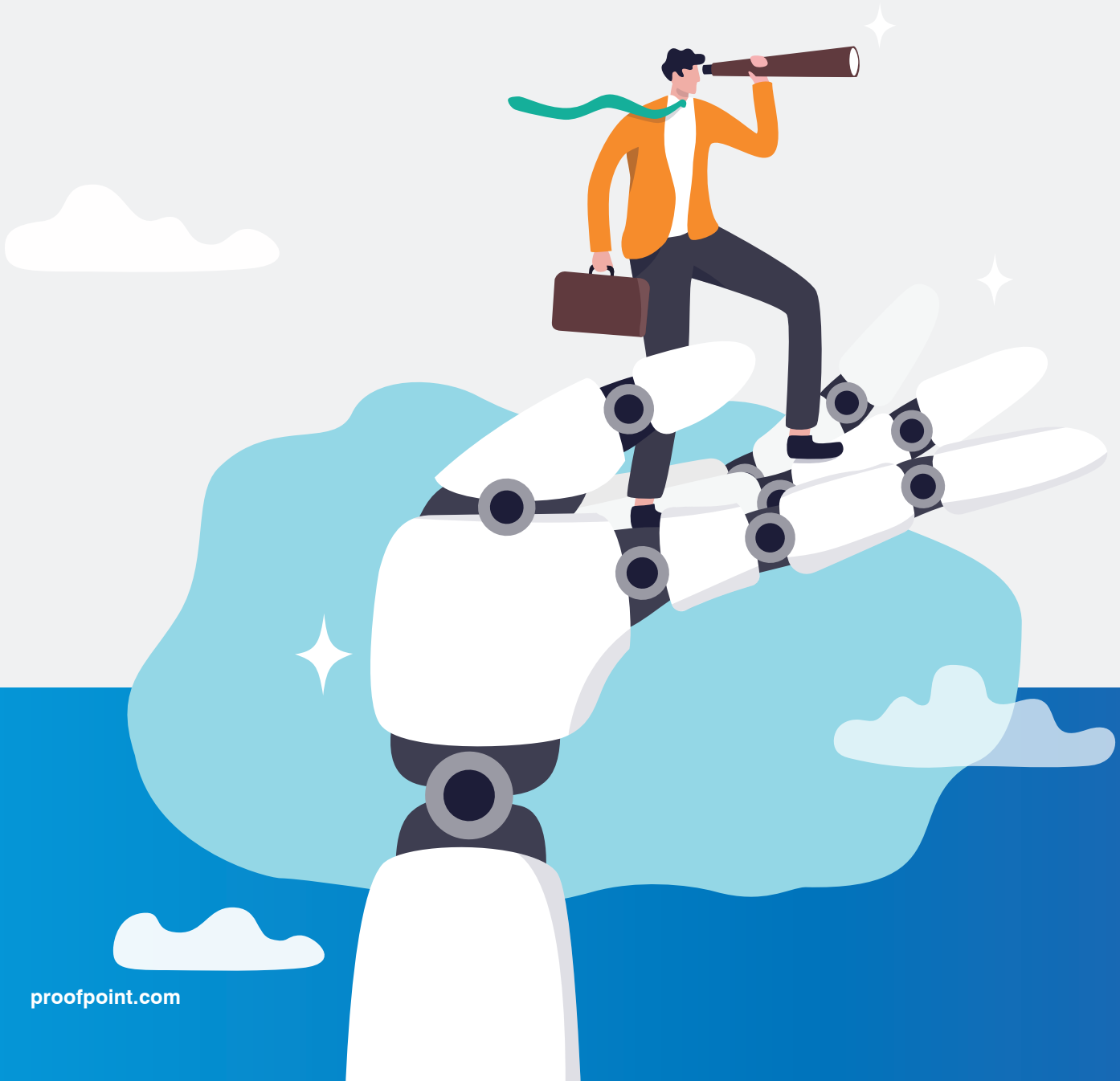
proofpoint.

EBOOK

Cybersecurity's AI Tidal Wave

Fact, fiction, and future trends for infosec teams



Introduction

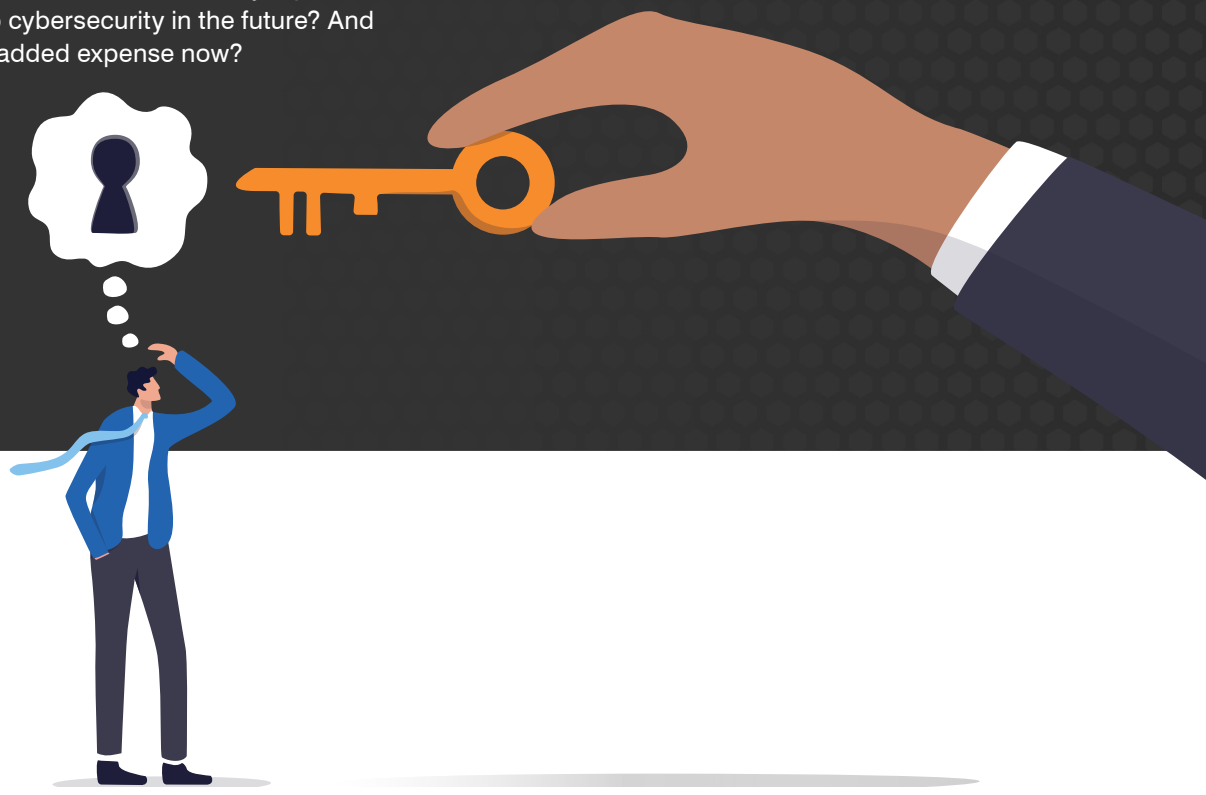
Artificial intelligence (AI) in cybersecurity has been around for well over a decade, and in that time a lot has changed. Years of trial and error using real-world data have helped researchers refine and advance AI models and algorithms to the point that, in many ways, today's AI is a miracle of engineering.

