead of incorporating a complete list of browser and scenarios in the malware, malicious actors only need to implement a few of them for the most frequently encountered brands. The AI algorithm uses this training and learns directly on the endpoint how to infiltrate also the less popular and not previously specified browsers.

Implement a self-destructive mechanism in the malware that is activated if an odd behavior is detected.

Induced by a login of non-standard user profile or a program, malware automatically activates self-destruction mechanism to avoid detection or to render further analysis impossible.

Detect suspicious environment

If the algorithm detects a virtual machine, sandbox or other tools used by malware researchers, it can alter the behavior of the malware or temporarily stop its activity to avoid detection.

Increase the speed of the attack

The speed of an attack can be crucial, especially in cases such as data theft. Algorithms can perform the extraction significantly faster than a human could, making it harder to detect and almost impossible to prevent – as the machine can copy the data out of the protected perimeter before the defenders are able to react.

Let other nodes in the botnet learn collectively and identify the most effective attack forms

Learning and sharing information via multiple nodes can be of an advantage to the attackers, as each of the enslaved bots can test different infiltration techniques and report back the results. It can also help malicious actors learn more about the targeted infrastructure in a shorter time-frame.

AI as a part of (targeted) attacks

When choosing their targets, attackers could utilize AI to:

Decide if the visitor is worth attacking

By monitoring the traffic to the infected website, the algorithm can learn and select those visitors who are the most valuable targets and serve them malware.

Identify a specific protective solution

The outside attacker can perform network reconnaissance of the target network and, based on the responses or lack thereof, AI could be used to infer information about the security solutions employed by the targeted organization.

AI as part of attacks in mobile environments

Misuse popularity of mobile apps

Machine-learning algorithm can identify the apps with popular mods and create its own mod to blend in. Unwary users might download such apps to their mobile devices and thus infect them with malware.

AI in attacks targeting IoT

IoT devices such as routers, security cameras, and various controllers are growing in number. Many companies which implement them underestimate the fact that these devices are small computers prone to vulnerabilities and exploitation. Moreover, cheap and poorly designed IoT products often lack basic security measures and/or commonly use weak default device credentials; both shortcomings can allow easy infiltration by malware.

Attackers targeting IoT devices could utilize AI to:

- Generate credentials and use them to infiltrate other similar IoT devices
- Find new vulnerabilities of the IoT devices
- If the IoT devices are part of a botnet, algorithm can be distributed across all the nodes for collective learning
- Learn the standard processes and behavior for given devices (or their groups), identify rival malware and kill, disable or cripple it

Even malicious AI has its limitations

Similar to any other field, even application of AI to malware and malicious activities has its limitations. Perhaps the most important one was documented in the deployment of the first cyberweapon used in the wild, the now infamous Stuxnet.

This malware family was very effective at infecting protected and even air-gapped environments, enabling it to spread not only in the targeted systems but worldwide. However, such an aggressive behavior caught the attention of security researchers, who eventually identified and dissected the threat.

This could also apply to future AI-powered attacks. With the growing number of infiltrations, these threats would also become more prevalent and thus visible attracting more attention on the defenders' side, ultimately leading to their detection and mitigation.

LIMITATIONS OF MACHINE LEARNING

At ESET we have been experimenting with various forms of machine learning since early versions of the product, developing an automated detection system that helps us protect our customers. However, in that process, we also learned the limitations of this technology:

Limitation #1: Training set

First, to use machine learning effectively a lot of input samples are needed, every one of which must be correctly labeled. In a cybersecurity application this translates into a huge number of samples, divided into three groups – malicious, clean and potentially unsafe/unwanted.

ESET researchers have now spent over three decades gathering and classifying, and more recently choosing the samples that can be used as training material for ESET's ML engine, Augur. However, even when an algorithm has been fed a large quantity of data, there is still no guarantee that it can correctly identify all new samples. Thus human expertise and verification is required.

Without this process, even a single incorrect input can cause a "snowball effect" and possibly undermine the solution to the point of failure. The same situation ensues if the algorithm only uses its own output data as inputs for further learning. Errors are reinforced and multiplied, as the same incorrect result reenters the solution in a loop and creates more "trash" – false positives (FPs) or misses of malicious items.

Another weakness of solutions that rely on ML/AI only is the situation when attackers decide to target a new platform - such as a new scripting or application macro language, or a new file format. In that situation, it can take quite some time to collect enough "clean" and "dirty" samples to create a training set to train the model on.

