



# IS THE AI HYPE PUTTING BUSINESSES AT RISK?



ENJOY SAFER TECHNOLOGY™

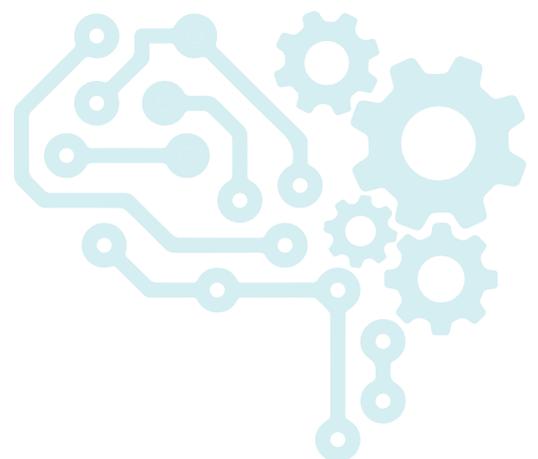
The media is awash with talk about the benefits of Artificial Intelligence (AI) and Machine Learning (ML) in cyber security. Next-generation vendors are increasingly bringing AI-based cyber security products to market in a big way, heralding the technologies as game-changers in the industry. With their ability to instantly detect any malware on a network and mitigate risks before they even start, the AI / ML sales pitch is certainly an attractive one for organisations to buy into.

**But the truth is that the sales pitch may be misleading. And the hype could actually be putting businesses at greater risk.**

In this paper, we discuss that while ML has proven to be a powerful tool in detecting malware for many years, the reality is that true AI does not yet exist. The marketing tricks of next-gen vendors are simply making matters all the more confusing for IT decision makers who need to build robust cyber security defences at a time when the threat landscape is becoming all the more precarious.

## CONTENTS

Conflicting opinions	2
Nothing new	3
The state of play today	4
Do AI and ML change the game?	6
Machine + Man	7
Looking beyond the hype	8



# Conflicting opinions

We surveyed IT decision makers in businesses across the US, UK and Germany about their attitudes and approaches to AI and ML in cyber security and it was clear that IT decision makers are confused and have conflicting opinions.

While a high percentage of respondents regard AI and ML as the silver bullet to solving their organisation's cyber security challenges, a significant number also argue that discussions around these technologies are purely just hype. US IT decision makers were most likely to consider the technology as a silver bullet to their digital defences, with 82% agreeing AI and ML would solve their cyber security challenges compared to 66% of German respondents (see figure 1).

Yet US IT decision makers were also more likely to consider the discussions around AI and ML as hype – 65% in comparison to the 53% of UK respondents and 40% of German respondents (see figure 2).

Do IT decision makers really know what to believe?

What's more, there is confusion over the terminology used as just 53% of IT decision makers said their company *fully* understands the differences between the terms AI and ML (see figure 3).

Figure 1: % of IT decision makers that agree AI and ML is the silver bullet to solving their organisation's cyber security challenges

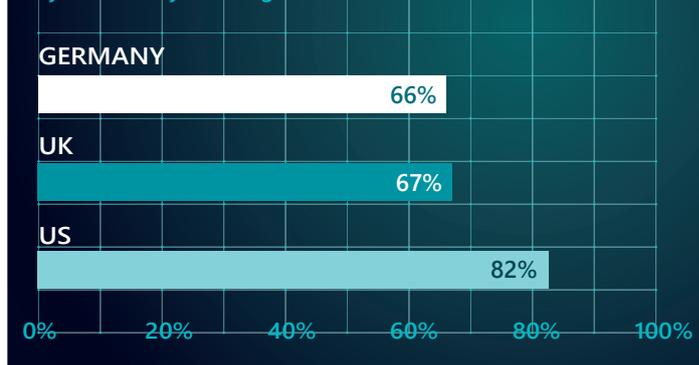


Figure 2: % of IT decision makers that agree discussions around AI and ML are just hype

