Insight from the Security for Business Innovation Council

The Machine Learning Cyber Imperative

Unlocking the Potential of Artificial Intelligence for Security

January 2019

An industry initiative
sponsored by RSA

TABLE OF CONTENTS

SBIC Report

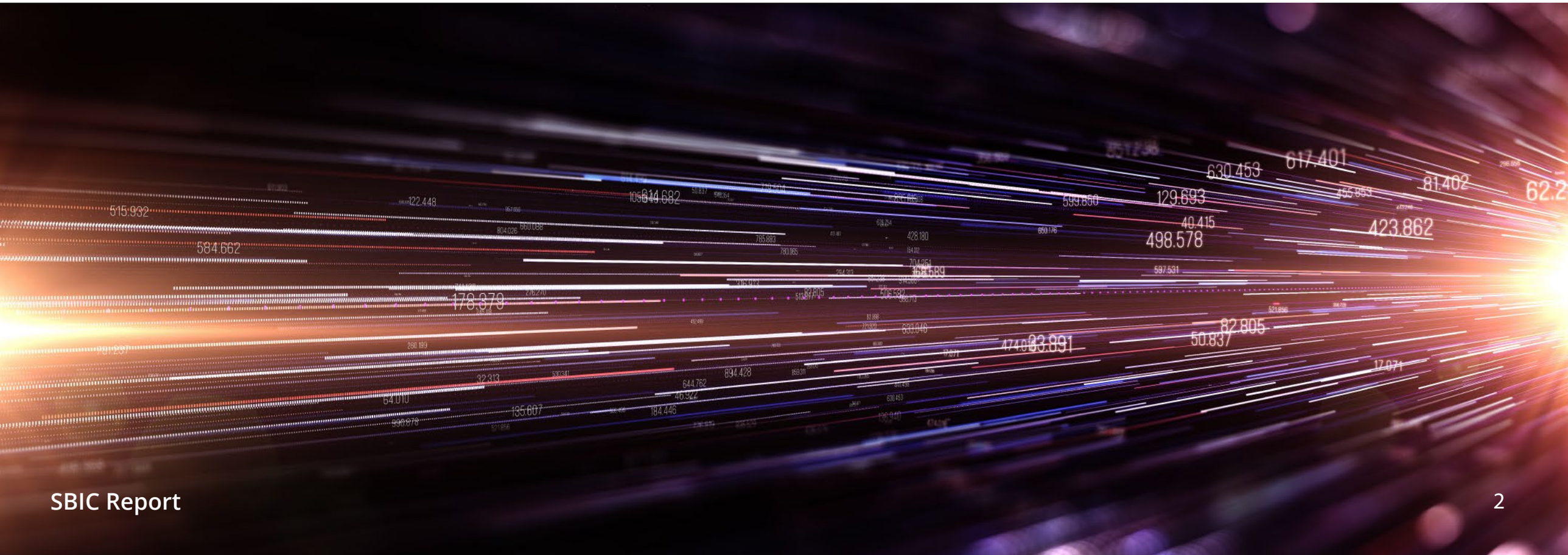
The term Artificial Intelligence (AI) carries some strong connotations. For some, it inspires images of technology gone awry, like the sentient HAL of the 2001: A Space Odyssey film. Or perhaps for the more “glass is half full” types, it inspires the idea of freedom from the mundane as embodied by the character Data from Star Trek: The Next Generation, who captured the idea that machines can be “one of us.” While we are a long way from reaching this level of conscious behavior for machines in the workplace, AI is finally coming into its own as a useful business tool.

While AI is a broad term encompassing myriad aspects of how non-human computation can work, one element of AI gaining significant momentum is its ability to empower systems to perform tasks they are ultimately better suited to handle than people. These tasks typically involve three things: massive amounts of data, different kinds of data, and the need to quickly process this data by effectively identifying similarities or differences within it. This is where Machine Learning, a discrete subset of AI, comes into play for the cybersecurity professional.

