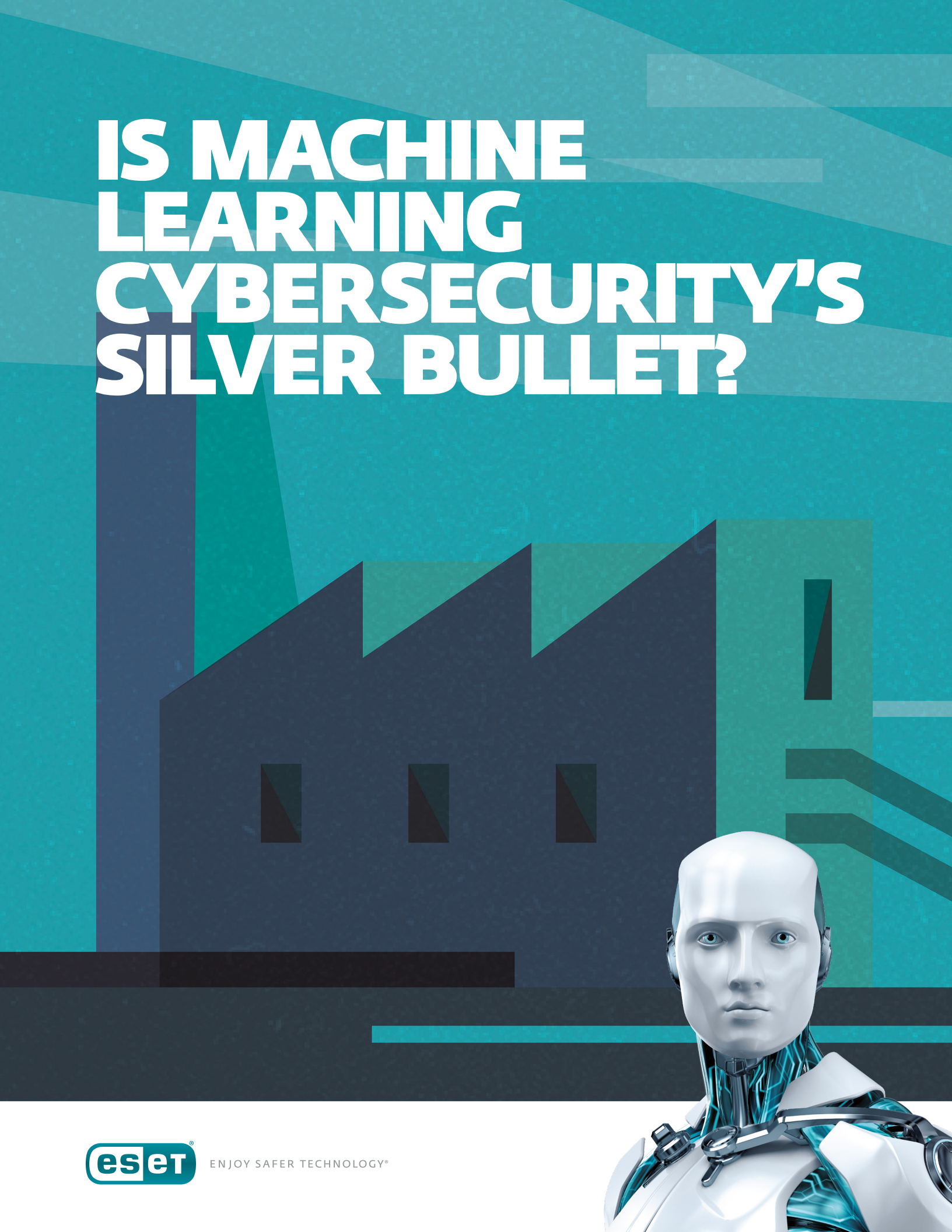
IS MACHINE LEARNING CYBERSECURITY'S SILVER BULLET?



ENJOY SAFER TECHNOLOGY®

FIGHTING POST-TRUTH WITH REALITY IN CYBERSECURITY

The world is changing in front of our eyes. Where facts, truth and honesty were once our most valuable assets, nowadays alternative-facts, post-truths and outright lies reign. Unfortunately, the cybersecurity business is no exception to this trend.

