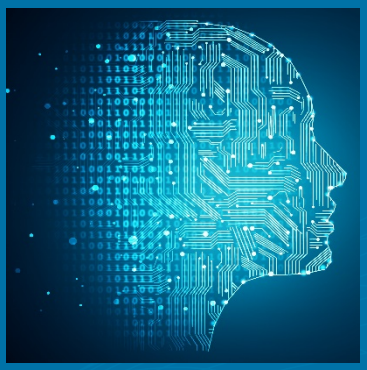
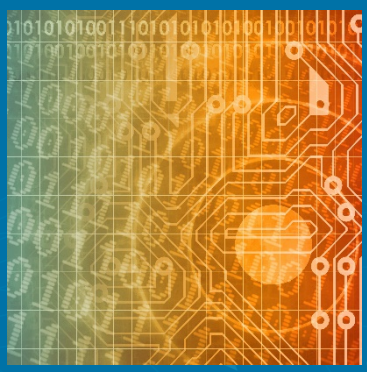



Artificial Intelligence and Machine Learning Applied to Cybersecurity

The result of an intensive three-day IEEE Confluence

6-8 October 2017

To download a copy of the paper and to provide your comments/feedback, please visit: https://www.ieee.org/about/industry/confluence/feedback.html

IEEE
Advancing Technology for Humanity

syntegrity

Executive Summary

In recent years, cybersecurity threats have changed in three important ways:

1. The adversarial motivation has changed. Early attack programs were written as a result of an individual's curiosity, more recent attacks are written by well-funded and trained militaries in support of cyberwarfare or by sophisticated criminal organizations.
2. The breadth and speed of attack adaptation have increased. Whereas the first attacks exploited software weaknesses found by hand, were propagated using "sneakernet," and affected single computers, today's attacks exploit weaknesses found automatically; are automatically propagated over the Internet, packaged even by unsophisticated attackers; and affect computers, tablets, smartphones, and other devices across the globe.
3. The potential impact of an intrusion has increased substantially. Globally connected devices and people mean that attacks affect not only the digital world as in the past but also the physical world through the Internet of Things (IoT) and the social world through ubiquitous social media platforms.

Our entire community needs to respond and develop the technology, and data structures, and the legal, ethical, legislative, and corporate governance mechanisms needed to secure an environment that is increasingly under siege.

The growing size of the attack surface presents both a threat and an opportunity [1]. The threat is that the rapidly increasing adoption of connected devices equipped with conventional security measures will render human security personnel incapable of defending the entire system. The sheer number of devices across the globe makes even a small percentage of failures and compromises a significant event, beyond the ability of human operators to cope with. Consider that for a population of 1 billion (10^9) devices, a 1 percent vulnerability represents 10 million devices. The opportunity, nearly a necessity, is for security artificial intelligence (AI)/machine learning (ML) to act as a force multiplier by augmenting the cybersecurity workforce's ability to defend at scale and speed.

The agility created by AI/ML augmentation of a cybersecurity system (henceforth, "security AI/ML" or "security AI/ML system") is two sided. Along with a rapid response to both detection and remediation comes the potential for an equally rapid corruption of systems. Computers do what they do really quickly, which can include doing the wrong thing. It is essential to keep in mind that, with the increasing use of AI/ML, bad actors and entities have AI/ML at their disposal as well.

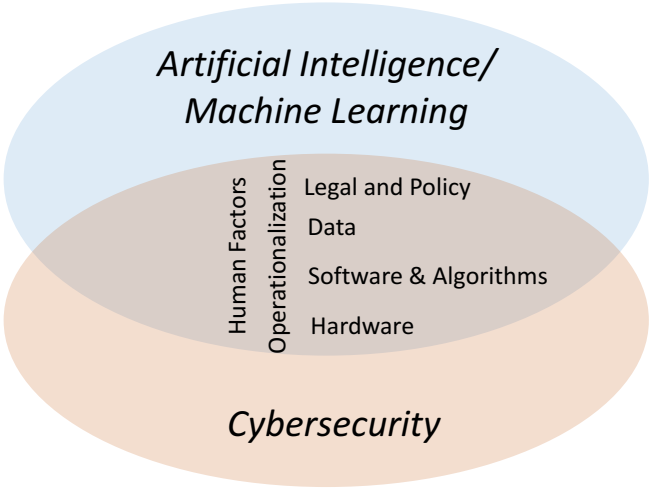
AI/ML systems are already able to identify and develop zero-day exploits, a part of the U.S. Defense Advanced Research Projects Agency (DARPA) 2016 Cyber Grand Challenge. While the technology was intended to help humans more rapidly identify and fix vulnerable systems, it is equally effective for adversarial use in finding and exploiting systems. Malware is already using AI/ML to detect when it is being monitored within a "security sandbox," and to alter its behavior to escape detection. Such a strategy is similar to Volkswagen's effort [2] to program around sandbox testing of diesel emissions. In both cases, adept coders created systems capable of behaving in innocuous ways when in a security sandbox but in a different, malevolent way when employed in operational systems.

In this trend paper, we'll address six different dimensions of the intersection of AI/ML with cybersecurity. They are: legal and policy issues; human factors; data; hardware; software and algorithms; and operationalization. These recommendations are intended for industry (I), academia (A), government (G), and standardization bodies (S). In addition to specific recommendations within each of these six dimensions, we make the following five cross-cutting recommendations, indexed by dimensions (1–5) and to whom they are targeted (I, A, G, and/or S):

- The future needs of cybersecurity will require an interplay of advances in technology (hardware, software, and data), legal and human factors, and mathematically verified trust (1, 2, 3, 4, and 5) (I, A, and G).
- It will require concerted business efforts to create products acceptable to the market, certified by established regulatory authorities (1, 2, 4, and 5) (I and G).
- If humans are to trust AI/ML, AI/ML-fueled cybersecurity must be based on standardized and audited operations (1 and 5) (I and S).