THE TIME IS NOW: GETTING UP TO SPEED ON MACHINE LEARNING

Originally coined in the 50s, the term Artificial Intelligence had a rocky start. Initially intriguing, billions of dollars in AI funding from the government in the 60s led to no real practical uses, which in turn led to funding being cut dramatically in the 70s. Interest in AI research ebbed and flowed for decades, sometimes hampered by computing power limitations and other times by the lack of any significant commercial product success.

But recently the concept of applying AI to business has re-captured the imagination of the global business and technology community. Why now? Fueled by vast increases in computational power, machine learning is bringing the once elusive possibilities of AI into sharper focus. As mathematical models are increasingly leveraged in probability, and design theory and AI algorithms are successfully embedded in other systems such as speech recognition and data mining applications, enterprises are taking notice and building AI into their near-term agendas.

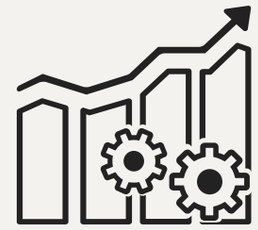


According to International Data Corporation, worldwide spending on AI systems is forecast to reach $57.6 billion in 2021.[1] McKinsey estimates total annual external investment in AI was between $8B to $12B in 2016, with almost 60% of that investment targeted at machine learning.[2] To further support the growing impact of machine learning, patents specifically targeted at machine learning techniques grew at a 34% Compound Annual Growth Rate between 2013 and 2017, the third-fastest growing category of all patents granted.[3]

> *"We think of machine learning as an approach, not merely a tool. Your organization should be at a certain maturity level for this approach to deliver the efficiencies it's capable of."*
>
> **Dr. Sunil Lingayat,**
> Chief of Cyber Strategy and Technology,
> T-Mobile

Machine Learning

What is it Good for?