Even worse, with all the recent advances in the field of artificial intelligence (AI) and machine learning (ML), cybersecurity is all the more complicated and thus confusing – opening opportunities for players who like to inflate their abilities and ignore the limitations.

Machine learning algorithms as a cybersecurity silver bullet? No need for updates, or the downplayed importance of false positives; those are just a few of the often used marketing tricks from the toolbox of these so-called "next-gen" – or as we call them – "post-truth" vendors.

Established vendors such as ESET, who have fought the cybersecurity fight for almost three decades, know the possible downsides of an over-reliance on machine learning. To bring more clarity to the murky waters of post-truth marketing, we have put together this paper focusing on the currents state of AI and all the ins and outs of ML.

The key outcome? True *artificial intelligence* doesn't exist yet and machine learning is still not mature enough to be the only layer standing between you and cyber attackers.

SUPERVISED VS. UNSUPERVISED MACHINE LEARNING

The idea of *AI* has been around for more than 60 years and represents the ideal of a generally intelligent machine that can learn and make decisions independently, based only on inputs from its environment – all without human supervision.

A step back from this as-yet unachievable AI dream, is machine learning, a field of computer science that gives computers the ability to find patterns in vast amounts of data, by sorting them and acting on the findings.

The concept might be a little newer, but it has still been present in cybersecurity since the 90s. In cybersecurity it primarily refers to one of the technologies built into a protective solution that has been fed large amounts of correctly labeled clean and malicious samples, thus learning the difference.

Thanks to this training and with oversight of humans – also known as **supervised** machine learning – it is able to analyze and identify most of the potential threats to users and act proactively to mitigate them. Automation of this process makes the security solution faster and helps human experts handle the exponential growth in the number of samples appearing every day.

Algorithms without similar "training" – fall into the category of **unsupervised** machine learning – are almost useless for cybersecurity. While able to sort data into new categories, they don't necessarily distinguish between clean items and malware. This makes them suited to finding similarities or anomalies in the dataset invisible to the human eye, but it doesn't make them better at separating the good from the bad.

LIMITS OF MACHINE LEARNING

At ESET we have been applying supervised machine learning for years. We just call it "automated detection".

To keep our detection rates high and false positives low, a team of experienced human supervisors evaluates items that are too divergent from other samples, and hence hard for ML to label. This approach allows us to avoid the pitfalls of false positives (FP) or misses which might occur on the way to a fine-tuned algorithm that works well with other protective technologies under the hood of our solutions.

But basically, there is no magic in machine learning. Under the supervision of our experts it learns how to extract features and find specific patterns in huge quantities of malicious and clean data. And it has helped us protect millions of users worldwide for years.

However, this technology comes with its own challenges and limitations that need to be addressed during the course of its implementation:

LIMIT #1
Training set

First, to use machine learning a lot of inputs 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 unwanted. We've spent almost three decades gathering, classifying and choosing the data that can be used as training material for our ML engine.

Where would a recently formed post-truth vendor get such data? Unless it resorts to the unethical use of competitor

research, there is no way to create a sufficiently large or reliable database, not even mentioning the labor required to sort such a database.

However, even when a ML algorithm has been fed a large quantity of data, there is still no guarantee that it can correctly identify all the new samples it encounters. Human verification is still needed. Without this, even one incorrect input can lead to a snowball effect and possibly undermine the solution to the point of complete failure.

The same situation ensues if the algorithm uses its own output data as inputs. Any further errors are thus reinforced and multiplied, as the same incorrect result enters a loop and creates more "trash" – false positives or misses of malicious items – that then reenters the solution.

LIMIT #2
Math can't solve everything

Some post-truth security vendors claim that similar situations can't happen with their machine learning algorithms, since they can identify every sample before it gets executed and determine whether it is clean or malicious just by "doing the math".

However, the famous mathematician, cryptanalyst and computer scientist Alan Turing (the man who broke the Nazi Enigma code during WW2 at Bletchley Park in England) proved that a similar approach isn't mathematically possible. Even a flawless machine would not always be able to decide whether a future, unknown input would lead to unwanted 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 theoretical computer science, where it originated.