Instead of depending solely on signatures and rules, AI can learn from threats and identify ones that it has never seen before. And applied the right way, it can be fast. When every second counts, it can be a game-changer for security teams working to stop attacks.

Yet with all that AI can do, there's still more work to be done. What's more, many in the industry are still confused about what it's capable of right now and how it will be used moving forward. Will AI tools wholly replace traditional approaches to cybersecurity in the future? And what tools are worth the added expense now?

Answers to these questions and others like them vary depending on whom you ask. Certainly, AI is an incredible asset when it comes to detecting and responding to threats and mitigating risks. However, that doesn't mean it's always the best approach – at least not yet. At this point, the key is to carefully think through how AI can be applied to different problems.

In this e-book, we'll explain where AI in cybersecurity works best. First, we'll give you some basics about what AI is. And then, we'll give you some background and details about how it works so that you can scrutinise AI claims and ask deeper questions about whether AI is the right choice for your needs.



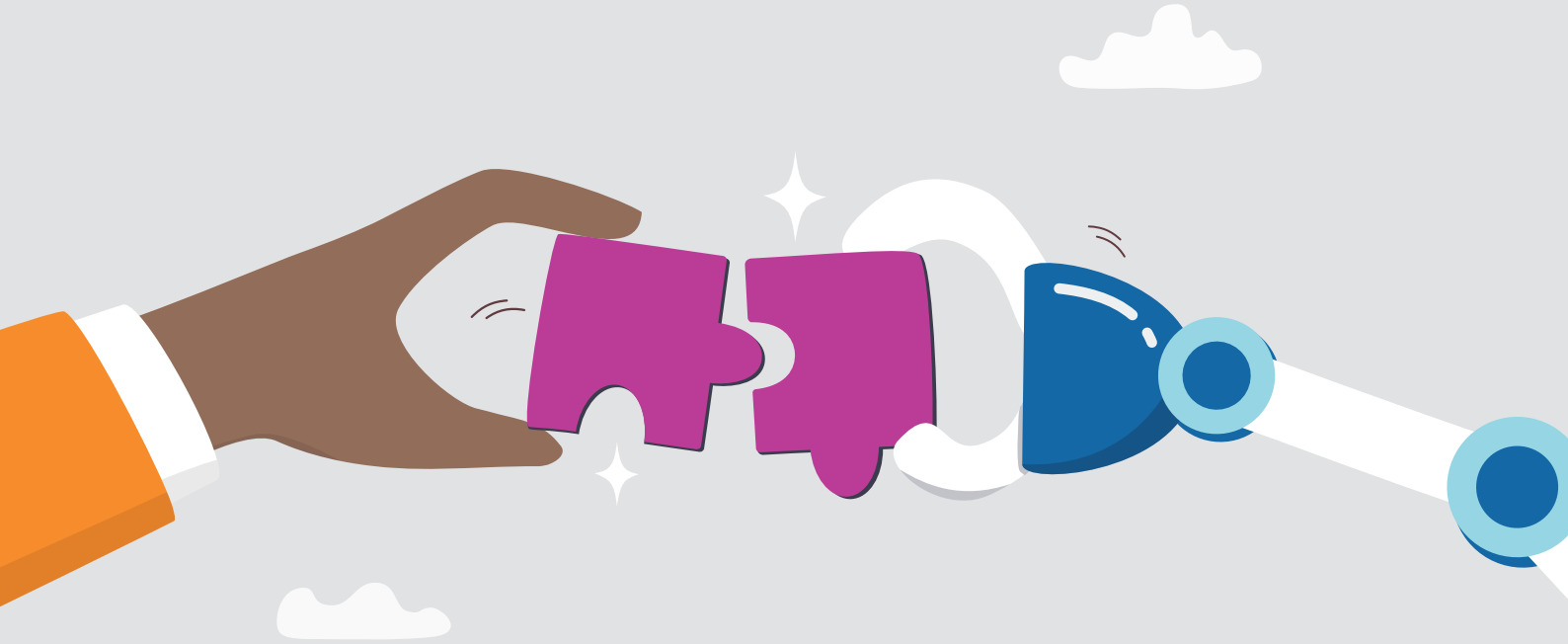
SECTION 1

The Lowdown on AI

When the topic of AI is discussed in cybersecurity, a lot of terms are used interchangeably. That's partly because there are many AI technologies that are all interrelated.

Here, we'll cover the three most frequently referenced ones so that later–when we switch between talking about AI and ML – you'll understand why.