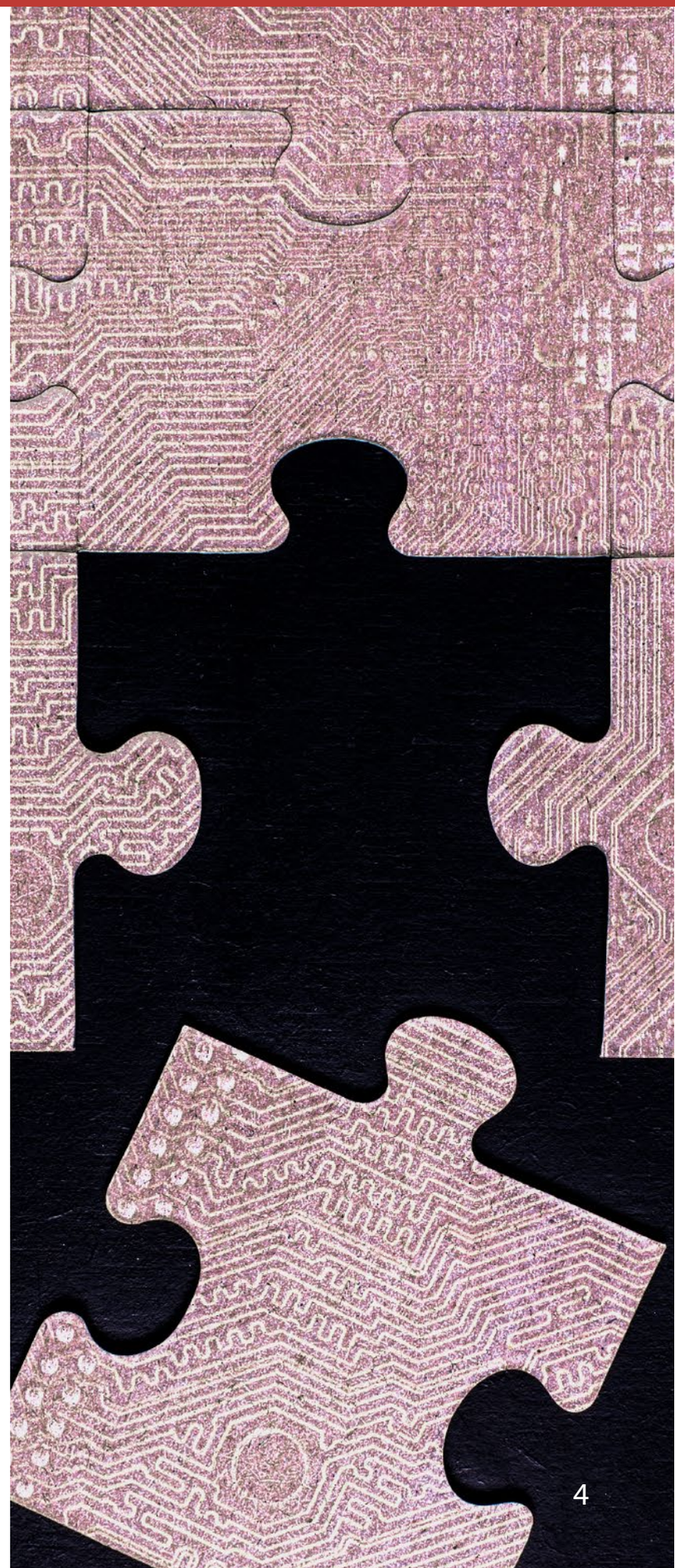


- **Pattern recognition.**
  Identifying phishing or malware.
- **Anomaly detection.**
  Spotting unusual activity, data, or processes.
- **Natural language processing (NLP).**
  Converting unstructured text into structured intelligence.
- **Predictive analytics.**
  Processing data and identifying patterns to make predictions and identify outliers.
- **Incident detection and response.**
  Improving and accelerating the process of aggregating security alerts and prioritizing remediation.
- **Unified view.**
  Providing an overarching status across the network.[4,5]

MACHINES LEARN AND SO DO YOU

Successfully applying a machine language approach to your cybersecurity operations takes thoughtfulness and preparation. There are many pieces to the machine language puzzle and your first step is to identify the problem you are trying to solve and then determine whether machine learning techniques will support the solution. If your answer is yes, next you'll need data and a lot of it. This data needs to be in workable form for the algorithms you'll put into play to find patterns in the data. This is why building your own machine learning solution requires the capability to select data, clean data, and test your algorithm choices and decision-making in a live environment. Testing will require iteration and adjustments will likely need to be made. In other words, like most cybersecurity processes, machine learning is in continuous mode. (See next page).

A machine learning approach to cybersecurity doesn't end with setting up and testing. While a key advantage to machine learning is to "see" patterns of threatening behavior in volumes of disparate data, it should also anticipate corrective actions. The "learning" part comes into play by seeing reoccurring patterns and automatically resolving them or providing guidance via a list of steps for a person to implement. "Using machine learning techniques benefits the decision making process. The resulting information informs not only cybersecurity decisions, but risk-management decisions and ultimately business decisions. Any advantage in making this process faster and more efficient must be a part of your security and business infrastructure," says Vishal Salvi, Chief Information Security Officer, Infosys.



FIVE STEPS TO PREPARE FOR CYBERSECURITY MACHINE LEARNING SUCCESS

While machine learning has gained more immediate traction in financial services and marketing, the benefits of preparing for the eventual machine learning immersion are clear. But the road to get there from here will require a major commitment to the right data and strategy, and sufficient resources of time and talent.

"Machine learning is hard to do right. A huge component is having access to quality data. If you are using machine learning for decision making, the data you use has to be good and unfortunately not all data is good. Machine learning doesn't occur in a vacuum, you need a clear understanding of how it fits into your taxonomy and framework," says Roland Cloutier, Senior Vice President, Global Chief Security Officer, Automatic Data Processing, Inc.