Limitation #2: Math can't solve everything

Some post-truth security vendors claim that some of these limitations do not apply to their machine-learning algorithms, as theirs can identify every sample before it is executed and determine whether it is clean or malicious just by "doing the math". However, as proven by the famous mathematician, cryptanalyst and computer scientist Alan Turing (the man who broke the Enigma code during WW2 at Bletchley Park in the UK) a similar approach is not mathematically possible.

Even a flawless machine would not always be able to decide whether a future, unknown input would lead to undesirable behavior – in Turing's case, one that would make the machine loop indefinitely. This is called the "halting problem" and applies to many fields other than just theoretical computer science, where it originated.

Fred Cohen, the computer scientist who formulated the definition of a computer virus, demonstrated how this principle applies to cybersecurity by showing another undecidable problem: it is impossible to say with absolute certainty whether a program will act in a malicious way if one can only analyze it for a finite amount of time. The same problem applies with future inputs or commands from the attacker that might push a program into the malicious sphere.

So don't believe if a vendor claims its machine-learning algorithm can label every sample prior to running it (that is, pre-execution) and decide whether it is clean or malicious. By utilizing such an approach, it would have to preventatively block a large amount of undecidable items – flooding company IT security departments with false positives.

The other option would be less aggressive detection with fewer false positives, yet if only machine learning technology is applied, it would shift detection rates far from the claimed "100%" silver bullet efficiency.

Limitation #3: Intelligent and adaptive adversary

Another serious limitation to machine-learning algorithms in cybersecurity is the intelligent adversary. Experience teaches us that counteracting cyberattackers is an endless cat-and-mouse game. The ever-changing nature of the cybersecurity environment makes it impossible to create a universal protective solution, one that is able to counter all future threats. Machine learning does not change this postulate.

Yes, machines have gotten smart enough to *defeat humans at chess*[3] and at the *Go game*[4], however these environments have binding rules. In cybersecurity, the attackers do not follow guidelines or accept limitations. They are even able to change the entire playing field without a warning.

One good example are self-driving cars. Despite heavy investment into their development, these smart machines cannot guarantee success in real-world traffic. They work in limited areas and specific environments. But imagine a situation when someone covers or manipulates the traffic signs or resorts to sophisticated malicious acts like making traffic lights blink at a rate beyond human eye recognition. With these types of deformations made to the most critical environmental elements, the cars can start making poor decisions, which can lead to fatal crashes.

In cybersecurity, steganography is an example of such adversary activity. Attackers hide malicious code into harmless files such as pictures. By burying it deep into a pixel setting, the machine can be fooled by the (infected) file, which in its altered form is almost indistinguishable from its clean counterpart.

Similarly, fragmentation can also lead to a detection based solely on a machine-learning algorithm returning an incorrect evaluation. Attackers split the malware into parts and hide it in several separate files. Each of them is clean on its own; only at the precise moment they converge on one endpoint or network do they begin to demonstrate malicious behavior. In such cases, pre-execution red flags are not present.

Limitation #4: False positives

Cybercriminals are known to work hard to avoid detection and their methods exceed the above-mentioned examples in sophistication. They use their skills to hide the true purpose of their code, by "covering" it with obfuscation or encryption. If the algorithm cannot look behind this mask, it can make an incorrect decision. Both passing a malicious item as clean, and blocking a legitimate one, have significant negative consequences.

While it's understandable why a missed malware detection represents an issue for a company, it is less obvious with false positives – errors made when a protection solution incorrectly labels clean items as malicious.

Not every false positive necessarily leads to a total collapse of a business's IT infrastructure. But some glitches can disrupt business continuity and be potentially more destructive than a malware infection. A false positive in an automotive factory that incorrectly labeled part of the production line management software as malicious could disrupt production and likely translate into massive delays and millions of dollars in financial and reputational damage.

But, false positives don't need to break critical processes to be highly unwanted for organizations or their IT security staff. With tens or hundreds of false alarms daily (which may well be the case with a security solution based purely on machine learning), admins would only have two choices:

1. Keep the settings strict and lose work-days of time dealing with the FPs.
2. Loosen the protective setup, which can reduce the detection capability and potentially create new vulnerabilities in the company's infrastructure – a scenario that can be provoked and exploited easily by an experienced attacker, if the security solution is too aggressive.