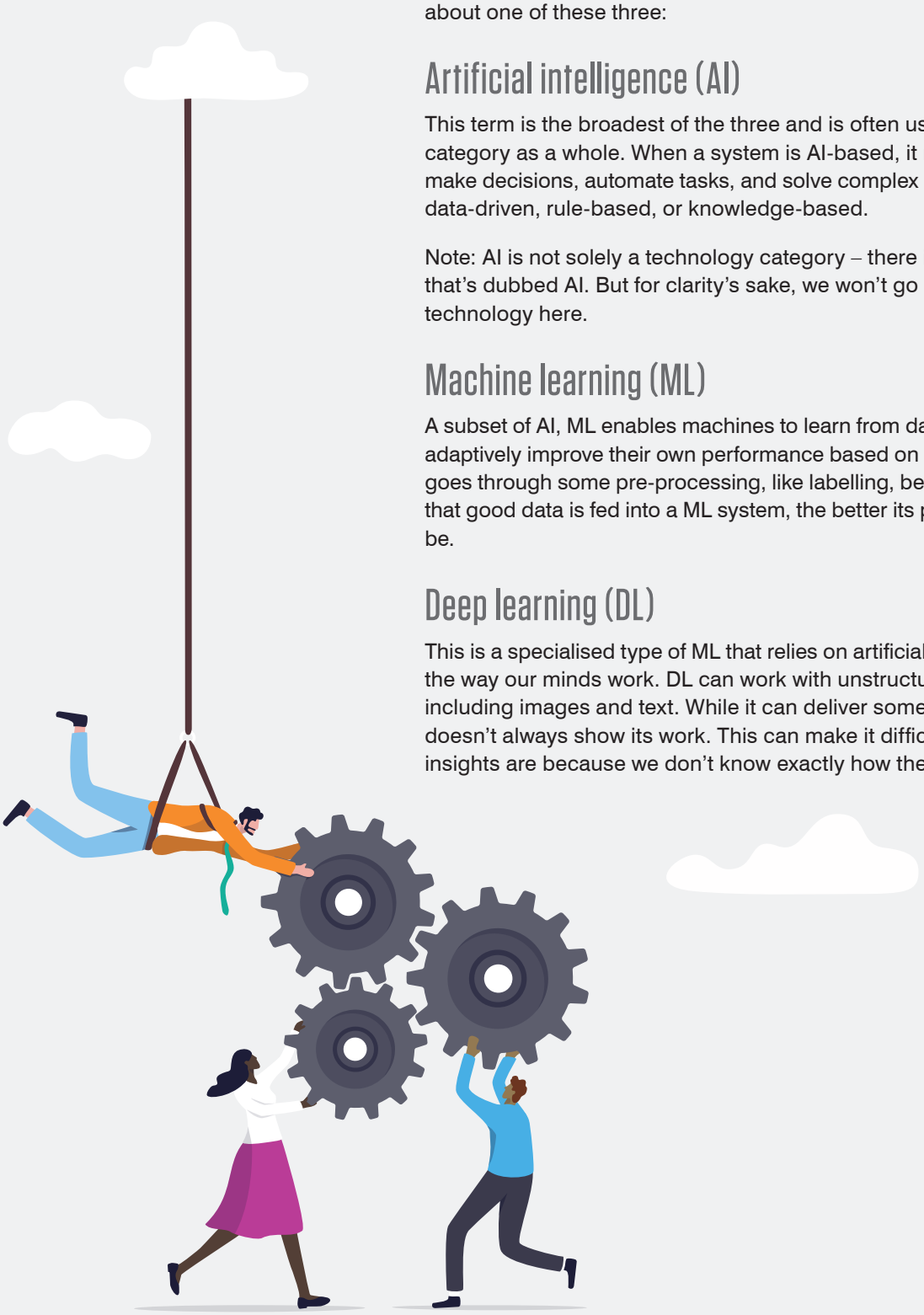
We'll also sketch out a few basics about how cybersecurity tools work so that you get a sense of what makes traditional tools different from those that are AI-powered.



Think big to small

Intelligent technologies share a lot in common, but each one is also unique. To keep the terms straight, it helps to picture them as a series of circles with each one encompassing the next. AI is the biggest, ML is a little smaller, and so on.

When vendors reference AI in their cybersecurity tools, they're typically talking about one of these three:



Artificial intelligence (AI)

This term is the broadest of the three and is often used to describe a technology category as a whole. When a system is AI-based, it mimics human intelligence to make decisions, automate tasks, and solve complex problems. AI systems can be data-driven, rule-based, or knowledge-based.

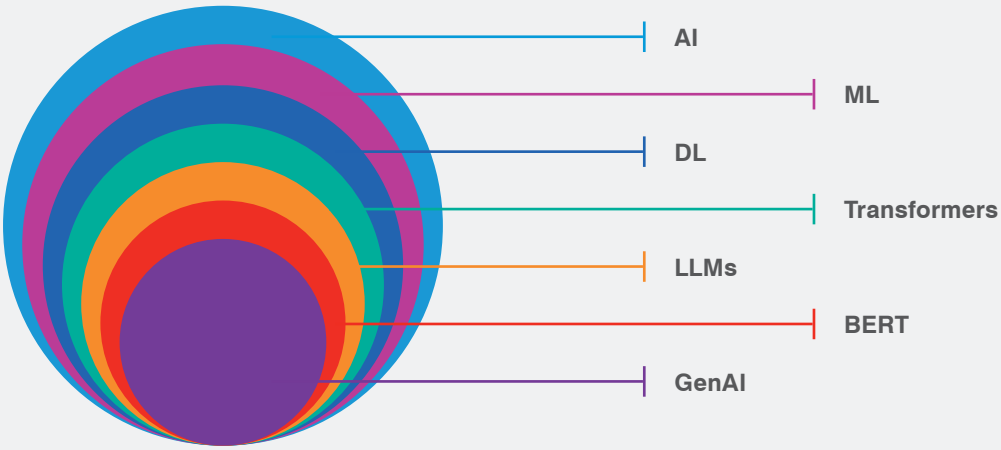
Note: AI is not solely a technology category – there is also a group of technology that's dubbed AI. But for clarity's sake, we won't go into the subcategories of that technology here.

Machine learning (ML)

A subset of AI, ML enables machines to learn from data on their own and adaptively improve their own performance based on data. However, data typically goes through some pre-processing, like labelling, before it can be used. The more that good data is fed into a ML system, the better its predictions and decisions will be.

Deep learning (DL)

This is a specialised type of ML that relies on artificial neurons, or nodes, to model the way our minds work. DL can work with unstructured, unprocessed data, including images and text. While it can deliver some very interesting insights, it doesn't always show its work. This can make it difficult to know how reliable its insights are because we don't know exactly how the system got there.