Commenting on the transformational possibilities of machine learning for cybersecurity, John Scimone, SVP, Chief Security Officer, Dell Technologies says, "We're excited by recent advances in machine learning and the opportunities they create to increase cybersecurity effectiveness. The threats we deal with are increasing in volume and complexity at a rate that risks further outpacing security programs. We need a means to disrupt this asymmetry rather than continuing to take an incremental approach. Machine learning may be one of the ways that we transform the effectiveness of our security efforts."

Five Steps to Machine Learning Success

1 Data collection.
Collect your data in advance and create a model with stored data.

2 Data cleaning.
Organize your raw data to ensure no data is missing, eliminate any inconsistencies and combine multiple data sources as needed. Remember that data cleaning is a time-consuming and iterative process, because fixing one issue often uncovers another.

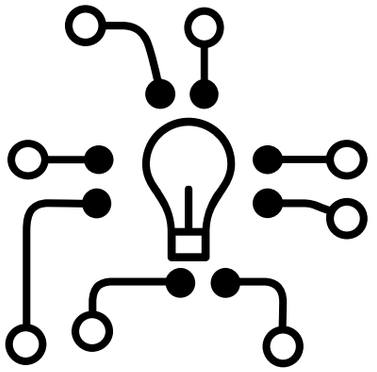
3 Feature engineering.
Once the data is ready, ensure you have the ability to extract the maximum information from the data itself.

4 Model building/model validation.
With clean data and a clear idea of the information you want from it, it's time to build your model based around statistical models and algorithms.

5 Deployment/monitoring.
What happened in the past doesn't necessarily mean the same things will happen in the future, particularly with cybersecurity incidents. Even after your deployment, monitor your models and periodically rerun through the build/validate step.[6]

IT'S NOT ROCKET SCIENCE, IT'S DATA SCIENCE

Figuring out how and when to use new technologies and approaches in your security organization is an ongoing challenge. With machine learning being a hot topic, we went to Kevin Bowers, Senior Technologist, RSA Labs to get some insight on the when, where, and how of machine learning.



Do you have some general advice for security professionals about how to start to incorporate machine learning into their security infrastructure?

Start small. Pick one project where you already understand the available data and build a simple model. Maybe cluster data into groups and see if that provides a new perspective and will bring up new questions. As the questions get harder you'll need to use more advanced methods, but you can get a lot of value out of simple models. So start small and make sure that what you've built has added value in some measurable way.

What's the best way to eliminate the "fear factor" when taking on something new like machine learning?

Begin by picking a project with a narrow scope. Don't create a model to detect all malicious files – choose one. For example, one that detects unsigned windows executables that make a network connection to an IP in a different country. You're more likely to succeed if the scope is tightly constrained. Next, deploy it in silent mode, which lets you see what would have happened without any real risk. You will find bugs in your model and other things that you didn't or couldn't anticipate. Training data can introduce bias that can't be detected until the model is run on new data. Lastly, give your model guard rails. Don't release it and start taking automated actions; have the output reviewed by a person and incorporate the feedback. Don't remove the guard rails until the model has performed consistently over time, and then make sure the guard rails can quickly be put back in place if needed.

Can you give us some examples of where machine learning can potentially be applied in a security organization?

Machine learning is great at picking up patterns in network traffic. The data is huge, but if you have examples of what you want to detect, a model can look for similar traffic. DNS tunneling, domain generation algorithms, and malicious RDP would all be good starting points. Identity assurance is another area and it's typically easy to leverage with other policies that can act as guard rails. Think about common locations, times of day and devices used to build a model for known behavior and then focus attention on authentication attempts that don't match. Financial fraud is another example. Let machine learning pull out the common factors and alert you when a new instance appears. This can be complementary to whatever rules and policies already exist. Lastly, machine learning can help identify and prioritize mundane tasks like vulnerability patching. Based on similarity between machines and vulnerabilities, it may be possible to build a model to identify which machines are most likely to be impacted so that scanning and patching can be prioritized.