- Regulators will need to protect research and operations and establish internationally recognized cooperative organizations (1 and 2) (S and G).
- Data, models, and fault warehouses will be essential for tracking progress and documenting threats, defenses, and solutions (3, 4, and 5) (S, I, and A).

Our recommendations can be applied at different time horizons. Operationalization takes the least amount of time and could be accomplished in no more than two years. This is similarly true for data and software. Addressing legal and policy issues takes more time, at least several years. Hardware, e.g., new processor architectures, typically takes more than five years to materialize.



The Six Dimensions of Intersection of AI/ML and Cybersecurity

In the following sections, we describe in detail the six most important dimensions related to the intersection of AI/ML with cybersecurity. We identify threats, challenges, and opportunities and make recommendations for each dimension.

1. Legal and Policy Issues: Building Trust through Accountability

AI/ML augmentation of cybersecurity systems may seem a highly technical topic best left to a small group of expert computer scientists. However, the most formidable challenges for the future of AI/ML are likely to be social in nature. While AI/ML promises to improve security by automating some aspects of defense, caution is needed for the creation, deployment, and use of these systems. Unless developed and used very carefully, AI/ML may irretrievably damage national security, economic stability, and other social structures. As such, it should not be viewed as a panacea, and our social structures (and the humans who rely on them) must be prepared for the inevitability that the systems will fail in both anticipated and unanticipated ways. Safety nets of legal and ethical constraints are needed. Building a world—a social, ethical, and legal context—that is ready for the incorporation of AI/ML matters as much as the creation of the technical systems themselves.

Despite (or perhaps because of) the considerable enthusiasm for AI in marketing circles, the meaning of the term is now ambiguous in common parlance. This fact should act as a warning to proceed with care as we enter an age of AI/ML. Creators and users of AI/ML should not be financially rewarded for shipping or implementing code prematurely without a thorough analysis and testing. While it is common for companies to ship code having known errors with plans to correct these in a future update, this model of shipping code does not work for AI/ML. The stakes of possible harm are simply too high with flawed AI/ML. In 2016, the Mirai botnet heralded the arrival of a new category of attack: distributed denial of service (DDoS) attacks carried out by botnets consisting entirely of vulnerable IoT consumer devices. Despite these devices having relatively little computing power, Mirai nevertheless succeeded in DDoS to some of the best-defended websites on the Internet [3]. Now, imagine the scale of damage a sophisticated, well-resourced, and security-compromised AI/ML might cause in the physical world.

In situations where a developer or operator loses control of security AI/ML and causes catastrophic harm, the public's trust in AI/ML will be shaken. A strong legal response will be needed to rebuild public trust in AI/ML. Up to now, however, courts and regulators in most countries have been slow to assess legal liability for harm arising from software malfunction. This reticence will need to change in the context of AI/ML. Courts and regulators will be willing to ascribe liability—and perhaps even criminal culpability—when corporate assets, humans, and infrastructure are physically harmed because of malfunctions or inadequate care in the creation, deployment, and use of AI/ML.

We should start preparing now for the AI/ML-caused disasters that will inevitably occur. Here are 10 things we can do now.

1. *Support the creation of enhanced regulatory structures.* As standards start to emerge from the technical community, regulators will begin to construct a proactive set of shared minimum baselines for reasonable conduct—what might be called "floors of care"—for security AI/ML. We have seen these floors of care already emerge generally in computer security enforcement. While consensus about the optimal design of AI/ML systems may still be developing, consensus around basic types of building, implementation, and security errors likely exists already. A set of tailored regulatory and enforcement measures will be needed to prevent low-quality or otherwise flawed security AI/ML from damaging the market for responsible builders and operators. For example, in the United States, enforcement will likely fall partially within the jurisdiction of the Federal Trade Commission (FTC) under its Section 5 FTC Act authority. However, in most instances, the FTC currently does not have standalone rulemaking and fining authority. As such, the legal evolution of certain regulators' authority (and budgets) is likely necessary for building trust in and policing AI/ML.
2. *Urge the creation of additional technical feedback loops for regulators.* An important step toward ensuring suitable regulatory approaches involves constructing formal technical feedback loops inside national legislative and regulatory bodies. In the United States, for example, Congress and regulatory agencies will serve as a starting point for most AI/ML policy. However, Congress currently lacks a funded office of technology assessment to offer technical feedback. As such, Congress should resurrect this technical body as the Office of Information Technology Assessment, with a budget and staff of technical practitioners knowledgeable about AI [4].
3. *Urge stronger legal protection for security research.* A single uncaught vulnerability in security AI/ML may result in significant harm. Similarly, training AI/ML systems will depend on the availability of high-quality security research. While rogue attackers require prosecution, legal systems should also be careful to facilitate, rather than burden, this essential security research. For example, in the United States, security researchers both inside and outside the academia require at least two corrective legal buffers as soon as possible: 1) congressional codification of the security research exemption granted by the Librarian of Congress to Section 1201 of the Digital Millennium Copyright Act [5] and 2) an amendment of the Computer Fraud and Abuse Act to provide clarity arising from statutory ambiguities regarding computer intrusion, absent definitions of key statutory terms, and judicial divisions in interpretation [4].
4. *Recognize that international legal and regulatory harmonization will present challenges.* As the recent negotiations over security tools and the restrictions of the Wassenaar Arrangement [6] have demonstrated, coordination and harmonization of regimes across borders and policy areas present formidable obstacles—and may also require years of negotiations. Discussions concerning AI/ML will also trigger a need to reconcile various legal frameworks from prior eras and across jurisdictions. Because of different legal approaches to privacy, security, and tort recourse for consumers in particular, the creators and operators of AI/ML may find themselves the subject of litigation in international forums where their AI/ML has allegedly caused harm. The contractual choice of forum provisions and limitations of liability will not be universally enforced. However, any discussion of statutory limitations of liability for AI/ML is premature at this juncture: it would erode public trust in these systems and create negative incentives for unsafe conduct by builders and users.
5. *Demand that criminal enforcers be wary of security AI/ML*. As the defeat devices employed by Volkswagen in its diesel cars remind us [7], computer code can be leveraged directly for purposes of avoiding regulatory requirements and facilitating criminality. The current set of tools available for identifying and prosecuting crimes facilitated by AI/ML may require reassessment. Regulators should consider offering avenues for both corporate and governmental whistle-blowers to report dangerous AI/ML systems in a manner shielded from legal consequences. Without these opportunities for whistle-blowing, dangerous AI/ML systems are likely to result in avoidable and severe levels of harm, which, in turn, will result in a breakdown of trust in AI/ML as a whole.
6. *Recognize that the use of security AI/ML for criminal enforcement has the potential to violate individual civil liberties guarantees.* As described in the "Human Factors" section, AI/ML systems are only as good as the human-curated training data and the strategic choices made for training methodologies. Even the highest-quality systems can produce false positive results. Particularly when full transparency into their functionality is absent, AI/ML systems do not meet the legal standards of individualized justice beyond a reasonable doubt for a criminal defendant, and this should not be used in lieu of thorough investigations and/or the independent judgment of a finder of fact (a judge or jury). Creating trust in AI/ML requires, first and foremost, preservation of traditional legal baselines of liberty and justice for citizens.