Glossary

Artificial intelligence (AI). AI-based technology mimics human intelligence to make decisions, automate tasks and solve complex problems.

Machine learning (ML). ML is a subset of AI. It enables machines to learn from data on their own and adaptively improve their own performance based on data.

Deep learning (DL). DL is a specialised type of ML that uses artificial neurons, or nodes, to mimic the human brain's learning processes.

Transformers. As a subset of DL, transformers use artificial neural networks to learn context and meaning by following the dependencies and connections between sequential data like the words in a sentence.

Large language models (LLMs). LLMs are built using transformer models and thus can understand how characters, words, and sentences function together.

BERT. This was one of the first LLMs. It helps machines understand the meaning of ambiguous language in text by using the text nearby to establish context.

Generative AI (GenAI). Generative AI is a type of LLM that uses various types of content – including text, imagery, audio, and synthetic data – to create new and unique outputs. ChatGPT and DALL-E are two examples.

Other key terms

Data poisoning. This is when attackers inject malicious data or inputs into a model's training dataset, which causes the model to make incorrect predictions, miss detections, or create a flood of false positives.

Hallucination. When an LLM creates false information or outputs that are nonsensical, that's a hallucination. Because LLMs are designed to produce coherent text, they can be difficult to spot.

Confabulation. This term is for when an GenAI generates a video, an image or an audio that is false or distorted – it's essentially a form of hallucination.

What's under the hood

It can often be hard to understand what makes one cybersecurity tool different from the next – and where AI fits in. It all becomes a lot clearer once you understand some basics about how they work.

One source of confusion has to do with the technology they use. AI tools can use rules and signatures, which are the foundation of traditional security solutions. But they don't always. And they definitely use ML and algorithms.

Static rules / signatures

A system that uses static rules or signatures is programmed by humans to make specific deductions or choices. So, for example, it may be told that certain behaviours on the network are malicious or that links with certain attributes are a threat. Then, these rules are used for a while until they're changed again by humans. The system doesn't automatically evolve on its own to detect new threats.

While rules might seem outdated, they're also fast, easy on computing resources and highly effective for certain aspects of threat detection. In fact, signals, such as email sender reputation and IP addresses, can even be as effective as AI for many detections.

And keep in mind that static rules and AI aren't mutually exclusive. Often, the best AI-powered cybersecurity tools integrate static rules and signatures to speed up threat detection. (That's something we'll get into later.)

ML / algorithms

Traditionally, when you program a computer to do a task, you use algorithms to tell it exactly what to do and how to do it, step by step. Then ML came along.

With ML, you don't explicitly program a system to do specific tasks. Rather, you feed it examples and provide programming so that it can learn from those examples. This enables it to figure out behaviours that all the examples have in common. It can then use these generalisations to make predictions and take actions in the future when it sees something new, but similar, to what it has seen before. In other words, it uses what it learns to improve its own performance.

SECTION 2

AI Is No Silver Bullet

It's clear that companies are all in on AI technology. In 2023, global spending on AI was projected to reach $154 billion – that's 26.9% more than what was spent in 2022.[1] But in this frenzy to avoid getting left behind, it's easy to get lost in a morass of hype and untested claims.

Certainly, AI marks a fundamental shift in how cybersecurity is done. In fact, it's an indispensable tool. And there are a lot of good reasons to jump on the bandwagon and invest. In a threat landscape that's evolving ever faster and characterised by stealthier attack techniques entering the fray every day, you need AI to give you an edge.

However, that doesn't mean AI will solve all your security problems. We're not there yet – and we may never be. There are technological limitations, which means that it's still important to think through how and where you use it.

1 IDC. "Worldwide Spending on AI-Centric Systems Forecast to Reach $154 Billion in 2023." March 2023.

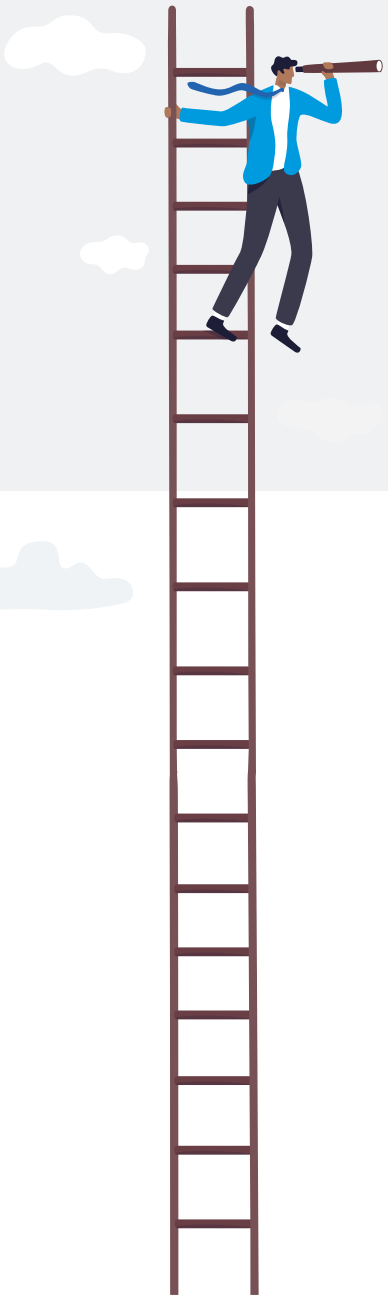
AI is good at many tasks

When it comes to cybersecurity, what's so different about AI and ML tools? What can we do now that we couldn't do before? This is a general overview of a few tasks that AI excels at:

- **Analyse large datasets and connect the dots between seemingly unrelated phenomena.** AI can uncover behaviour patterns and flag suspicious activities that might otherwise be missed. It can generalise about what it learns and recognise similar threats.
- **Make predictions and infer intent.** AI can predict whether something is a threat or not. It can deduce, or "infer," if there's threat – even if it has never seen that particular threat before.
- **Prioritise risk.** Teams are often inundated with security alerts, which can lead to threats being overlooked. AI can sift through the barrage of notifications and single out the ones that need immediate human review.
- **Automate maintenance, hygiene and response.** AI can take over many tedious, but essential, security tasks. This frees up the security team to work on more pressing challenges.