Do I need a data scientist to use machine learning in my organization?

If you're just starting, you probably don't need a data scientist right away. Getting and organizing the data and developing the right features are 95 percent of the work. If you have people who can ask good questions, understand security, and have a modest statistics background, you would be well on your way. Once you've moved up the maturity curve a data scientist can be an excellent resource for model tuning and optimization.

WILL MACHINE LEARNING EAT YOUR JOB?

The 2018 (ISC)2 Cybersecurity Workforce Study reported that the shortage of cybersecurity professionals has reached nearly three million[7] globally, with survey respondents currently in security roles saying they would like to spend less of their professional time on administrative tasks and more of it on high value responsibilities including threat intelligence analysis, penetration testing and forensics.

Can machine learning be the answer?

One recent real-world example of the human-machine team working together to deliver superior results comes from the medical field. While children often have fevers, in one of approximately every thousand cases a fever can be deadly and requires special care. An experienced physician's judgment can pick out that one in a thousand case approximately 75 percent of the time. To improve this outcome, a Philadelphia hospital started to use quantitative algorithms from electronic health records to determine which fevers could be life threatening. While the computers more successfully identified the dangerous cases nine times out of ten, they did so with ten times the number of false positives. When experienced medical teams reviewed the computer-generated results, their experience, coupled with the algorithm's list of worrisome cases allowed them to bring detection rates from 86 percent by just using the algorithm to 99.4 percent with a joint computer/doctor team.[8]

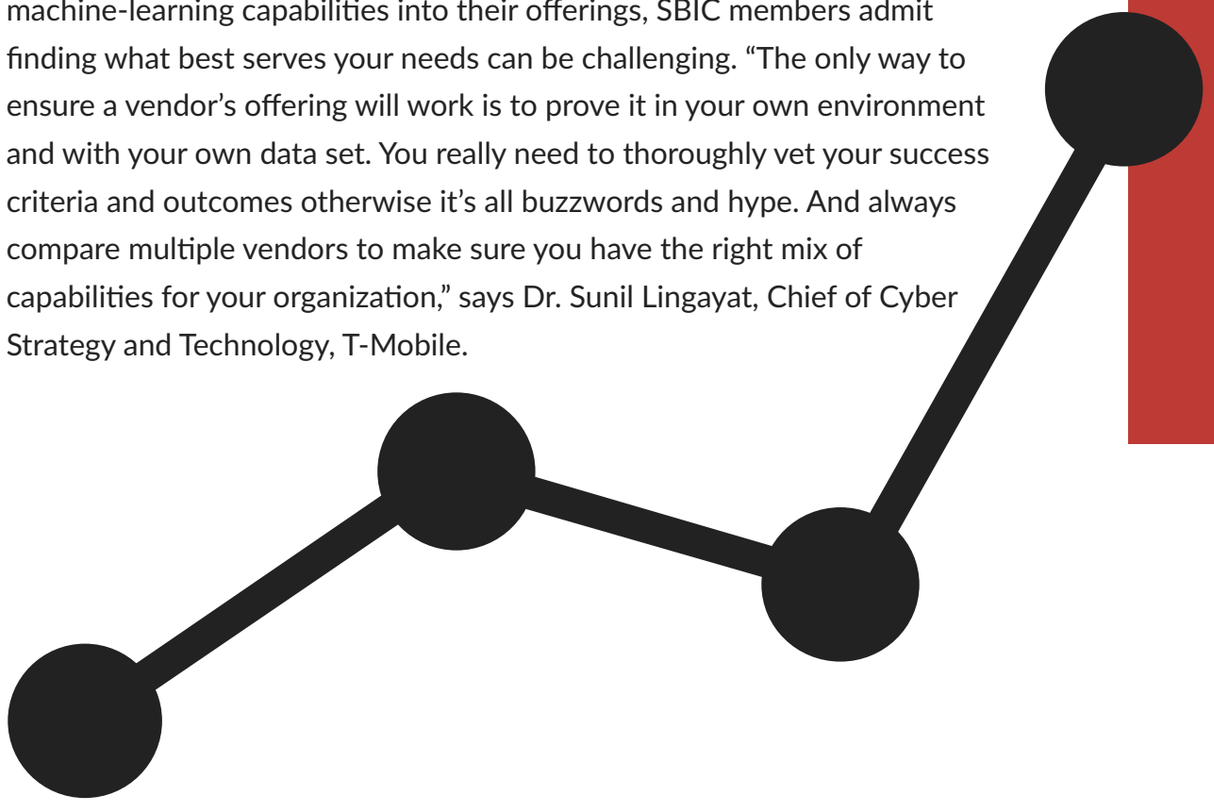
"Machine learning will definitely not replace people. You might have fewer people working on, for example, supervised machine learning tasks, but only the combination of machine learning and trained staff can derive the best value out of this approach. Machine learning is becoming essentially a labor-saving device that allows security professionals to perform the high-level assignments that they are best at."