Limitation #5: Machine learning alone is not enough

Building effective cybersecurity defenses for a company network is similar to protecting a house. Homeowners wanting to keep their properties safe need to set up as many protective layers as possible – e.g. strong fences, security cameras, loud alarms and motion detectors for the dark corners.

The approach in a business environment is similar. It would be unwise to rely solely on one technology – even if it is the newest machine-learning algorithm. With all its limitations, use of other protective layers is necessary to keep endpoints and crucial parts of the network safe.

In today's IT security environment, perimeter security alone would not be enough. Therefore additional and more elaborate tools are needed, such as endpoint detection and response (EDR) systems, threat intelligence as well as tools that allow for quick and reliable analysis of suspicious items, to provide enterprise security departments with necessary logs and forensic information.

Consequently, if a company aims to build reliable and strong cybersecurity defenses, it should select a balanced array of solutions and tools offering multiple complementary technologies with high detection rates and a low number of false positives. Or if we want to revert back to the home security metaphor – a complex security system that detects the thieves but will not sound the alarm when a neighbor's cat walks across the lawn.

MACHINE LEARNING BY ESET: THE ROAD TO AUGUR

At ESET we love ancient history – the company is named after an Egyptian goddess, after all – so antiquity was naturally where we looked when it came to naming our machine-learning engine. In ancient Rome, "augur" was a term used for religious officials who observed natural signs and interpreted these as indications of divine approval or disapproval of a proposed action.

The analogy with cybersecurity is not hard to draw, but in contrast with the alchemy-natured augurs back then, ESET's Augur engine bases its decisions on science, mathematics and previous experience.

Three trends that helped us shape the ESET Augur engine:

1. Arrival of big data and cheaper hardware
   Thanks to this shift, machine learning was made more affordable – be it for medical purposes, autonomous cars or detections in cybersecurity.
2. Popularity of machine-learning algorithms
   A growing number of successful real-life applications of ML led to a surge of investment into this field, rapid development of new capabilities, and boosts in academic as well as practical research, all contributing to wider availability of this technology.

3. **High-quality training material**
   Three decades of fighting black-hats and their "products" enabled ESET to build a latter-day "Library of Alexandria" of malware. This vast and highly organized sample set contains millions of extracted features and DNA genes of every sample ever analyzed in our VirusLabs, forming a great foundation for a precisely chosen training set, necessary for Augur's development.

However, the boom in these areas has also created new challenges. Our experts had to hand-pick the best performing algorithms and approaches, as not all machine-learning algorithms and technologies are equally applicable in the highly-specific security environment.

After much testing, we have settled on combining two methodologies that have proven effective so far:

- **Processing with deep-learning methods**
  Combination of long short-term memory (LSTM) and fully connected neural network
- **Multi-model processing (combining supervised learning methods)**
  Consolidated output of precisely chosen classification algorithms combining various ensemble methods, vector machines and decision trees

This combination of classification algorithms and deep-learning methods also increases Augur's resilience against adversarial activity. As documented in a *recently published paper*[5*] focusing on adversarial machine learning, an error-generic or error-specific evasion attack that would force an engine with structure similar to Augur's to misclassify a sample, would require a more complex strategy.

How Augur processes samples (see Figure 9)

The ESET engine emulates the behavior of a sample providing a set of features and sequences for further processing. Along with this procedure, a deep DNA analysis outputting numeric features of the sample is run. All the gathered data is then combined into a so-called xDNA.

In the following step, the gathered information is analyzed via Augur, using both the deep-learning methods and multi-model processing. We recently extended Augur with a subsystem for classification of samples, based on their binaries. In this subsystem a sample is first disassembled and the output is used for feature extraction. The outcome is then vectorized and fed to the neural network.

Each of the subsystems mentioned in the previous paragraph produces a separate probability value, which is consolidated by Augur into a final value labelling the sample as **clean, potentially unwanted/potentially unsafe, or malicious**.