Not all AI is created equal

As security vendors continue to flood the market with new AI-powered tools, it can be difficult to tell what – if anything – makes one better than all the others. What sets apart good AI from great AI? In basic terms, there are two big factors: the training models that are used and the size and quality of the datasets. Here's why.



Datasets matter

Fundamentally, an ML model's performance depends on the source and quality of the examples it is fed. That's because ML learns from examples and patterns. So, the more data – and the higher the quality of that data – the better. If your data is wrong, your models will be biased. And if the dataset is too small, it may not reflect the real-world threats in your environment.

How many examples or much data you need will vary depending on the type of task you're doing. Sometimes a system needs tens of thousands, hundreds of thousands, or even millions of examples before it can reliably identify malicious behaviour in an environment.

The cycle of ML models: rinse and repeat

Essentially, a vendor's competitive advantage comes from the efficiency of the "factory" it has assembled to build, monitor, and maintain its ML models. In other words, there needs to be a thoughtful process behind the scenes.

Different types of ML models require varying degrees of human input. This is just a simplified sketch of what it takes to keep one type of model primed:

- First, you collect the examples that you want the system to learn from, and you label each one. This is a task best performed by expert threat analysts working with data scientists, rather than data scientists alone.
- Next, you feed those examples into an ML algorithm to construct a model.
- Before you use that model, you need to perform a series of iterations to ensure that the model performs at the level you want it to.
- Then, as new examples come in – which is all the time because the threat landscape is constantly changing – you have to do this process again and again, often multiple times per day.

If that sounds time-consuming and labour intensive, it is. And it's one reason why ML isn't always the right choice for solving a problem and why a vendor's breadth of capability matters.

When it comes to general-purpose AI applications, finding data is easy. It's all over the internet. But threat data – especially data that's well-suited for the type of ML model a vendor intends to use – is scarcer. It's a lot harder to collect samples of malware than it is to get data that's used in applications such as image and natural language processing for applications like ChatGPT.

There are a couple reasons for this. For starters, not much attack data is publicly available because most security vendors hold on tightly to the threat data they collect. That's for good reason. Not only does it have obvious competitive advantages, but threat data is sensitive and it comes with a bevy of privacy concerns. As a result, few vendors have a dataset large enough to train (and retrain) models accurately.

Training models matter

Training models – and how they're applied – come a close second to data in terms of importance. The reason why is that an ML model is fundamentally a summary of a dataset.

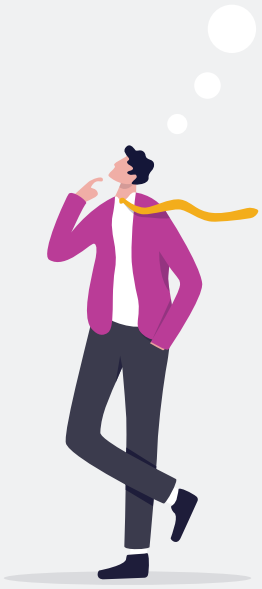
And bigger models aren't necessarily better. If a model is targeted on a very specific area of the threat landscape, it doesn't need to be big. However, when it comes to general computation ability, bigger does tend to lead to better results – as long as you have sufficient data to train it on. If you don't, sometimes a model can essentially memorise data and just parrot back what it has been given.

A smaller model that's trained on known "bad" emails and files will likely be faster and cheaper than other approaches, but it won't have the data to be able to adapt to a constantly changing threat landscape. In contrast, a larger model that uses a self-learning approach will likely be more adaptable, but it will take much longer to train and it will be more computationally (and therefore financially) more costly.

Two are often better than one

When you compare the strengths and weaknesses of ML and signatures side by side, you can see why some cybersecurity tools combine them. Rules, signatures, and other signals can be a powerful complement to ML. They offer:

- **Model training.** Rules provide ready-made labels for datasets.
- **Speed.** When AI is used alone, it can be slow to learn about new cybersecurity trends. Rules and signatures ensure a system can respond quickly to the threat landscape.
- **Context and domain knowledge.** ML models can lack or struggle to learn context and domain knowledge. For example, rules, and signatures can encode policies, regulations, standards, and best practices that apply to particular business domains. They can also capture specific traits, behaviours, and patterns of threats that are relevant to an organisation, industry or geography.



Not everyone uses AI intelligently

AI can be expensive, so it's important to think through how and where you apply it. That requires asking some key questions.

Is applying AI to this problem necessary?

AI and ML undertake multiple advanced processes at once, which requires a huge amount of computing power. But sometimes a problem simply doesn't call for that kind of effort. So when a cybersecurity vendor touts its tool's AI capabilities, that doesn't necessarily mean that the tool is better. AI might just make it slower and more costly.

Is this best place to use AI?

In cybersecurity, every second counts. A central challenge is to identify threats in real time and block malicious content before it's delivered. If AI slows down processing time or if it's relegated to scanning potential threats after they're delivered, then that's a major drawback.

Would this task be better solved with ML or signatures?

ML excels when it comes to processing vast amounts of numbers, and abstracting patterns and concepts. It's also language-agnostic. Together, this means that it's great when it comes to analysing content and user behaviour – something that's especially important for finding BEC threats. However, ML struggles when it comes to detecting malicious URLs and attachments without a sandboxing stack. And it generates a lot of false positives when it's used alone.

In contrast, signatures are right every time and they don't generate any false positives. But they require high domain expertise. And they're limited in that they don't detect anything outside of the specific behaviour that they have been created for. Because they're more surgical, rules and signatures often a good tactic for addressing known and common cyber threats.

Machine Learning

- automatic
- moderate/low domain expertise
- high true positive (TP) rates on similar samples
- low, but non-zero false positive (FP) rates (1:10000)
- weakly interpretable

Signatures

- manually intensive
- high domain expertise
- low TP rates on new samples
- 100% TP rates on known samples
- 0% FP rates
- interpretable

Some pros and cons of machine learning and signatures.

SECTION 3

What's Hype, What's Reality

Hype doesn't always match the reality once AI technology is deployed in complex environments. That's why it's important to go in with your eyes wide open. Here, we'll dispel some common misconceptions.