Ralph Salomon,
Vice President Enterprise Security,
SAP Global Security

HELP COULD BE ON THE WAY

With the anticipated growth in the machine learning market, it will only be a matter of time before a pervasive line of machine-learning infused products, services, and solutions will be available to any sized company hoping to increase their cybersecurity efforts. For smaller companies, it may be a package of machine learning based “analytics on-demand” that provide capabilities they have neither the expertise nor resources to build themselves. For larger organizations using cloud-based applications, more customizable services will be readily available. And larger entities such as defense, military, and critical infrastructure players are likely to develop their own portfolio of machine learning-driven analytics services.

While a number of technology companies are beginning to embed machine-learning capabilities into their offerings, SBIC members admit finding what best serves your needs can be challenging. “The only way to ensure a vendor’s offering will work is to prove it in your own environment and with your own data set. You really need to thoroughly vet your success criteria and outcomes otherwise it’s all buzzwords and hype. And always compare multiple vendors to make sure you have the right mix of capabilities for your organization,” says Dr. Sunil Lingayat, Chief of Cyber Strategy and Technology, T-Mobile.



> *“Whenever new technology innovations emerge, threatened legacy vendors will often try to figure out how to bolt the new tech on to their legacy solutions. This can create a lot of noise and a wide disparity between those who are seeking to stay relevant by adding buzzwords like AI and ML into a legacy product line and those emerging capabilities that are truly innovative and deliver new value. You need to filter out that noise to find the gems of innovation that are of real value for your organization.”*
>
> **John Scimone,**
> SVP, Chief Security Officer,
> Dell Technologies

WHAT COULD GO WRONG?

One downside to leveraging commercial products is that the same products are available to the bad guys. With enough time and effort, cybercriminals can reverse engineer the products to generate misleading results that can help them circumvent capture. While there may be a trade-off in providing cybercriminals with another tool they can possibly turn against us, it's worth noting that predictions say cybercrime will cost the world $6 trillion annually by 2021.[9] Since that is trillion with a "t," that trade-off seems more than tenable. The only option to stay ahead of the potential threats is to embrace the machine learning approach as an essential weapon in the cybersecurity arsenal.

Another potential hurdle is the reality that when you create data sets, you often use multiple data points from multiple assets, essentially creating "new" data assets. This can create compliance issues, particularly in the privacy area where businesses are already challenged to adhere to the new EU GDPR regulations. Putting controls and safeguards in place now to protect any newly developed data assets and adhere to a "privacy by design" approach in machine learning deployments will pay off as your organization's reliance on machine learning grows.

Another thing to keep in mind is that as you deploy a machine learning model, it becomes a data asset in and of itself and must be secured and protected as any other asset would be as they will inevitably be targeted and exploited. It is important to ensure security by design at every stage of design, test, and implementation, including all supporting processes and tools.