For instance, Fred Cohen, the computer scientist who formulated the definition of a computer virus, demonstrated how it 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 emerges with future inputs, or specific settings that might push a program into the malicious sphere.

So how does this apply to cybersecurity? If a post-truth vendor claims its machine learning algorithm can label every sample prior (or pre-execution) to running it and decide whether it is clean or malicious, then it would have to preventively block a huge amount of undecidable items – flooding company IT departments with false positives. The other option would be less aggressive detection with fewer false positives. However, if only machine learning technology is applied, it would shift detection rates far from the claimed "100%" silver bullet efficiency.

LIMIT #3
Intelligent and adaptive adversary

On top of the abovementioned challenges connected with any application of ML to cybersecurity, there is another serious limitation: the intelligent adversary.

Experience teaches us that counteracting cyber attackers 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 any future threat. And machine learning doesn't change this. Yes, machines have gotten smart enough to *defeat humans at chess*[1] or even at the *Go game*[2], however these games have binding rules while in cybersecurity, the attackers don't stick to any. What's worse, they are even able to change the entire playing field without warning.

Let's take self-driving cars as an example. So far, despite heavy investment into development, these smart machines can't guarantee success in real-world traffic, i.e. beyond limited areas with an environment. Now imagine that someone covers all the traffic signs, manipulates them 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 critical elements, the cars can begin to make poor decisions which can end in fatal crashes, or simply immobilize the vehicles.

In cyber security, steganography serves as a great example of adversary activity. Attackers just need to take malicious code and smuggle it into harmless files such as pictures. By burying it deep into a pixel setting, the machine can be fooled by the (infected) file, which is now almost indistinguishable from its clean counterpart.

Similarly, fragmentation can also lead to a detection 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. Pre-execution red flags are simply missing in such cases.

LIMIT #4
False positives

Cybercriminals are known to work hard to avoid detection and their methods exceed the above-mentioned example

1 https://www.technologyreview.com/s/541276/deep-learning-machine-teaches-itself-chess-in-72-hours-plays-at-international-master/
2 http://www.cnbc.com/2017/05/23/googles-alphago-a-i-beats-worlds-number-one-in-ancient-game-of-go.html

in sophistication. They try 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. Either labeling a malicious item as clean or blocking a legitimate one have negative consequences. While it's easy to understand why a missed detection poses a problem, so called false positives – errors made when a protection solution incorrectly labels clean items as malicious might be even worse.

Sure, not every false positive necessarily leads to a total collapse of a business's IT infrastructure. But some glitches can disrupt business continuity and thus potentially be even more destructive than a malware infection. Just imagine an automotive factory halting production because its security solution labeled part of the production line's software as malicious and subsequently deleted it – a "glitch" likely to translate into massive delays and millions of dollars in financial and reputational damage.

False positives don't need to break critical processes to be highly unwanted for organizations and their IT security staff. With tens or hundreds of false alarms daily (which may well be the case with a security solution set to an extremely aggressive mode), admins would only have two choices:

1. Keep the settings strict and waste time dealing with the FPs.
2. Loosen the protective setup, which at the same time would likely create new vulnerabilities in the company's systems.

Now how difficult can it really be for experienced attackers to provoke and exploit the latter scenario if an aggressive solution were in place?

BALANCING DETECTION AND FALSE POSITIVES

Of course, it would be easy to achieve 100% detection - by flagging every sample as malicious - or 0% false positives – by labeling every sample as clean - but it is mathematically impossible to reach both at the same time. Thus, the goal in malware protection is to achieve an equilibrium of sufficient protection from malicious items and false positives minimized to a manageable level.

This can be achieved via the following:

Human involvement

Some IT environments require 24/7 monitoring, and a responsible person who can react almost instantaneously to any suspicious activity or security notification. This is certainly the case for sensitive systems, such as a car factory or other production lines, but cannot be applied to all systems.