7. *Use the legal lag for technical standards creation.* A lag always exists between progress within the scientific community and the catching up of legal and regulatory mechanisms. However, this lag sometimes represents a positive feature rather than a bug. For security AI/ML, the legal lag creates a window of opportunity for builders, operators, and international organizations such as the IEEE to initiate a policy conversation with regulators to craft shared minimum baselines, or floors, for reasonable care of the AI/ML. In other words, these conversations should occur in advance of any catastrophic global AI/ML incident that will likely trigger reactionary and potentially aggressive regulation.
8. *Correct currently imperfect security indexing and reporting structures*. Our current systems of assessing vulnerability and issuing advisories suffer from deficiencies in scalability and accuracy [8]. Similarly, patching and disclosure practices vary across entities in ways that sometimes place consumers, national security, and business partners at avoidable risk. In preparation for the expedited pace of flaws that security AI/ML will uncover and report, these deficiencies require immediate remediation. Corrections will necessitate, among other things, standardizing security advisory formats to the greatest extent possible and enforcing failures to accurately disclose (and patch) flaws in a reasonably timely manner. These corrections to existing structures will pave the way for the types of transparency disclosures (the limitations of selected data sources; training, strategy, and end of life plans; and other key characteristics) that will distinguish various AI/ML systems from one another.
9. *Support the robust enforcement of security by design.* Preserving public trust in security AI/ML requires that, first and foremost, these systems be constructed as securely as possible from the beginning. Security cannot be retrofitted; adding security after code has shipped in a vulnerable state inevitably introduces new vulnerabilities and undesirably increases complexity. For this reason, the current focus on post-breach enforcement should be replaced with a focus on security by design and security processes. Unfortunately, lessons from past generations of security products warn that such products are sometimes themselves vulnerable, placing their users at greater security risk rather than better defending them [9]. In such circumstances, regulatory enforcement action should require immediate correction: levying fines and/or requiring the removal of unsafe AI/ML from the marketplace.
10. *Engage in discussion and the choice of ethical design.* The creation of policy guidelines around issues of ethical design presents an opportunity for organizations such as the IEEE to expand their current discussions. Engaging the broader technology and user community in issues of ethical design for security AI/ML will facilitate public trust and nudge improvements among builders and operators. In the case of an individual builder, a robust strategy for legal risk mitigation should involve working with counsel to document the corporate decision-making process around ethical design choices. For example, as the WannaCry worm demonstrates [10], the presence of a remote (albeit accidental in this instance) "kill-switch" and constant human oversight are two strategies for mitigating harm when code malfunctions. Ultimately, regulators will analyze whether there is proof of ethical and safer design choices. Such documented choices aimed at reducing risk to innocent third parties demonstrate a degree of care that is likely to mitigate findings of liability.

2. Human Factors: Building Technical and Human Trust

In 1983, Stanislav Petrov, a Soviet officer, helped avert nuclear war. Petrov had been on duty at the Serpukhov-15 secret command center outside Moscow when the attack detection algorithms running his systems warned that the United States had launched five intercontinental ballistic missiles at the U.S.S.R. [11]. Instead of reporting the alarm to his superiors, Petrov paused [11]. Although he knew that the algorithms had processed over 4,000 variables, his years of experience (and his awareness that the system had been deployed in a hurried manner) counseled caution [11]; he did not trust the system [11]. Deeming the notification a false alarm, he chose not to report it. Later, forensic analysis showed that Petrov's distrust of the system was well founded. The predictive algorithms had, indeed, been confused. The alarm had been falsely triggered by the sun's reflections from clouds [11], a data input the system's programmers had apparently not adequately anticipated.

While not every security situation is as serious as a nuclear standoff, this incident serves as an important reminder that the future of security AI/ML will rely not only on technical trust but also on human trust. Even the best-engineered systems can fail. The key question, then, becomes whether they will "fail well," that is, in a manner that preserves humans' trust in security AI/ML and minimizes harm. Indeed, it will be these human trust factors in the operationalization of AI/ML systems that will dictate their adoption rates.

