WEBROOT
Smarter Cybersecurity™

A Layman's Guide to AI, Machine Learning and Its Importance to Endpoint Security

Making Sense of an Evolving Technology

Written by
Hal Lonas, CTO, Webroot
David Dufour, SVP Engineering, Webroot
George Anderson, Product Marketing Director, Webroot

Introduction

"Progress is impossible without change, and those who cannot change their minds cannot change anything."

– George Bernard Shaw

Beginning around 2007, traditional endpoint security was becoming ineffective. Stopping infections was based around finding a user with an infection (patient zero), creating a detection signature (inoculation) and then updating every device to stop any further infections (eradication). The ineffectiveness was a direct result of the volume, variety, and velocity of infections. These factors completely overwhelmed the 'patient zero' approach. There were simply too many patients and not enough inoculations.

While patient zero vendors valiantly did more, and managed to stop whole 'families' of infection using heuristics and advanced signature detection techniques, the fundamental problem didn't go away. Too many devices were getting infected and the cost of remediation was so significant that organizations were creating remediation re-imaging budgets just to cover infection costs.

As an industry, endpoint protection vendors needed to change their minds and do something new to change the game. Webroot was the first vendor to do so by introducing a totally new cloud-based way of countering malware with machine learning at its core.

In October 2011, Webroot launched Webroot SecureAnywhere® in the US retail consumer market. It was the first of the so-called 'next generation' endpoint security solutions, with a revolutionary architecture designed to harness machine learning and high automation to handle the volume and variety of attacks customers were facing.

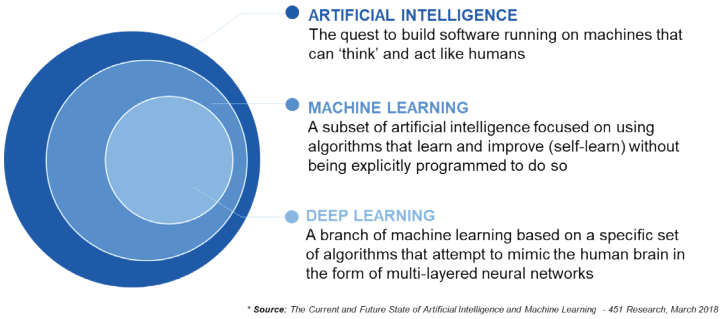
Our considerable experience selling antivirus solutions meant we knew we had to change how we thought about predicting, preventing, detecting and remediating malware – and the best ways to do all of that given the threat landscape we faced. This brief guide will, we hope, provide you with both a snapshot of what Webroot does today to harness machine learning to predict, prevent and protect you against malware, plus give you a better understanding of why this technology is being put to use by next-gen vendors to try to differentiate themselves in a crowded endpoint security market.

Computers Replacing Humans

Webroot has been 'doing machine learning' for more than a decade, and we do consider this a major key differentiator for our own and our threat intelligence partners' solutions. In fact, it's highly likely your current organization is benefitting from Webroot machine learning via our BrightCloud® Threat Intelligence services as we contribute near real-time threat intelligence (TI) data to over 85 other network and security vendors.

However, for many small to medium-sized businesses (SMBs), that doesn't really seem to matter. They have probably heard the terms artificial intelligence (AI) and machine learning (ML), but aren't sure how these advancements are keeping them safe. Then, the many managed service providers (MSPs) we help to provide SMBs with security services are not as knowledgeable about how this technology works, or how it helps their customers either?

The phrase we hear a lot from our customers is 'it just works'. MSPs focus on 'real-life' issues (not the how or why, but the what). They want to know, does the technology really work or not?



* **Source:** *The Current and Future State of Artificial Intelligence and Machine Learning - 451 Research, March 2018*

However, given the amount of 'hype' surrounding AI and machine learning, it's worth knowing a thing or two about these efforts.

Artificial intelligence and machine learning are not the same thing. Marketing campaigns and news articles blur the line and often confuse people into thinking that they are. They are not, and it's important to know the differences so you can understand how each helps to make your cybersecurity stronger.