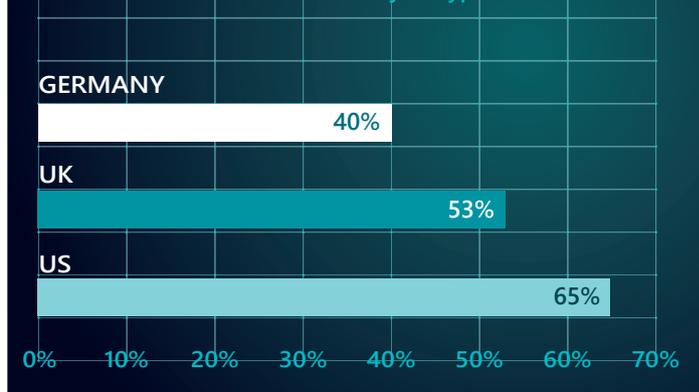
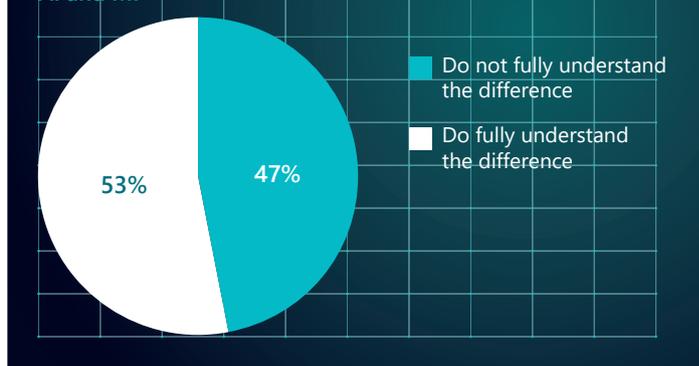


Figure 3: % of IT decision makers that say their organisation understands the differences between AI and ML



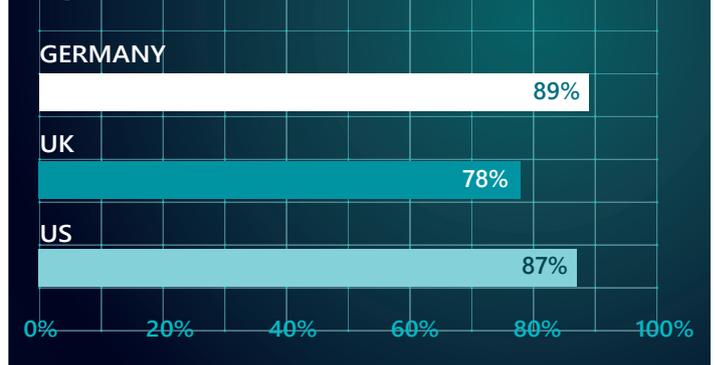
# Nothing new

Unfortunately, the terminology used in today's media and marketing materials is often misleading. In many cases, the term 'ML' is wrongly interchanged with 'AI'. Put simply, AI happens when machines conduct tasks without pre-programming or training. ML, in comparison, relies on training computers, using algorithms, to find patterns in vast amounts of data and identify data based on rules and information it already has. ML is nothing new; it has been present in cybersecurity since the 90s.

What's more, the majority of IT decision makers we surveyed have already implemented ML in their cybersecurity strategies with 89% of German respondents, 87% of US respondents and 78% of UK respondents saying their endpoint protection product uses ML to protect their organisation from malicious attacks.

There needs to be greater clarity around the claims marketing teams at next-gen vendors are making. The threat landscape is becoming an even more complex environment to navigate as hackers try new ways to gain access to company networks. The hype surrounding AI and ML as the silver bullet to solve cyber security challenges muddles the message for those making key decisions on how best to secure their company's networks and data.

Figure 4: % of IT decision makers that say their end-point protection product use ML to protect the organisation from malicious attacks



Unfortunately, the terminology used in today's media and marketing materials is often misleading

# The state of play today

It is important to note that ML is a powerful tool in the fight against cybercrime, especially malware scanning – helping to detect potential threats to users who can proactively mitigate them much more quickly.

ML primarily refers to one of the technologies built into the protective solution that has been fed large amounts of correctly labelled clean and malicious samples to essentially learn the difference between the good and the bad. Thanks to this training, it is able to analyse and identify most of the potential threats to users and act proactively to mitigate them.

The technology's ability to detect threats quickly and mitigate the rising number of samples emerging every day. This is what makes the technology so appealing to IT decision makers.