1. *Trust-building through transparency and preemptive risk assessment.* In the 1980s, in his treaty negotiations with the Soviets, US President Ronald Reagan often quoted the Russian proverb "trust but verify." It's still a useful touchstone as we discuss security AI/ML. Trust in AI/ML can be substantially buttressed through the builders' transparency in disclosing strategic choices, updating processes, and providing contingency plans to assist their systems in "failing well."
    a. *All training data are not equal.* The first layer of necessary transparency involves the human processes for selecting the data used to train security AI/ML. As social scientists and statisticians have amply demonstrated [12], the selection of data sets will potentially suffer from a number of sampling errors and biases. Every training sample will have a certain degree of sampling error, and this error requires analysis and disclosure to avoid creating a false sense of confidence in a particular training methodology. Different training methodologies will vary in success based, in part, on the extent of this sampling error. Builders of AI/ML systems should also disclose the extent of any affirmative steps they have taken to avoid sampling bias in selection. In other words, they should articulate why they are confident that the sample used for training data is, in fact, accurately representative of the entire population of real-world deployment situations that the AI/ML system is likely to encounter. For example, one infamous AI/ML training failure occurred in March 2016 when Microsoft introduced Tay, an "AI chatbot" on Twitter. Within one day, Twitter users "taught" Tay to spout racist and Nazi propaganda [13], a highly undesirable outcome from Microsoft's perspective. Most importantly, as with every rigorous scientific process, the measurement and selection processes with respect to training data should be replicable by independent third parties. Replicable measurement processes, along with what social scientists call "interrater reliability" checks, build confidence and trust. It is through this level of rigor, planning, and transparency that builders can reassure both users and policymakers that their systems are well built and thus, to the greatest extent possible, protected against malfunctioning in catastrophic ways.
    b. *Building needs to be accomplished with attackers in mind.* As with all code, the question related to an AI/ML security compromise is "when" and not "if." Yet, as daily headlines about data breaches remind us, both the public and private sector still struggle with even rudimentary questions of security, and legal accountability has been slow. Adversaries will attempt to fool systems as built and try to repurpose systems for their own nefarious interests. Builders and operators of AI/ML systems must recognize and plan for this unfortunate and inevitable security reality, preparing technical incident response capabilities and corporate processes for mitigating harms to third parties caused by compromised AI/ML.
    c. *Risk management should leverage humans in the loop—as a feature, not a bug.* A second trust-building disclosure involves an honest acknowledgment of the limitations of security AI/ML systems and their risks. Although security AI/ML presents a potentially game-changing improvement for extending the capacity of computers and humans to jointly defend against attackers, as Petrov's story cautions, malfunctions carry significant risks and potentially devastating consequences. In particular, the more sensitive the deployment context, the more important it becomes to retain human oversight as a part of the decision loop. Some contexts may even prove too fragile for the use of AI/ML. When appropriate, properly designed and implemented AI/ML can leverage preexisting and new knowledge to assist in more effectively securing systems at a speed and efficiency beyond human abilities. However, the implementation of AI/ML systems should not be viewed as an excuse to eliminate humans or limit the exercise of necessary discretion and judgment. Indeed, humans should remain the ultimate arbiters for all decisions that may have potentially catastrophic consequences.

       As the DARPA Cyber Grand Challenge organizers explain [14], careful planning beforehand was required to constrain the competitors' security AI/ML systems and predict their possible malfunctions. Accurately predicting and avoiding harm constitute a dispositive component of building trust in AI/ML system capabilities. Similarly, the competitors in the DARPA Grand Challenge demonstrated that, even when two systems appear on their face to be parallel in their functionality and training data, the builders of each have made different key strategic behavioral determinations. They have also potentially employed dissimilar training methodologies [14]. Consequently, individual AI/ML systems will behave differently, even in the same deployment environment and relying on the same training data. These strategic choices by builders should similarly be disclosed to generate trust in AI/ML. Disclosure will assist the market (and, later, legal enforcers) in more accurately assessing

suitability for particular deployments and the extent of care that went into the construction (or selection) of particular AI/ML systems.

2. *Trust building through accountability.* It is inevitable that some security AI/ML will malfunction, just as Petrov's system did. To maintain trust in light of this expected malfunction reality, builders and operators should strive to self-audit, third-party audit, and build systems that fail safely in ways that limit harm. However, some builders' attempts at self-audit and correction will prove inadequate. In these circumstances, preserving trust in AI/ML will necessarily lead to regulation, enforcement, and legally mandated damages recoveries by harmed third parties.

We offer the following four recommendations to assist with the development of human trust in security AI/ML.

1. *Participate in standards development.* Because standards usually chase industry innovation, we recommend that the academic community, standards-setting organizations such as the IEEE, builders of AI/ML systems, and regulators convene standards meetings on an ongoing basis to articulate the minimum floors of care required in building and operating security AI/ML. Examples are the IEEE Cybersecurity initiative [15] and IEEE Standard for Ethically Aligned Design [16].
    a. In particular, this group of interdisciplinary experts should issue recommendations with respect to data selection for systems training, borrowing methodologies from social science and statistics research regarding sampling bias and errors. Identifying data blind spots and articulating floors of care in an interdisciplinary manner will build trust in AI and security.
    b. Further, this body of interdisciplinary experts should continue convening on a regular basis to engage with evolving practices in AI as they are implemented in individual security AI/ML systems. In particular, these experts should perform postmortem analyses of systems that malfunction due to identifiable design or strategic choices made by their creators and users.
2. *Assist in demystification.* Academia, industry, and regulators should each independently engage with the daunting process of public education to demystify the benefits and limitations of security AI/ML. As one important example, the public currently lacks a set of narratives that realistically assesses the functioning of AI/ML. Current narratives either err on the side of unrealistically utopian visions or dramatically dystopian ones leading to, for example, the extinction or enslavement of humanity.
3. *Regularly perform robust self-audit.* Builders and operators of security AI/ML should engage with existing standards of care, such as those reflected by International Organization for Standardization (ISO) standards [17], and analyze their organizations for the existence of robust and rigorous self-audit and technical governance processes. Each AI/ML builder and operator should, in particular, ensure that security by design principles are in place throughout the organization. Because of the severe risks presented by malfunctioning AI/ML, each organization should make certain that a designated ethics officer is in place who has expertise in both AI and security. This ethics officer should work closely with the chief information security officer (CISO), general counsel, and other C-suite executives to craft meaningful accountability processes that accurately assess the limitations of AI/ML.
4. *Regularly perform robust external audit*. Rampant data breaches and vulnerabilities remind us that all code contains errors. In addition to self-audit mechanisms, third-party technical audits offer a key verification method for security AI/ML safety. Robust regulatory enforcement presents another necessary trust-preserving audit mechanism for the future of AI/ML.