What is artificial intelligence?

Artificial intelligence interacts with people, for instance emulating a human as a 'chat bot'. The AI component is that interactive component—the thing you can touch, feel, and see. AI technology is very nascent in its cybersecurity application and we expect great things from it in the future. There are of course many hurdles still to be overcome in making a computer act like a human. Right now the programming is restricted to applications like driverless cars, but that is a defined set of circumstances and a singular application being programmed, and still it has involved a lot of software engineering.

True AI would be far more self-learning in its interactions. A good way to define it is that AI is the creation of software running in a machine that can 'think and act independently' and, in doing so, completely emulate a human being.

What is machine learning?

You can think of machine learning as artificial intelligence's nerdy cousin. Machine learning models are designed to analyze data collected behind the scenes, with no human interface. Machine learning is the heavy science. It's where all the data crunching takes place.

A good way to think of machine learning is as a subset of AI focused on using, as some vendors call it, 'math'. But in reality we are talking about algorithms that self-learn and improve their findings and results without being explicitly programmed to do so. Machine learning is now used extensively in cybersecurity, but has an effective and proven track record with only a few vendors.

What about deep learning?

To be thorough, we need to mention deep learning. It's another major technology that Webroot uses. Deep Neural Nets have been around since 1975, but only started to emerge around 2007 with the increased

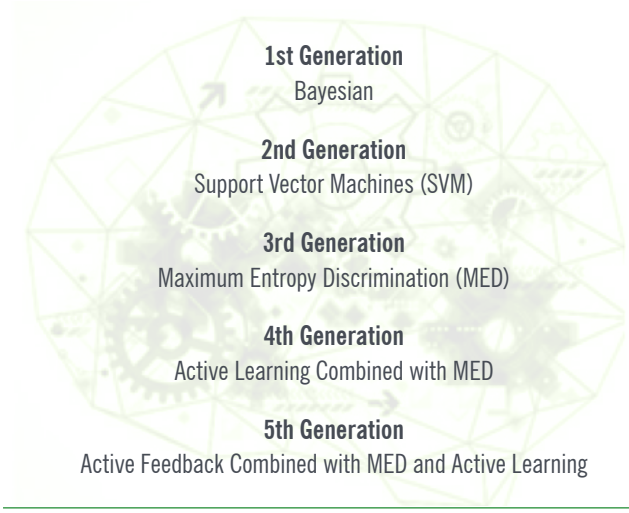
availability of affordable and powerful hardware. This subset of machine learning is about improving the 'training' of machine learning models further by mimicking the human brain with multi-layered neural networks to get 'better' models.

The best and only way to counter malware today

If you strip away the superfluous, the issues Webroot and others are trying to solve using machine learning are clear. Malware and other threats are constantly evolving, their volume is mostly increasing, and the ability to predict and stop zero-day threats is essential.

Machine learning is currently the best and, from Webroot's perspective, only way to tackle these issues. With the right quality and quantity of data you can train and use machine learning to learn directly from data and predict the likelihood of malware, a behavioral anomaly threat, and lots more.

Machine Learning is the best way to do this, as it adapts automatically to changing and evolving environments, a trait that's so essential when todays' attacks are polymorphic and in constant change to avoid detection. Lastly, it's an issue of scale. Because, unlike humans who are limited in capacity, get tired, make mistakes and get overcome by volume, machine learning is tireless, highly scalable and makes far fewer mistakes.



Fifth-Generation Machine Learning

Advanced machine learning using multiple sources means objects like URLs, IPs, files, apps and other data components are now classified far faster and more accurately than could be done by any human army of threat researchers. It allows Webroot to generate and host hundreds of classification models to cover different threat types and content languages. It allows us to publish millions of updates every day on new threats and reputations of existing URLs, IPs, apps and files.

Machine learning's biggest plus is that it's so highly automated. It requires minimal human interaction to produce the highly accurate and timely threat intelligence outcomes that our customers need across multiple threat vectors, often in milliseconds.