Whitelisting lockdown

In restrictive environments – such as bank employee terminals, where identical devices run only a limited set of applications – admins can opt for whitelisting. This allows them to create a detailed list of authorized actions and software. Anything off the list gets blocked, regardless of whether it is clean or malicious.

This "whitelisting lockdown" reduces the attack surface significantly and minimizes false positives, but it also shrinks the functionality of the system and is not applicable universally. Another limit to this approach is that blocking automatic updates may lead to endpoints running a vulnerable version of the app.

Less restrictive approaches to whitelisting, or "smart" whitelisting, have defined exceptions for updaters, paths or file names.

As businesses use their unique mix of software within their networks, it is therefore up to them to decide how restrictive its security systems should be in order to achieve the desired level of protection.

Minimal functionality

If the system can be stripped down to minimal functionality, it lowers the attack surface, but leaves a lot of legitimate activity and files out. On the other hand, for some businesses a false positive would have a higher cost than a potential infection, which forces them to take the risk.

Well-tuned security solution

The most effective way to protect general-purpose systems, networks and/or endpoints is to deploy a well-tuned security solution and to supervise it with experienced administrator(s) who can take care of the rare cases when FPs occur.

NECESSITY OF UPDATES

Emerging cybersecurity vendors criticize their established counterparts for depending on regular updates of their virus databases as well as their engines. As an alternative, some of them offer a solution based solely on machine learning (ML) algorithms that acquire all the data on clients' local machines and in their security environments, resulting in one "perk": No updates necessary.

But is that really an advantage?

Solutions that protect systems locally can be very effective and relatively successful in countering threats. However, this is only true for:

a) Specific environments with very limited functionality; or

b) Systems that are strongly averse to change or are – partially or totally – isolated from connections to the outside world.

However, the vast majority of endpoints in small, medium and large companies don't operate in a restricted environment like that. They need to communicate with contractors, clients and potential partners, as well as with each other; which requires a near-constant internet connection.

So even if the security algorithm is good at learning from the user and his network, without the global context provided by updates to its virus database, it can have difficulty correctly identifying incoming external data as clean or malicious. This can lead not only to an increase in the rate of false positives, but in the worst case scenario, to a "miss" – an infection caused by mistaking malware for a clean item.

Based on data from tens of millions of nodes, ESET protection systems combine human oversight with the latest technologies to provide real-time updates to whitelists and systems, which can then properly label suspicious or unfamiliar items with a high degree of accuracy.

There are other benefits too:

- **Lower company-side hardware demands**
  Any of the analyzed samples may already have been evaluated by other endpoints in the global network, they don't require reevaluation.
- **Building a reliable threat database stored in the cloud**
  By sharing with all recognized endpoints, this can protect users from a wider array of malicious items than a ML algorithm that only learns from a very limited number of machines.
- **Updated solution can cover extraction methods and samples, whenever machine learning cannot do so on its own.**

MACHINE LEARNING BY ESET **THE ROAD TO AUGUR**

Despite all the above mentioned limits of machine learning, we see the value of this technology. That's also the reason why our experts have been playing with machine learning for more than 20 years – with neural networks making their first appearance in our products in 1998.

One of our early efforts was an automated expert system, designed for mass processing. In 2006, it was quite simple and helped us process part of the growing number of samples and cutting the immense workload of our detection engineers. Over the years, we have perfected its abilities and made it a crucial part of the technology responsible for the initial sorting and classification of the hundreds of thousands of items we receive every day from sources such as our worldwide network ESET LiveGrid®, security feeds and our ongoing exchange with other security vendors.

Another ML project has been running under ESET's hood since 2012 placing all the analyzed items on "the cybersecurity map" and flagging those, which required more attention.

ESET's current ML engine could have difficulties to materialize without three main factors:

1. With the arrival of big data and cheaper hardware, machine learning was made more affordable.
2. Growing popularity of ML algorithms and the science behind it led to their broader technical application and availability to anyone who was willing to implement them.
3. After three decades of fighting black-hats, we have built a latter-day "Library of Alexandria" equivalent – of malware. This vast and highly organized database contains millions of extracted features and DNA genes of everything we've analyzed in the past. This was a great foundation for our carefully chosen mix that has become Augur's training set.