Cybercrime

will cost the world

$6 trillion

annually by 2021[9]

CONCLUSION: ACT TODAY, REAP TOMORROW'S REWARDS

The role of AI and specifically it's more accessible subset, machine learning, is making dramatic inroads in business-driven security. Machine learning sits in the cybersecurity sweet spot – mountains of data, sophisticated data analytics, and vast increases in computing power and cloud services are setting the stage for a new and valuable role for electronic brain power in the global cybersecurity battle.

Machine learning, while still on the early part of the maturity curve, is both exciting and inevitable. The knowledge you gain and moves you make now can propel you ahead on the AI cyber curve, setting you up to capitalize on this powerful technology to bring your business a future of new advantages.

About the Security for Business Innovation Council

While most security teams have recognized the need to better align security with business, many still struggle to translate this understanding into concrete plans of action. They know where they need to go, but are unsure how to get there. This is why RSA has convened a group of top security executives from Global 1000 enterprises called the Security for Business Innovation Council and is publishing their ideas in a series of reports. Together we are driving an industry conversation to identify a way forward. Our hope is that these documents will provide your organization with valuable techniques for improving information security.

Recent SBIC industry briefs have explored related topics including:
Translating Security Leadership into Board Value, Taming Cybersecurity Regulation Mayhem, The Evolution & Revolution of the CISO, and The CISO's Guide to Cybersecurity Risk Management and Measurement.
You can check out and download these reports on the SBIC page, www.rsa.com/sbic.

REPORT CONTRIBUTORS

Peter Beardmore
SBIC Chairman/Dir. of Marketing for Digital Risk Management Solutions,
RSA

William Boni
Corporate Information Security Officer and Vice President,
Enterprise Information Security,
T-Mobile USA

Kevin Bowers
Senior Technologist,
RSA Labs

Roland Cloutier
Senior Vice President, Global Chief Security Officer,
Automatic Data Processing, Inc.

Dr. Martijn Dekker
Managing Director, Chief Information Security Officer,
ABN Amro

Ben Doyle
Chief Information Security Officer Asia Pacific,
Thales

Jerry R. Geisler III
Senior Vice President, Global Chief Information Security Officer,
Walmart, Inc.

Dr. Sunil Lingayat,
Chief of Cyber Strategy and Technology,
T-Mobile

Timothy McKnight
EVP & Chief Information Security Officer,
Thomson Reuters

Kevin Meehan
Vice President and Chief Information Security Officer,
The Boeing Company

Ralph Salomon
Vice President Enterprise Security,
SAP Global Security

Vishal Salvi
Chief Information Security Officer,
Infosys

John Scimone
SVP, Chief Security Officer,
Dell Technologies

SOURCES

[1] IDC, Worldwide Semiannual Cognitive Artificial Intelligence Systems Spending Guide, 2017

[2] McKinsey & Company, Artificial Intelligence The Next Digital Frontier? 2017

[3] IFI CLAIMS Patent Services, 2018

[4] The Recorded Future, Machine Learning: Practical Applications for Cybersecurity, 2018

[5] Oltsik, Jon, January 2018, "Artificial Intelligence and Cybersecurity: The Real Deal," Enterprise Strategy Group, (January 25, 2018)

[6] Kanal, Eliezer, June 2017, "Machine Learning in Cybersecurity," Software Engineering Institute, Carnegie Mellon University, (June 5, 2017)

[7] (ISC) 2, 2018 Cybersecurity Workforce Study, 2018

[8] Siddiqui, Gina. "Why Doctors Reject Tools That Make Their Jobs Easier." Scientific American, October 15, 2018. https://blogs.scientificamerican.com/observations/why-doctors-reject-tools-that-make-their-jobs-easier/ (accessed October 25, 2018)

[9] Cybersecurity Ventures, Cybercrime Damages $6 Trillion By 2021, 2017