**However, ML – if it's done properly – comes with problems and limitations that marketing materials seem to brush over.**

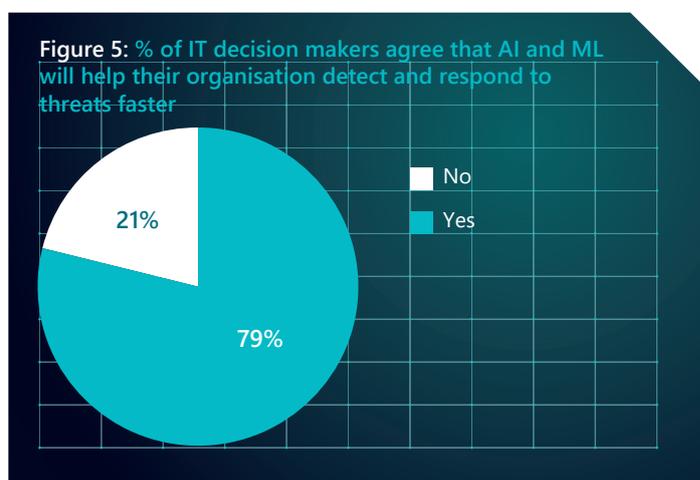
## 1. Machines need supervised training

First, to use ML you need a lot of inputs – and every one must be correctly labelled. In a cybersecurity application this translates into a huge number of samples, divided into three groups – malicious, clean and potentially unwanted. At ESET, we've spent almost three decades gathering, classifying and choosing data to train our ML system.

What's more when an 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.

We've heard some next-gen security vendors claim that similar situations can't happen with their machine learning algorithms, since they can identify every sample before it runs and determine whether it is clean or malicious just by doing the "math".

But once again, there are some flaws to this claim – flaws that simply confuse matters for IT decision makers.



## 2. Maths isn't enough

The truth is that even a flawless machine would not always be able to decide whether a future, unknown input would lead to unwanted behaviour<sup>1</sup>. If a next-gen vendor claims its machine learning algorithm can label every sample prior 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 (errors made when a protection solution incorrectly labels clean items as malicious).

Of course, not every false positive necessarily leads to the collapse of a business' IT infrastructure. They can, however, disrupt business continuity and thus potentially be even more destructive.

The role of a human is critical here. ML systems need the option to notify teams when they come across something they haven't seen before and ask for help from a human.

<sup>1</sup> Known as the halting problem, proven by the English mathematician, computer scientist and cryptanalyst Alan Turing, who broke the Nazi Enigma code during WW2

---

### 3. Hackers break the rules – machines don't

Malware is always evolving and black-hat hackers are continually learning. You need to keep up if you want to protect your business from their sticky fingers. The typical sales-pitch from a next-gen vendor praises machine learning as the solution to make businesses fit for the fight and clever mathematics means that one can predict an attacker's every move. But sadly, no matter how smart a machine learning algorithm is, it has a narrow focus and, as we discussed, learns from a specific data set and rules.

The simple fact is that, by contrast, attackers don't play by any rules. What's worse, they are able to change the entire playing field without warning.

A hacker can learn context and benefit from inspiration, which no machine and no algorithm can predict – no matter how sophisticated they might be. Malware writers are also able to hide the true purpose of their code by "covering" it with obfuscation or encryption.

For example, an attacker could bury malicious code into the pixel settings of a harmless image file. They could also split the malware into parts and hide it in several separate files. Each of the files appears clean on its own – it's only when they converge on one endpoint or network that they begin to demonstrate malicious behaviour. If the ML algorithm cannot look behind these 'masks', it can make a wrong decision, labelling a malicious item as clean – causing a potentially dangerous miss.

Machine and human, then, need to work together in order to proactively prevent and mitigate malicious activity.



**Attackers don't play by any rules. What's worse, they are able to change the entire playing field without warning.**

# Do AI and ML change the game?

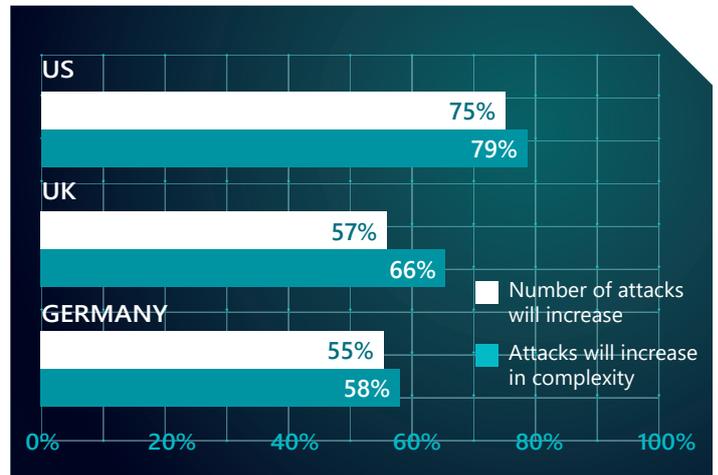
Though ML, with human intervention, is undoubtedly a powerful tool in a business' defence strategy, experts and researchers have also warned about the ways in which attackers have begun to adopt ML techniques to improve or automate malicious activity. Sadly, where there is an advance in technologies that assist in the battle against cybercrime, there is also potential for cybercriminals to use AI to advance their malicious activities.