These developments as well as other internal ML projects helped us gain experience, and piece-by-piece paved the way for what we have today – a mature, real world application of machine learning technology in the cloud, as well as on client's endpoints that we call Augur.

However, the boom of the above named factors has also brought challenges. We have had to pick the best performing algorithms and approaches, as not all machine learning is applicable to the highly specific cyber security universe.

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

1. **Neural networks**, specifically deep learning and long short-term memory (LSTM).
2. Consolidated output of six precisely chosen **classification algorithms**.

Not clear enough? Imagine you have a suspicious executable file. Augur will first emulate its behavior and run a basic DNA analysis. Then it will use the gathered information to extract numeric features from the file, look at which pro-

cesses it wants to run and look at the DNA mosaic in order to decide which category it fits best – clean, potentially unwanted or malicious. At this point, it is important to state that unlike some vendors who claim they do not need unpacking, behavioral analyzing or emulation, we find this crucial to properly extract data for machine learning. Otherwise – when data is compressed or encrypted – it' just an attempt to classify noise.

The used group of classification algorithms has two possible setups, each aiming for different outcome:

The more aggressive one will label a sample as malicious if most of the six algorithms vote it as such. This is useful mainly for IT staff using ESET Enterprise inspector, as it can flag anything suspicious and leave the final evaluation of the outputs to a competent admin.

The milder or more conservative approach, declares a sample clean, if at least one of the six algorithms comes to such conclusion. This is useful for general purpose systems with less expert overview.

We know visuals are everything today, so if the previous explanations weren't clear enough, chart on the next page might help.

Okay, so let's move away from theory and look at the real world results of ESET's machine learning approach as applied to the recent malware attacks misusing the EternalBlue exploit and pushing both the WannaCryptor ransomware and CoinMiner malware families. Apart from our network detection and effective flagging by our other ML system, the Augur model also immediately identified samples of both families as malicious.

What's more interesting, we also ran this test with a month old Augur model that couldn't have encountered these malware families anywhere before. This means, the detections were based solely on the information learned from the training set. And guess what? They were both correctly labeled as malicious.

30 years of progress and innovation in IT security have taught us, that some things don't have an easy solution, especially in cyberspace, where change comes rapidly and the playing field can shift in a matter of minutes. Machine Learning, even when wrapped up in shiny marketing speak, won't change that anytime soon. Therefore, we believe that even the best ML cannot replace skilled and experienced researchers, who built its foundations and who will further innovate it.