Because of Webroot's very early adoption of machine learning, we have fully harnessed a fifth-generation machine learning approach to analyze and produce rich sources of contextual threat intelligence that directly increases the accuracy and capabilities of our own and other vendor partners' security protection.

It's All about Data

When it comes to machine learning and AI, it's important that your vendor has experience and access to both current and historical data. Webroot is fortunate that, for a company of our size, we have a disproportionately large access to both historical and current data to feed our models.

Webroot analyzes half a trillion security events per day, linking and pushing them through our models to enhance our analysis. We have a lot of access to information that new players in the cybersecurity space simply do not. Data quality and volume are both vital to training up a model, but so is the processing power to make it actionable in a timely way.

Webroot uses AWS as our primary Infrastructure as a Service (IaaS) partner. We are currently their tenth largest data business worldwide. We also access the San Diego Supercomputer Center at the University of California that lets us leverage up to 1 terabyte of RAM and 40-50 computing nodes for help with our modeling.

Out of all of this, Webroot publishes over 1,000 machine learning models per day that have typically used over 10 million data points and 20-50 million model parameters.

Why is that rate of modelling important? Well, timeliness is what allows us and our threat intelligence partners to consume and directly benefit from our machine learning models and provide our business customers with better cybersecurity.

Being actionable is an important Webroot edge, too. It's pretty easy to tune new models, but not so easy to deploy the models in a way that allows customers to get meaningful, actionable data from them. Deploying models in the cloud allows us to react much more quickly than if we had to deploy them to endpoints.

What can be Achieved Today – Hype vs Reality?

For everyone reading this guide, this is generally what we've heard:

"Almost all of my technology decisions are based on whether it reduces headaches and is an innovative tool for my customers; so if machine learning does that, I'm all for learning more. I'd be happy to read up on it, but my customers don't have time to read or care about it."

At Webroot, we believe artificial intelligence and real machine learning are able to help all of our MSP, business, and threat intelligence partners in the following key areas:

» Real machine learning and artificial intelligence help create new prediction, prevention, and detection capabilities for the security stack while at the same time decreasing costs and reducing the time to detect and remediate threats.

» Real machine learning helps detect emerging, unexpected threat behaviors quickly, thereby helping security teams, or security orchestration solutions, take action.

» Real machine learning delivers considerable value toward personnel augmentation by building on the skills of human analysts (e.g., it can automate remedial tasks, or simply work around the clock while employees go home and sleep).

Hype vs Reality

Unfortunately there is no shortage of hype around AI and machine learning. Here are a few of the 'fake news' items we've heard recently:

"Sixth-generation artificial intelligence."

There's no such thing. There is, however, a fifth generation of machine learning, and some companies like Webroot are testing sixth-generation capabilities as well.

"Data sources don't matter."

Actually, the source of data does matter. You know not to trust just any old fly-by-night vendor of anything to give you a solid product. You have to do your research and ensure a certain level of quality and reputation. The same should hold for threat intelligence vendor and the data they use and deliver.

"It doesn't matter how long a company has been doing machine learning."

How long a company has been working with machine learning and artificial intelligence is crucial. Quality models take time to tune, and historical data helps guide predictive assessments to prevent emerging and as-yet-undiscovered threats. You can't spin up a new model and expect it to be effective in a week, or even a month. Maturity is a good thing.

This is by no means an exhaustive list of the misinformation out there, but it should provide you a good start against snake oil salesmen. Unfortunately, there is no silver bullet in cybersecurity. No single technology will stop 100 percent of threats.

Employees are going to click on malicious links and use recycled or easy-to-guess passwords, and cybercriminals are going to continue coming up with highly creative ways to get around defenses. After all, the threat landscape is unpredictable.

However, don't be afraid to ask questions of your vendors to see how machine learning and AI are embedded into your security solutions to help protect customers and streamline your business. At the end of the day, you are providing a lifeline to your organization or security clients.