3. Data: New Information Frontiers

In 2014, the International Data Corporation reported that the amount of data was doubling every year and would reach 44 zetabytes (44×10^{21} bytes) by 2020 [18]. This figure includes data from individuals, devices, technical networks, social networks, and various applications. As security AI/ML requires large and diverse data sets for effective training and the networks that the AI/ML will be applied to produce significant amounts of real-time data, it is clear that data represent a critical dimension.

To be effective, security AI/ML algorithms must be trained on large, diverse training data sets. As such, the effectiveness of the algorithms is directly proportional to the quantity and quality of the data. While large training data sets are often available, one challenge is the completeness of the data. Existing devices and networks were not originally designed with instrumentation and measurement as an integral feature; therefore, the data available from these devices and networks are not capturing critical conditions.

Additionally, data sets are often incomplete because individuals and organizations are influenced by liability and reputational concerns and withhold data about potentially embarrassing cybersecurity events that could reduce customer and investor confidence. Consumer privacy concerns, government policies and regulation, and protection of proprietary information also contribute to incomplete data sets.

Relevancy and integrity are additional factors associated with data. While simulated data sets are convenient to generate, they are often artificial because they do not properly encapsulate reality and the human dimension of adversarial actions. Additionally, to be effective, data sets must be continually updated so they include the most recent evolution of threat results. Data that do not include the most recent attack data cannot be effective against those attacks. Data integrity affects both the effectiveness of and confidence in AI/ML. Data collection techniques, by their very nature, often include unintended human and technical biases. Understanding, documenting, and sharing those biases are important to ensure AI/ML effectiveness and operation. Data integrity also affects human confidence in AI/ML. If the AI/ML training data set is incomplete, includes questionable biases, or is, in general, not fully understood, then confidence in the entire system is diminished. Preprocessing of the data prior to use for training can also alter data integrity and reduce confidence.

Beyond the actual data used for the training and operational employment of AI/ML in cybersecurity applications, storing, sharing, and ensuring the integrity of the data impact the effectiveness of and confidence in the respective systems. No centralized, standardized, and qualified data warehouses for cybersecurity data currently exist that allow broad sharing across industry, government, and academia.

Because data in the cybersecurity domain continue to grow at an increasing rate, it is important to consider alternative algorithmic approaches that abstract threat anomalies from the data level to higher-level ensemble indicators. Characterizing common attack patterns will allow AI/ML models to focus on features that predict outcomes. Additionally, rare threat events, while potentially devastating, are often underrepresented in a probabilistic model that encompasses all threats. As a result, there is a need within the AI/ML development community to devise a feature-engineering approach. This will allow AI/ML systems to analyze common attack patterns, then generate representative attack scenarios, subsequently analyze those patterns to identify variations, and ultimately update and improve the algorithms.

It is well known within the military that, while operations are planned with great precision, the enemy gets a vote on the final outcome. Security AI/ML models are complex, and a sophisticated adversary can determine the boundaries of the model and potentially exploit these boundaries. The fundamental challenge is that detection-driven data potentially create a false representation of an attack landscape, and the models are then updated to prevent only attackers who are willingly or unwittingly transparent [19]. Primary data gathered from professional attackers show a completely different landscape than the corresponding landscape inferred from detections [20].

We offer the following recommendations for the data dimension.

1. The sponsorship of *data warehouses*, with support from analysts, can maintain data quality and facilitate feature engineering. Government and industry are both capable of providing financial support and leadership for coordinated data management.
2. A sponsored data warehousing organization should drive a move toward *international data storage standards* to facilitate information sharing across organizations. These standards should be sufficiently flexible to evolve as threats, models, and networks change over the ten-year horizon.
3. If we try to harmonize rules and standards, there could be a race to meet the lowest common denominator. This may lose data granularity. We should look at “norms” in addition to standards. We should also use the format of data, or metadata, to ensure trust and interoperability among organizations with different security AI/ML. Governments should establish *regulations, rules, and norms on new, frontier data sets* for smart cities, smart cars, and the IoT. Care should be taken regarding how data are handled: Can personal data be collected? Do vendors get access?
4. Academia should be invited to work on a framework. We need *cross-disciplinary research* across AI/ML, cybersecurity, data science, human-factor cyber engineering, the social sciences, and the work of futurists. Research is required for contextual and inferential data collection, using data formatting as well as sensors to collect data.
5. *Economic incentives* should be introduced so that sensors are in place *to collect data*. Governments at all levels should decide how to collect and make use of their data. Information should be collected that

respects set policies. To leverage data for cybersecurity, while maintaining privacy and security considerations, the cybersecurity community should investigate sharing data through trusted third parties.