It is important to note that, unlike some of the post-truth security vendors, ESET utilizes unpacking, behavioral analysis as well as emulation as part of our sample processing. We consider this to be a crucial step to properly extract a sample's features before they can be fed to the Augur engine. If only data from the analysis of compressed or encrypted samples were used, algorithms would attempt to classify noise, leading to largely meaningless results. (see Figure 9)

---

* Battista Biggioa, Fabio Rolia, Wild Patterns: Ten Years After the Rise of Adversarial Machine Learning, (2018), 6-7

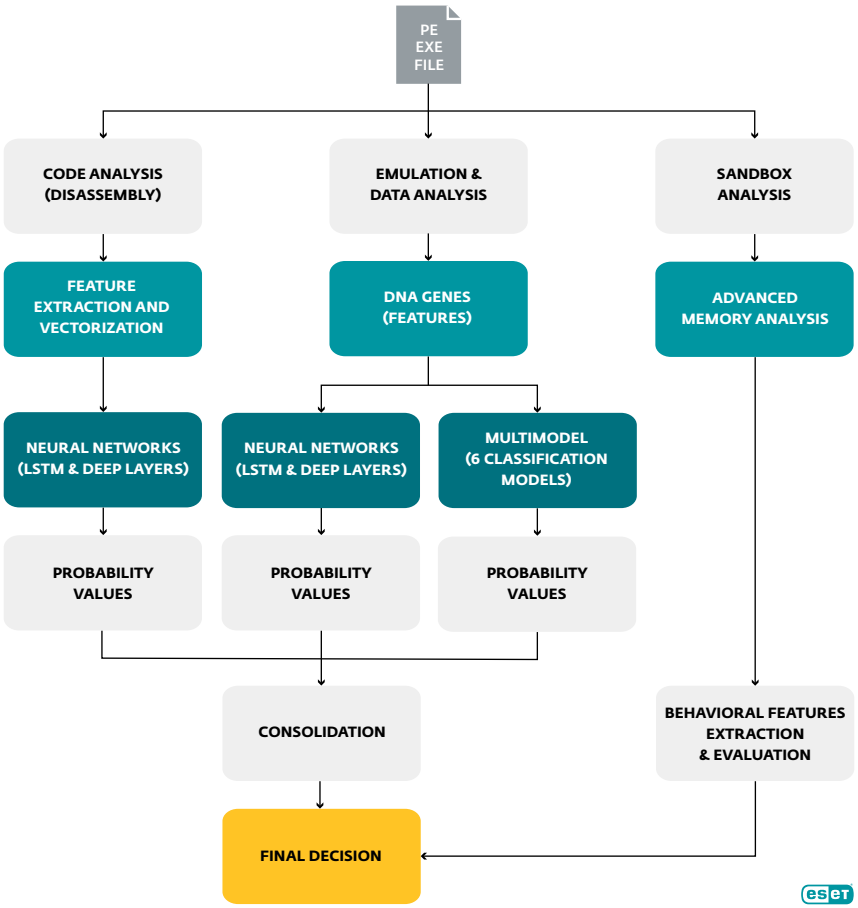


Figure 9 // Scheme detailing ESET's machine-learning engine "Augur".

To also offer a real-world perspective, we ran a series of tests to demonstrate the effectivness of Augur's analysis. We fed the old Augur model - originating in the first months of 2017 - with samples of well-known malware families that wreaked havoc in business environments later in 2017: WannaCryptor.D**, Diskcoder.C ((Not)Petya), Diskcoder.D (BadRabbit) and Crysis (one of the most intriguing ransomware families targeting enterprise and SMBs today).

Malware sample	Number of samples	Number of samples detected by Augur	Detection ratio
Win32_Diskcoder.C	16	10	62.5
Win32_Diskcoder.C (in-memory)	86	85	98.8
Win32_Diskcoder.D	17	14	82.4
Win32_Diskcoder.D (in-memory)	20	20	100
Win32_Filecoder.Crysis	113	112	99.1
Win32_Filecoder.Crysis (in-memory)	30	30	100
Win32_Filecoder.WannaCryptor.D	15	13	86.7
Win32_Filecoder.WannaCryptor (in-memory)	67	67	100

** Important to note, *Win32/WannaCryptor.D* as well as *Win32/DiskCoder.C* utilized a leaked NSA exploit, EternalBlue, to infiltrate victims. This spreading mechanism has been blocked by ESET's other technology – Network Attack Protection – which dropped the packets of the initial communication, effectively closing the vulnerability in the SMB protocol and blocing any attack leveraging the EternalBlue exploit irrespective of the final payload.