What You Need to Ask – Endpoint Security Vendors

So what sort of questions should you be asking and how can you test the real benefits of machine learning technology against all of the buzz?

1. **Ask questions about the data they learn from**
   How does the vendor get their data? Do they have historical data to track the behavior of a website or URL from the last 60 days, year, or 10-year period? How is the data fed into the security solution(s) they offer? How many attributes or "features" in machine-learning speak do they collect and use to classify things?
2. **Ask questions about update frequency**
   While quality of data is paramount, so is the time it takes to turn that data into something useful and actionable. In cybersecurity, time is of the essence. So, how often does the vendor update their machine learning models? Ideally this should be done at least daily, if not multiple times a day. The longer the period between model updates, the larger the window of vulnerability and the opportunity for spectacular failure
3. **Ask questions about the depth of machine learning defenses they offer**
   Does the solution only offer protection against files or processes? If so, you probably want other security protection layers in place.

   According to Verizon's most recent Data Breach Incident Report 2018. "JavaScript (.js), Visual Basic Script (.vbs), MS Office and PDF10 tend to be the file types found in first-stage malware. They're what sneaks in the door. They then drop the second-stage malware. In this case, it's predominantly Windows executables…Once the first unwelcome guest is in, it's much harder to catch the rest before they execute and wreck the place."

   Malware vectors in the same report were 92.4 percent email, 6.3 percent web and 1.3 percent other, so phishing and the depth of the other web defenses provided are critical too.
4. **Ask how they handle unknown files and internet objects, and when and how threat researchers interact with the machine learning**
   We're not yet to the point where machines can run without human oversight. Human threat analysts need to review unknowns, edge cases, and models' overall behaviors. This is how they fine-tune the algorithms. Their oversight helps avoid false positives. Threat researchers should be thought of as machine learning's teachers.
5. **Ask how the product handles a threat that does get through**
   Does it track what took place on the computer? Can it roll the computer back to a pre-infected state? What is the extent of the remediation? For instance, one well known vendor uses the Volume Copy Shadow Service (VSS) as the remediation back-up, a Windows area that is commonly the first place erased by ransomware or malware!

This is again not an exhaustive list, but it's a good start. The thoughts to hang to when considering AI and machine learning are:

» Data quality, history, and volume are all key to training up an effective and efficient model.

» Consumability, getting fast and easy access to the models, is vital so the security solution is providing timely and actionable protection.

Conclusion

It's pretty easy for vendors to claim they use machine learning in some way. It's not as easy to collect the right data, get accurate machine learning deployed, work out how to train and update models, tie in humans, and to allow customers to glean immediate and meaningful, actionable data from them.

For that reason Webroot believes in a cloud-based machine learning approach with always-on prediction, prevention, detection and remediation. We did not want to be tied to the old self-contained antivirus model where every machine is an isolated island protected by a static defense. That simply means you breach one, you breach them all!

By monitoring continuously and collectively in real-time, everyone benefits from everyone else's data immediately. This is particularly relevant in a world where everyone and everything is connected to the internet and events are happening at internet speed. Even if that internet connection is not possible, having alternative protection built-in that allows that device to run safely and securely until it reconnects is a highly secure way to operate.

For the past nearly seven years, Webroot has been effectively and efficiently protecting millions of consumers and hundreds of thousands of businesses using machine learning to predict and stop malware and lots of other threat vectors. The success of our approach is validated by the trust of not only our customers and their clients, but also many other security and networking vendors who rely on our machine learning and threat intelligence to help protect millions of their customers too.

World Headquarters
385 Interlocken Crescent
Suite 800
Broomfield, Colorado 80021 USA
+1 800 772 9383

Webroot EMEA
6th floor, Block A
1 George's Quay Plaza
George's Quay, Dublin 2, Ireland
+44 (0) 870 1417 070

Webroot APAC
Suite 1402, Level 14, Tower A
821 Pacific Highway
Chatswood, NSW 2067, Australia
+61 (0) 2 8071 1900