6. *Mechanisms to measure the confidence level of data's relevance and accuracy should be established*. Will the market help ensure data accuracy and relevance, especially when people are paying for the data?
7. Data collection and feature engineering should focus on cybersecurity attributes that have a reasonably small probability of being manipulated by bad actors. *Oversampling techniques can increase the presence of threat cases in the training set*. Where positive examples are rare, the number of features in any model should be limited.

4. Hardware for AI/ML and Cybersecurity

Nathaniel Fick, chief executive officer of Endgame, has stated that "the attackers' advantage is getting ever stronger. Companies have growing attack surfaces driven by device proliferation: the IoT, mobility, automation and AI, and infrastructure as a service (IaaS). Meanwhile, barriers to entry to creating and deploying sophisticated cyber weapons continue to fall." The network is no longer defined by the electronic equipment within the physical protection of buildings and campuses. Today, the network consists of human users connected by mobile devices anywhere in the world and autonomous devices broadcasting sensor information from remote locations. This is a very large and varied attack surface to manage and defend against adversaries deploying sophisticated cyberattacks aided by AI bots. The problem appears impossibly hard to solve, with many leading CISOs admitting that they no longer view cyberattacks as a question of *whether* they will be hacked, but rather *when*.

Hardware is an integral part of this solution in three ways. The first is by integrating security into hardware device designs. The second is by creating hardware network architectures that can intelligently monitor the network's security state. The third is by creating hardware that allows AI/ML systems to solve more complex problems by eliminating existing compute barriers.

Because IoT and mobile devices usually lack the computational power needed to run advanced security software, security must be embedded within the hardware of the devices themselves. The devices must become the front line of defense, or they will be used to enable attacks. This was shown in the October 2016 DDoS attack [21], [22] in which millions of DVRs and webcams were converted into botnets by the Mirai malware and then used to launch a continuous and massive stream of traffic that resulted in shutting down Netflix and other major websites. The ability exists to mitigate these attacks or make them more difficult by implementing hardware-based security features such as ARM's TrustZone technology [23], which supports secure end points and a device root of trust. These features are essential for an AI-based system, if IoT devices can manage to defeat simple attacks and provide an AI algorithm not only to understand the current state of the network but also to find and defend against anomalies. In the highly competitive, low-cost environment of the IoT, it is hard to convince device manufacturers to commit design time and resources to implementing these features. This is clearly shown in the case of Meltdown and Spectre, where simple security fixes could have prevented large-scale security flaws; but there was no incentive for industry to find and implement those fixes. Government agencies, standards organizations, and consumers must act in concert to demand that security be integral to these devices; end-point/edge devices must also be strengthened to make them harder to compromise by deploying at least parts of an AI/ML system on the edge devices themselves.

A model to follow could be the 1890 establishment of Underwriters Laboratory (UL) to develop standards for electrical wiring because of the potential to create fires. The National Fire Protection Association was also founded at that time to initiate fire codes and promote laws for fire safety. Both of these efforts helped to create a demand for certified equipment. Consumers wanted to be assured that the devices they bought would not be a danger to their houses and families. All they had to do was look for the "UL" seal. This, in addition to product safety standards legally imposed by appropriate regulatory bodies, forced manufacturers to add safety features to their products to sell them. A similar UL-like seal for security is needed. Unfortunately, despite having been discussed for years and some recent efforts being made, the idea has never been implemented at scale. We need to treat cyber incidents in a manner similar to traditional safety incidents.

Effectively using AI/ML to defend against cyberattacks requires the ability to monitor network security health, assess threats to the network, and provide solutions to cyber analysts to defeat the attack. Monitoring the

network and assessing threats require information in the form of telemetry. Networks should have imbedded hardware monitors that can broadcast the status of different devices in the network to a central defense AI/ML system and so detect and defeat threats before they damage the network. The challenge here is that, to create such a system, the computer and network architecture must be designed with security in mind. It is not enough to simply place monitors into hardware; thought must be given to what information is needed and how best to deploy the monitors to ensure adequate coverage of the network as well as real-time alerting of attacks as they occur—and, of course, the security of such a system. The National Science Foundation and DARPA have begun investigating what this next-generation network would be, but more needs to be done. Industry and academia must also step up and explore what this system would look like and how it would function. This research will help us enormously, not just to deploy an AI/ML solution but to deploy the *right* solution.

Finally, today's computer architecture was designed to do complex calculations on relatively small amounts of data. This architecture is not suited to the type of computations performed by modern AI/ML systems. AI/ML algorithms find clusters of data or associations to connect observed information together and so provide context for the observations. This context allows the machine to understand the perceived world and make decisions about how to respond to what the system is observing. To accomplish this, AI/ML algorithms process a large amount of data and perform relatively simple operations (e.g., matrix multiplications) on those data. This is a fundamentally different processing paradigm from what is common today. Because of this disconnect, AI requires a large amount of computing hardware to do the training, thereby precluding the real-time threat assessment and response required by cybersecurity for new threats. To solve this problem, computer architects need to fundamentally change their approach to computing. We need to take a more data-centric approach, focusing on how data flow through a processor, and a less processor-centric approach, which focuses on how computations are done. Academia, funded by government agencies and industry, can lead the way by experimenting with new and novel outside-the-box architectures. Innovative approaches are the only way to shake up a field that hasn't effectively changed in the last 50 years. Without a new architecture, AI/ML will be unable to solve large-scale problems such as those in the cybersecurity application.