Hype: AI is better than ML

Reality: ML has a limited scope that's focused on analysing and learning from large volumes of data. When a problem is well-defined, ML is often faster and less expensive to implement than some other categories of AI. It's all about finding the right tool for the job.

Hype: ML detections make signatures and rules obsolete since they aren't brittle

Reality: Generally, this is true. Rules are much more brittle than ML. But ML can still be evaded with targeted evasion techniques. In fact, good rules or signatures are perfect for well-characterised known threats. Think about a file hash – it always and only matches the file that it's intended to catch. It has a 100% true positive rate and 0% false positive rate for that file, except in the extremely rare case of a checksum collision. (MD5 collisions are much more likely than SHA1/SHA256 but the latter is still a theoretical possibility.)

In contrast, ML is mostly based on statistics and numbers. It's good at generalising and identifying patterns that mean something might be a threat. So it's not surprising that tools for email security that use ML alone suffer from high false positive rates. They deal with a vast majority of one class of data – like emails – and a minority of other classes of data. That's why tools often combine ML with traditional signatures and rules.

Catch-22: AI and the false-positives issue

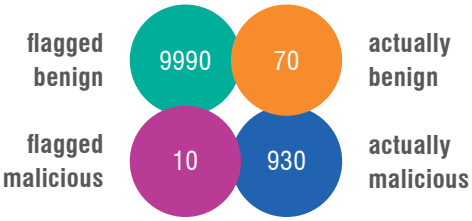
A perfect system would have a 100% detection rate and 0% false positive rate. In reality, there's always a tradeoff between those two factors.

- With AI, a high true positive rate will always correspond to a high false-positive rate and vice versa.
- Given the scale of today's threats, a seemingly small false positive rate can quickly overwhelm security teams. In other words, false positives can be almost as harmful as false negatives.
- In the example below, a false positive rate of 1:1000 means that 10,000 benign samples will result in 10 false alerts for security teams to track down and verify. In a more typical environment, with hundreds of thousands of samples, that's hundreds of false alerts.

Risk = likelihood x cost

What's the cost to your organisation of 1 FN? How often does that to happen?

What's the cost to your organisation of 1 FP? How often does that to happen?



Hype: ML is inherently better than humans at finding and stopping intrusions

Reality: Yes – ML can show impressive detection results when it comes to finding and stopping intrusions. And this can be tracked back to some of its strengths. Firstly, it generalises, which means it excels at detecting related-but-different samples or techniques. It's also faster than humans, and it scales better.

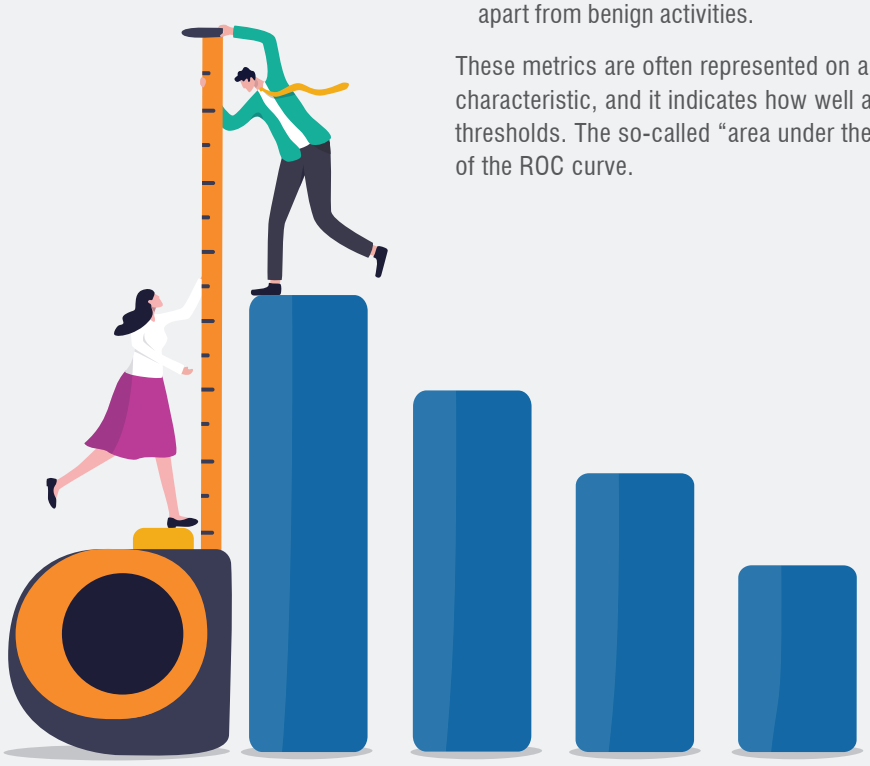
But ML falls behind when it comes to some key tasks. For starters, it's usually not very good at extrapolation. So, if there's a novel corner case or something that's not in the training set, ML won't understand it. It's also not great at "explainability," which means that when ML tries to show its work, its efforts are often superficial. Compare that to humans, who are very good at understanding the thought process behind the steps that were taken to identify an attacker's behaviour.

How ML model performance is measured

With AI/ML, you need to have an objective mechanism for evaluating your models. This is where performance metrics come in – they give you a sense of how good a model really is. Experts typically use two key metrics to judge ML models:

- **Accuracy.** This is the proportion of correct predictions made by the ML model out of the total number of predictions. It tells you how well the model can correctly tell cyber threats apart from benign activities.
- **Precision.** This is the proportion of true positives (correctly identified cyber threats) out of the total number of positive predictions (of all identified cyber threats). Precision measures how well the model can avoid false positives.

These metrics are often represented on an ROC graph. ROC stands for receiver operating characteristic, and it indicates how well a model detects cyber threats at different thresholds. The so-called "area under the curve" (AUC) measures the overall performance of the ROC curve.