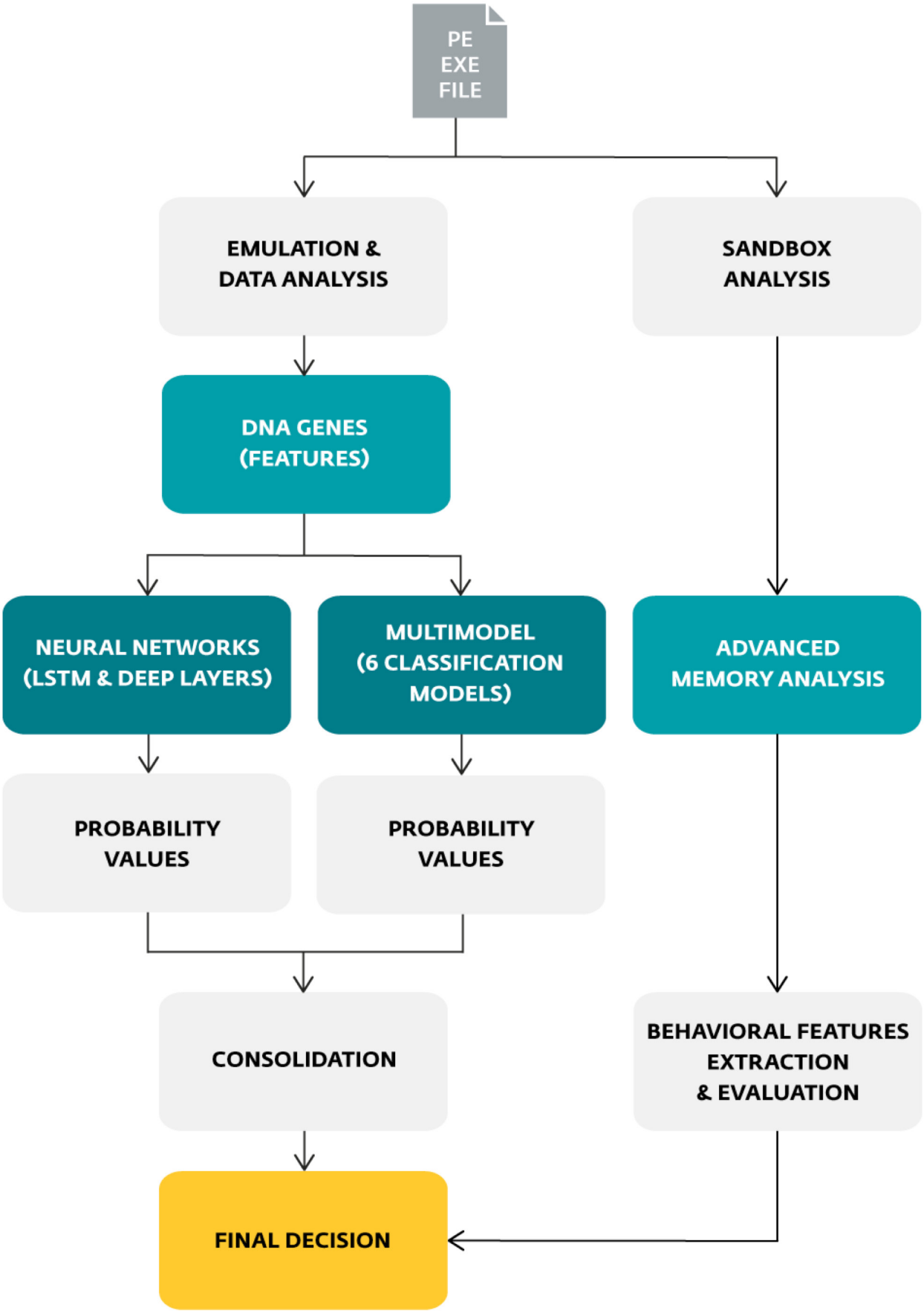
CONCLUSION

Building effective cybersecurity defenses for a company network is similar to protecting your own home. If you want to keep it safe, you will try to have as many protective layers installed as possible – a strong fence, a set of security cameras, a very loud alarm and motion detectors for the dark corners.

In a business environment, it would be unwise to rely solely on one technology – even if it is a machine learning algorithm. With all the limitations to ML mentioned in this paper, it is clear, that the use of other means is also necessary to keep users safe. Remember, avoiding protective solutions is a cybercriminal's daily bread. Moreover – as has been proved again and again in the past – any feature or system can be circumvented given enough effort.

Therefore a company aiming to build reliable and strong cybersecurity defenses should opt for a solution offering multiple complementary technologies with high detection rates and a low number of false positives. In other words – reverting back to the home metaphor – one that catches thieves but doesn't react when a neighbor's cat walks across the lawn.

Thanks to 30 years of research and development, ESET can offer fine-tuned mix of time-proven protective technologies and its advanced machine learning engine named Augur.


PE
EXE
FILE
EMULATION &
DATA ANALYSIS
SANDBOX
ANALYSIS
DNA GENES
(FEATURES)
NEURAL NETWORKS
(LSTM & DEEP LAYERS)
MULTIMODEL
(6 CLASSIFICATION
MODELS)
ADVANCED
MEMORY ANALYSIS
PROBABILITY
VALUES
PROBABILITY
VALUES
CONSOLIDATION
BEHAVIORAL FEATURES
EXTRACTION
& EVALUATION
FINAL DECISION

welivesecurity

news, views and insight from the ESET security community

WeLiveSecurity.com is where ESET experts are. The site is an editorial outlet for internet security news, views and insight. It covers relevant breaking news and aims to cater to all skill levels by offering video tutorials, in-depth features and podcasts.