AI/ML can also be used to design better hardware. It is difficult to create hardware that functions predictably and securely because those attributes traditionally depend on the experience, foresight, and knowledge of human designers. AI can be integrated into current design tools, like those produced by Cadence and Mentor Graphics, in such a way to find common design mistakes or errors early in the development cycle. This would be a significant aid to the human designers. AI/ML is able to explore more possible failure modes and can look for complex failure mechanisms buried in a design that would otherwise be missed. Eliminating hardware faults can go a long way toward making the network secure because hardware faults and design errors are among the most reliable targets for exploits. Based on a 2015 study by MITRE, 2,800 cyberattacks could be traced back to seven classes of hardware bugs. Eliminating these bugs using AI/ML in the design process will close several attack avenues used by hackers. The electronic design automation community will need to invest in developing these tools, and their users will have to provide fault information so that an AI/ML system can learn from those mistakes. This effort should be mostly industry focused, with the government playing a supporting role in encouraging the development of these systems.

We make the following recommendations regarding hardware.

1. Investing in new memories and interconnects will *more efficiently process large data*. Currently, anywhere between 40 and 96% of time/energy is spent moving data around, and between 4 and 60% is spent processing [24], [25].
2. Solving important, real-world problems will require many more graphics processing units (GPUs), central processing units (CPUs), application-specific integrated circuits (ASICs), and field-programmable gate arrays (FPGAs) than are practical*. Improving data movement (see 1, above) will enable new AI algorithms.*
3. The IoT needs *security standards*, developed by a standards body such as the National Institute of Standards and Technology or the IEEE. Another organization (akin to the UL) and regulators should enforce adoption.
4. *Educating the public about the value of certification* and creating a market function to force hardware manufacturers to incorporate security, including in accelerators such as GPUs, FPGAs, and tensor processing units (TPUs), are essential.
5. Academia, industry, and government should develop *a methodology for building a secure hardware* (we might call it "design for security").

6. Industry should establish *an affordable means for security testing and certification*. Today, such laboratories are so expensive that most companies do not use them.
7. *Security middleware to monitor a system* and issue alerts using current hardware monitors should be developed along with new ones to determine system security.
8. Experts should devise *certifications enabling manufacturers to regard security as a contributor to profits and allowing consumers to differentiate in their purchasing behavior based on security robustness*.
9. Universities should *incorporate security into the hardware development curriculum of system design courses* and include hardware into cyber analysts' and programmers' training.

5. Software And Algorithms for AI/ML and Cybersecurity

Countering cybersecurity attacks in a completely autonomous way, using sophisticated AI/ML algorithms and without human supervision, is both appealing and controversial. Security AI/ML software observes system usage, estimating in real time whether there is a threat. To enable ML systems to construct a detailed model of a scenario, developers are challenged to quickly understand normal and threatening scenarios and their associated feature space at a high level. Five basic principles have guided this analysis of how corporations, government agencies, and other institutions should best deploy AI/ML software and algorithms to address growing cybersecurity threats.

1. For both technological and policy reasons, *a completely autonomous system* for detecting and responding to threats *is not always an appropriate option*. Balancing the benefit of human versus machine—given that they both make mistakes—should be used to decide who or what makes the decision.
2. The underlying technologies of cybersecurity and AI/ML are evolving rapidly; therefore, *an adaptable AI/ML framework* must be developed. Focusing on a specific methodology or algorithm, such as deep learning, would be unwise because developments in a few years are likely to supersede it. For the same reason, the search for a single "proven" cybersecurity model is a chimera.
3. AI/ML approaches to cybersecurity *must be problem specific*. A successful approach will feature more than one model, operating in sequence, in any conceivable circumstance.
4. AI/ML models for cybersecurity will be *applied in two phases*. The first phase will involve developing an understanding of the normal historical landscape of network data traffic, extracting actionable insights about threats, and learning to identify anomalies in network traffic. The second phase will consist of applying an understanding of "normal" to identify anomalous situations requiring human interaction and action against known threat profiles.
5. AI/ML for cybersecurity is similar in nature to the application of AI/ML for fraud: both are adversarial and ongoing. In either case, perpetrators will modify their behavior when their actions are detected and thwarted, necessitating *constantly evolving countermeasures*.

Because typical cybersecurity data sets are extremely large, networks for data delivery and the processing of ML models must be capable of efficiently handling staggering amounts of diverse data. The scarcity of such networks today is a major hindrance to progress in the field. Achieving such networks for real-time analytics requires even more careful software design and algorithms.

Additionally, AI/ML can be applied to cyber networks in either a proactive or a passive (forensic) way. This distinction merits explicit inclusion in planning and design. Proactive models leverage insights gained from historical analysis to continually monitor network activity against known indicators of attack patterns. As a new input arrives, it is compared to all known patterns of attack. As knowledge of these patterns deepens (a function of both the data and an analysis of historical information), a more aggressive approach for reacting to suspicious activity can be employed.

In contrast, passive models collect sufficient data to enable the post hoc analysis of attacks that were unanticipated in kind. This allows an organization to use a tip from another domain to learn about how an attack was carried out and possibly also to be able to attribute the attack to a specific operator. Collected data should include those that provide broad visibility into enterprise activities, as a way to understand how malicious software can spread, as well as deep visibility into specific system activities, as a way to understand how malicious software executed its attacks. The first usually requires capturing network activity, while the second usually requires capturing system activity on each system.

Natural language processing (NLP) makes it possible to derive actionable insights from previously inaccessible data. Analyzing unstructured text with NLP enables the extraction of key actors from past cyber incidents, news stories, analysis reports, and many other similar text sources. Knowledge Graph technology enables the discovery of nonobvious secondary and tertiary relationships by connecting individual nodes and also provides insights into sequences of events. It is possible to deepen our understanding of the cyber landscape to identify precursors to threats and more readily determine deviations that could indicate hazards.

Cybersecurity is highly dynamic because the underlying technologies are evolving rapidly, and the offense and defense are locked in a threat–response–threat coevolution. This dynamic and constantly evolving landscape requires constant vigilance and updates to threat classification, identification, and response.