The results show that despite the Augur model being months older than the malware samples, the file detection ratio is fairly high, in some cases even flawless. However, the most important point for every business is that - even if the file was executed - Augur was able to correctly identify its malicous nature in memory and would give defenders a chance to stop the threat before it could cause damage within company infrastructure. We also need to stress, that Augur is only one protective layer implemented in ESET products, with a variety of other technologies stepping in if necessary.

Augur in ESET products

The power of Augur is already available to ESET clients on multiple fronts. Each endpoint and device that has ESET LiveGrid® enabled benefits from Augur's ability to analyze emerging threats in a fraction of the time possible for human analysis.

ESET's enterprise clients will also have Augur at their disposal via two enterprise-grade products:

1. ESET Enterprise Inspector (EEI) is ESET's Endpoint Detection and Response (EDR) tool. It works by collecting real-time data about ongoing activity on endpoints, which is then matched against a set of rules to automatically detect suspicious activities. The gathered information is processed, aggregated and stored in a searchable form, creating an overview of unusual and suspicious activities. EEI also provides the enterprise security team with information for forensic investigation of past incidents and offers response capabilities, to mitigate the presence of threat actors (advanced persistent threat or APT) in the network. Augur is integrated into EEI scanning and is essential to the process of flagging suspicious activities and samples.
2. ESET Dynamic Threat Defense (EDTD) is a system of multiple cloud components, allowing a customer's infrastructure to request information about previously analyzed samples directly from ESET's internal database. If a unique (never-seen-before) sample is part of an inquiry, it is uploaded to ESET's servers for deep analysis and evaluation – via ESET engine including Augur. Results are instantly returned to the customer.

Both of these products are designed to leverage advantages of the Augur engine and to work in sync with ESET's endpoint products.

CONCLUSION

As documented in this white paper, machine learning has many implications for cybersecurity. Unfortunately, this includes seasoned cyber attackers, who we presume will start to use this technology to protect their malicious infrastructure, improve malware they create and to find and target vulnerabilities in company systems.

We need to stress that as of now, there is no known evidence of machine learning being used to "power malware" per-se. Yet, the hype around the topics and growing number of news stories revolving around massive data leaks and cyberattacks fuels fears in company IT departments of what is yet to come.

According to the results of ESET's survey, conducted in the most advanced markets (United States, United Kingdom, Germany), the vast majority of IT decision makers are concerned about the growing number and complexity of future "AI-powered attacks" as well as the increased difficulty of detecting them. As a result, these defenders are implementing security solutions whose marketing materials promise advanced and reliable detection mechanisms utilizing "AI".

Development in the threatscape and these circumstances are the reasons why established security vendors – such as ESET – constantly improve their protective layers and incorporate machine learning (or AI in broader terms) into their solutions. Tests of ESET's Augur engine show how powerful the combination of a multi-layered solution and ML can be - even if facing dangerous global threats such as WannaCry, NotPetya, BadRabbit or Crysis ransomware.

With the rapid advances in the AI field, it is difficult to predict when a shift towards attacks making wide-scale usage of machine learning happen, and at what point malware families will become enhanced by technologies broadly described as "artificial intelligence". But if at that time proper counter-measures and security tools utilizing the technology are in place, the impact and inflicted damage can be significantly reduced.

EXECUTIVE SUMMARY

As "AI-powered" attacks are becoming one of the trending topics amongst enterprise users, ESET offers its view on the potential malicious uses of machine-learning algorithms (or AI if you will) without creating further hype around the topic. To put things into broader perspective, ESET augments its stance by describing the limitations of this technology, as seen by our research and development teams. The paper also presents results of a survey conducted on behalf of ESET among almost 1000 US, UK and German IT decision makers, with focus on the use of AI/ML in business security and concerns connected with this field. The final part of the paper is dedicated to ESET's implementation of this cutting-edge technology in its multi-layered engine and its enterprise and consumer portfolio.

HYPERLINKS

1 http://approximatelycorrect.com/2018/06/05/ai-ml-ai-swirling-nomenclature-slurried-thought/

2 http://www.virusradar.com/en/Win32_TrojanDownloader.Swizzor/detail

3 https://www.technologyreview.com/s/541276/deep-learning-machine-teaches-itself-chess-in-72-hours-plays-at-international-master/

4 http://www.cnbc.com/2017/05/23/googles-alphago-a-i-beats-worlds-number-one-in-ancient-game-of-go.html

5 https://arxiv.org/abs/1712.03141