Hype: The type of ML model is the most important factor in the performance of a system

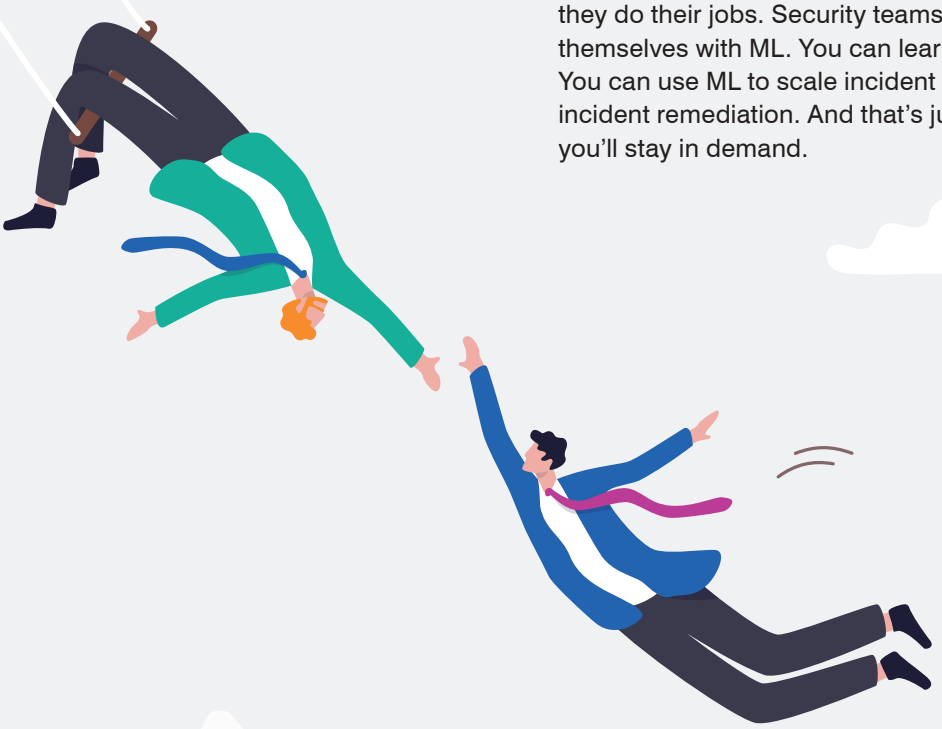
Reality: There are a lot of ML concepts that are tossed around in the industry in an effort to make some ML systems sound like they're better than others. "Deep learning," "gradient boosting," and "Bayesian methods" are just a few. In truth, the type of ML model is not the most important factor.

As we said above, ML model's dataset is what matters, and the training model itself comes a close second. That's because an ML model is essentially a summary of a dataset. Its behaviour is inherited from a dataset's richness and relevance, which is why plentiful, contemporary examples are so important. Another key factor is how the data is labelled and distributed. Better labels mean better performance.

Hype: ML replaces the need for skilled employees

Reality: ML has been around for a while now, and yet security incidents only continue to rise. If ML could replace skilled employees, we wouldn't keep seeing headlines about the security workforce shortfall. In 2023 the cybersecurity workforce gap reached hit an all-time high – there are 4 million more jobs than there are skilled workers.[2]

Even so, that doesn't mean ML won't impact security professionals and how they do their jobs. Security teams should make it a priority to upskill and scale themselves with ML. You can learn how to interpret ML detections in context. You can use ML to scale incident discovery. You can provide a human touch in incident remediation. And that's just for starters. The trick is to stay curious and you'll stay in demand.



2 ISC2. "ISC2 Reveals Growth in Global Cybersecurity Workforce, But Record-Breaking Gap of 4 Million Cybersecurity Professionals Looms." October 2023.

SECTION 4

7 Key Questions

As the market explodes with new entrants in the space, it's increasingly important – and difficult – to carefully evaluate the effectiveness of each AI/ML-based offering.

Even the most reliable cybersecurity vendors may have aggressive or misleading marketing that clouds good buying decisions. Their ML models may have flaws or limitations that they are not aware of or do not disclose. And sometimes, the sales rep may simply not fully understand AI technology.

By asking the right questions, you can get a better handle on how they'll keep your people, data, and environment safe.

Red Flags

Too often, AI/ML claims are vague or exaggerated because the technology is complex and marketers may not fully understand it. Moreover, statements that are technically true can still be misleading or not applicable to your unique environment.

So keep an eye out for some common red flags:

- **Misleading claims.** Using the term "AI" or "ML" to describe any kind of automation or data analysis – even if it does not involve any learning or adaptation from the system.
- **Ambiguity.** Claiming that a product or service is powered by "AI" or "ML" without providing any evidence or explanation of how the technology is used or what value it adds.
- **Buzzwords.** Using hype and buzzwords like "smart", "intelligent", "advanced", or "revolutionary" to describe a product or service that uses AI/ML, without specifying what makes it so.
- **Big promises.** Making vague or unrealistic promises about what a product or service that uses AI/ML can do.
- **Fear tactics.** Appeals that exploit an asymmetrical technical understanding of AI and ML.

1: Why is AI/ML important for this security problem?

On one hand, AI/ML can help analyse large amounts of data and find anomalies, trends, and behaviours that indicate potential attacks. And they can automate response and mitigation of security incidents.

But they can also require enormous amounts of computing power depending on the size and complexity of the learning model. What's worse, execution time can be much longer than less complex approaches such as rules and signatures.

Your vendor should have tested its models against other analytical techniques to ensure that the AI is more than a marketing gimmick. So ask for an explanation about exactly why AI is right for your use case.

2: Where do you get your training data?

Not only should vendors have access to a lot of data, but they should also have access to wide-ranging and diverse data. And the data should reflect the world you want the system to learn from. You want the data to accurately represent the threats that are targeting the group or industry or users that you're trying to protect. Few cybersecurity vendors have a dataset large enough to train (and re-train) models accurately. So it makes sense that this is an area where new market entrants often struggle.

3: How often do you update your ML model to adapt?

Cyberattacks are constantly evolving. And ML models need to adapt to be able to see these new threats. That's why regularly updating them is critical. ML models used in production detection stacks should ideally be updated multiple times per day. If they're not, they'll no longer reflect real-world data, which means they're less effective over time. In data science circles, this issue is known as model drift.

Beyond model drift, sometimes attackers try to fool or manipulate ML models by injecting malicious data into them, which is called data poisoning. Or they might craft inputs that are misclassified by a model, like adding noise to malware samples, to evade detection.