As such, there is growing concern among IT decisions makers over the way in which attackers will use AI. However, our research shows that, once again, the level of concern varies across the different regions, leading us to question how informed IT decision makers feel about the potential risk AI-powered attacks pose.

For example, US IT decision makers are much more concerned about how AI will increase the number of attacks their teams will have to detect and respond to, and they are more likely to think that AI will make attacks more complex. In comparison, fewer IT decision makers in the UK and Germany thought AI would have this impact on the attacks their organisation faces.

The truth is that hackers could use ML to help profile the victims before it attempts to infect. This may include checking to see if a victim's machine is running in a virtualised environment or being run in such places as a malware analyst's machine. Another persistent question is whether more of these types of attacks could happen in the near future.

Thanks to the scalability and growing efficiency of ML systems - and logically the AI that might follow - it could become significantly more effective and easier to carry out labour-intensive cyberattacks. This includes attacks involving social engineering, such as spearphishing. By automating the non-trivial tasks that attackers need to perform prior to launching these targeted operations, future use of AI could potentially enable more adversaries, and with less effort, to conduct them. Attackers might also be capable of launching sophisticated spearphishing



attacks en masse, while realistic chatbots mimicking 'friends' could add new layers to the threats, too.

Furthermore, vulnerabilities in ML-based systems themselves could be ripe for exploitation. For example, this could take place through data poisoning, whereby attackers work out how the algorithms are set up or where ML gets its training data from, hackers can compromise and manipulate data to mark what is recognised as 'good' or 'bad'.

# Machine + Man

---

The ever-changing nature of the cybersecurity environment makes it impossible to create a universal protective solution, based solely on ML. With a purely ML-based cybersecurity solution, it only takes one successful attack from malicious actors to open up your company's endpoints to a whole army of cyber threats.

This is why other protective layers, as well as humans, need to be involved when implementing ML systems.

In order to keep detection rates high and false positives low, a team of human supervisors can evaluate items that are too divergent from other samples, and hence difficult for ML to label.

Thanks to rigorous training and supervision of humans, ML is able to analyse data, find patterns and identify most of the potential threats posed to organisations. Automation of this process speeds up the security solution and essentially helps your IT teams handle the growing number of samples they see every day.



# Looking beyond the hype

---

While it's nice to believe that a 'silver bullet' to solve all our cybersecurity challenges exists, it's simply not true. Despite what the shiny marketing materials might say, true artificial intelligence does not yet exist and machine learning is still not mature enough to be the only layer standing between you and the cyber attackers.

Over-hyped claims are simply confusing IT decision makers and potentially putting businesses at greater risk. In today's business environment, it would be unwise to rely solely on one technology to protect your networks and data. It's important that businesses are aware that ML has its limitations in order to understand the ways in which you can ensure you've properly secured your organisation.

In the building of your robust and reliable cybersecurity defence, you need to fully understand the challenges your business faces and then consider the solutions that will best meet your specific needs. Every business is unique so a universal solution isn't going to cut it. Multi-layered solutions, combined with talented and skilled people, will be the only way to stay a step ahead of the hackers as the threat landscape continues to evolve.

If the past decade has taught us anything, it's that some things do not have an easy solution, especially in cyberspace, where change comes rapidly and the playing field can shift in a matter of minutes. Rather than looking to AI and ML as the silver bullet, look beyond the hype and focus on what is right for your business; where are the most vulnerable points and how can you make sure these don't act as backdoors to malicious actors? Do you know where your most sensitive data resides and how are you ensuring it is protected?

Machine learning, alone, is not the answer. It is invaluable in detecting malware but it's critical that we manage IT decision makers' expectations of what the technology is capable of doing. The game can change at any point and you need to make sure you have properly deployed and managed your defences to keep the bad guys out.

Despite what the shiny marketing materials might say, true artificial intelligence does not yet exist and machine learning is still not mature enough to be the only layer standing between businesses and the cyber attackers