4: What is your training distribution compared to my deployment distribution?

Your vendor should have trained and tested its ML models with datasets that are large and diverse enough to measure expected performance for real-world threats targeting your particular environment.

But typically, vendors train their models based on an environment they know – and they don't know your environment. So there's always a mismatch. They should be able to provide a good explanation of how they bridge that gap.

5: What metrics do you use to quantify performance?

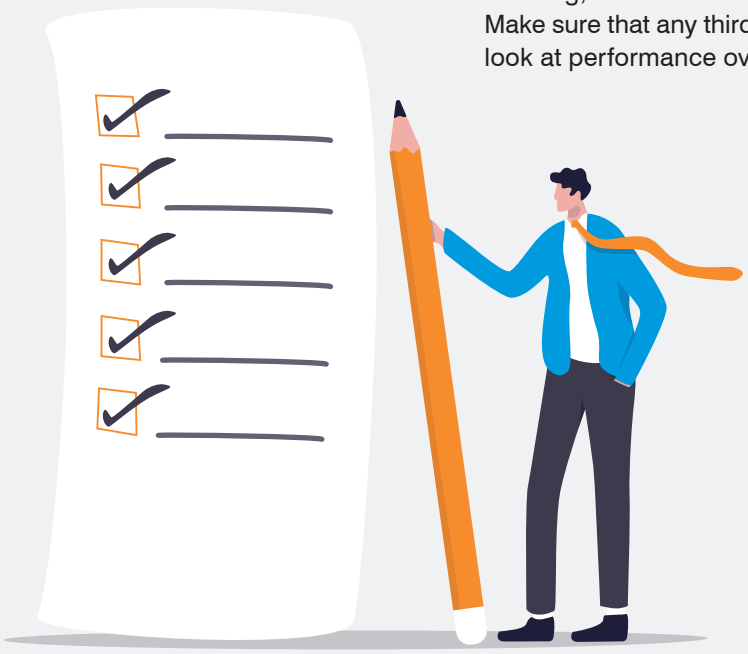
Experts typically use two key metrics to judge ML models: accuracy and precision. These metrics are often represented on an ROC graph. Some vendors also use metrics like efficacy and efficiency. So make sure you understand all the metrics the vendor uses in its ROC curve and discuss all the implications for your security team.

6: Do you supplement the ML with blocklist/ safelist, signatures, or other mechanisms?

To solve the ROC curve, most vendors supplement AI/ML detections with rules, signatures, and other means. You should understand what mechanisms the vendor uses to solve this issue. A detection stack that relies too heavily on AI could be a problem. For many threat scenarios, the technology is slower, less efficient, less effective, and less reliable than other defensive layers.

7: Has this been validated by a reputable third party?

Third-party validation is critical. An unbiased assessment can help you compare and choose the best solutions for your unique needs and environment. But keep in mind that validation is not a one-and-done event. Cyber threats are always evolving, so ML models, products, and services need constant improvement. Make sure that any third-party assessment of the vendor is recent and be sure to look at performance over time.



SECTION 5

Conclusion

As cyber threats continue to evolve at an unprecedented pace, many organisations are turning to AI in hopes of keeping up. While these advanced technologies hold immense promise, they're also more complex and far less efficient than traditional threat detection approaches. The tradeoff isn't always worth it.

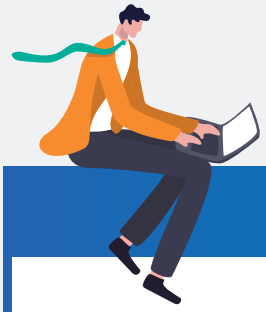
On the one hand, the technologies can help analyse large amounts of data and find anomalies, trends, and behaviours that indicate potential attacks. And they can automate response and mitigation of security incidents.

But depending on the size and complexity of the ML model, they can also be computationally intensive to maintain. What's worse, execution time can be much longer than less complex approaches such as rules and signatures.

It's important to keep in mind, too, that the cybersecurity vendors with the "best" algorithms or models aren't necessarily the best at finding and stopping threats. Rather, it's the ones that have the most and highest quality raw threat data. But that's not all. They also need to harvest and refine their data from the right resources, and their team must accurately label and understand that data. When it comes to AI, the old saying "garbage in, garbage out" endures for a reason.

From the beginning, Proofpoint has used AI to give you complete and constantly evolving protection against a wide range of external threats. Our NexusAI engine includes trillions of data points to continuously protect your people and your organisation against attacks – before they reach users' inboxes.

Whether you're evaluating a potential vendor for the first time or considering new tools and platforms, you should know how AI and ML fit into the equation. To learn how Proofpoint uses AL and ML, visit www.proofpoint.com/uk/solutions/nexusai.

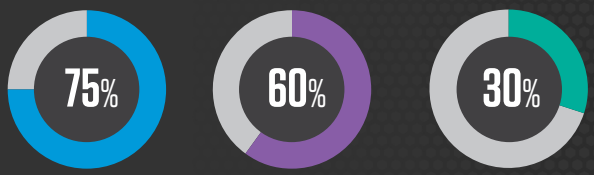


Why Proofpoint

Every day, we analyse more than:

2.6B EMAILS	**49B** URLS	**1.9B** ATTACHMENTS
1.7B MOBILE MESSAGES	**430M** WEB DOMAINS	**143,000** SOCIAL MEDIA ACCOUNTS

We are trusted by more than:



OF THE FORTUNE 100 | OF THE FORTUNE 1000 | OF THE FORTUNE GLOBAL 2000

8,000 ENTERPRISES

200,000 SMALL BUSINESSES

LEARN MORE

For more information, visit **proofpoint.com**.

ABOUT PROOFPOINT

Proofpoint, Inc. is a leading cybersecurity and compliance company that protects organisations' greatest assets and biggest risks: their people. With an integrated suite of cloud-based solutions, Proofpoint helps companies around the world stop targeted threats, safeguard their data and make their users more resilient against cyber attacks. Leading organisations of all sizes, including 85 percent of the Fortune 100, rely on Proofpoint for people-centric security and compliance solutions that mitigate their most critical risks across email, the cloud, social media and the web. More information is available at **www.proofpoint.com**.

proofpoint.

0503-006-02-01 03/24