Finally, the adversarial nature of the cyber domain presents a modeling challenge that is also an opportunity. Cyber competitions, in which teams act and react to others, are valuable laboratories to explore interactions. The goal of these experiments is to imitate processes by which an adversary learns of defensive measures and then preempts evasive measures. Understanding an adversary's strategy, then, helps refine the models.

We make the following recommendations regarding software and algorithms.

1. *ML should be used as a tool to enhance and extend human cognition.* If models reduce the burdens of routine activity and identify potentially risky activity, the probability of threat avoidance increases. ML shows significant promise in support of forensics, intrusion detection, and attack response.
2. Academic, industry, and government partnerships should *develop game-theoretic models for a deeper understanding of the motivations and behaviors of threat actors*.
3. *Every appropriate form of data should be aggressively leveraged.* NLP techniques can be used to extract artifacts from unstructured data, and Knowledge Graph technology can be leveraged to identify nonobvious relationships between entities while recognizing the data sampling concerns set forth in the "Human Factors" section of this trend paper. These will identify precursors to threat incidents and support automatic detection of nefarious activity.
4. *Systems should be architected around the uncertainty of cyber defense*. Less focus should be given to specific threat indicators (often unknowable) and more to understanding what is different or anomalous. This requires a deeper understanding of what "normal" looks like, so unusual indications can be detected more rapidly and with greater fidelity.
5. ML models are not static; they must adapt as threats develop. To keep pace with developing threats, a system requires the attention of ML scientists. *ML systems need a ready-made development environment*, with easy data access, to facilitate experiments with feature sets and functional forms. It must be simple to push models into production.
6. Academic, industry, and government partnerships must *foster cooperation on modeling advances for particular cyber challenges*. Government and industry organizations should fund academic research and provide sufficient guidance on specific problems requiring creative technical approaches. Similarly, government and industry should *encourage data sharing, so models can be trained* with the most comprehensive data possible.
7. ML focuses on statistically based methodologies, but these are not always appropriate for understanding the dynamics of an adversarial system, as in cybersecurity, where threat actors modify behavior when it becomes ineffective.
8. Models must *adapt quickly to dynamic threats*. Complex models that take weeks to modify, train, and push to production will be too brittle to provide adequate protection. Hybrid techniques that enable quick changes that protect against rising threats could augment robust, carefully trained systems.
9. The effective implementation of an ML-based cyber strategy requires *close integration of diverse expertise*. Cyber and ML experts must collaborate to understand the nature of threats, so implicit uncertainties can be explicitly modeled. Field leaders must find ways to increase professional collaboration.

6. Operationalization: Putting It All Together

The world has finite resources to dedicate to improving cybersecurity, a fact that will inevitably lead to issues of resource allocation. Imagine a future meeting to create an industry or government road map for research and the development of security AI/ML. We believe the participants would agree that properly developed and deployed AI/ML would be highly desirable to give the good guys at the meeting an advantage over bad actors.

But there would be disagreement over which good guy's business model needs protection first—or which nation's laws should provide the template for cybersecurity law and policy.

The counterpoint to the growing size of the cyber physical attack surface is that its growth represents enormous opportunities. Through hardware improvements and proliferation, over the coming decade, organizations will be able to integrate AI/ML into cyberspace operations in ways they would not have anticipated even five years ago. AI/ML will help create integrated meaning from hundreds and thousands of disparate data streams; support automated, real-time prevention platforms; and augment humans' decision-making ability.

Substantial opportunities exist for determining how humans learn to trust AI/ML systems and the entities that use AI/ML. The logical extension of such research is to examine how humans (once they have learned to trust the outputs of the AI/ML systems they interact with) cope with violations of that trust—such as incorrect outputs, lost data, data aggregation across systems that violate privacy expectations, and adversarial manipulation of learning strategies to poison "trusted" systems. This knowledge will ultimately become the rules of the road for a long-term cyber-enabled society. There is a call for collaboration among researchers in fields of personal and organizational trust and the designers, developers, and trainers of AI/ML systems.

As discussed in the section "Human Factors," trust in the technology may require substantial financial support and attention by key decision makers. As it evolves, AI/ML is more likely to reach conclusions or perform actions that humans do not fully understand or that differ from the results of typical human judgment. Handled poorly, recommendations or actions by AI/ML increase the probability that the AI/ML industry will recreate, rather than learn from, experiences such as the nuclear power industry's handling of nuclear plant accidents.

Security fatigue is likely to be a challenge unique to each industry segment. Probabilistic AI/ML systems will need to learn while avoiding misclassification in terms of frequency or severity (in the eyes of the user, not the security specialist) that could lead to distrust and disbelief—electronic versions of the boy who cried wolf, in a sense. The punishment in the story was that the boy was eaten; the outcome in this discussion could be reduced business growth due to general distrust of computer technology.

It is easy to forget the consuming public while industry sectors vie for leadership in cybersecurity or other aspects of computing. There will be new and traditional challenges to the integration of AI and ML into cybersecurity. Repairing or mitigating vulnerabilities will remain a challenge. Most users either do not know or do not have a way to report discovered vulnerabilities. In other instances, involvement in and additional automation of repairing might be rejected by organizations unable to accept much deviation in compatibility and performance.

While the solid political support of small businesses suggests they will have a place at the security AI/ML table, small businesses may be disadvantaged by a lack of data sets or resources to collect such sets. This presents opportunities for larger organizations to productize larger AI/ML solutions or for new organizations to step into the marketplace with meaningful and useful data sets.

Current use cases, such as fraud detection in the banking industry and diagnosis in the health-care industry, serve as enablers for the future operationalization of AI/ML in the cybersecurity domain. Although not all use cases and current AI/ML algorithms are designed to be employed in real-time environments, they serve as foundations for real-time detect–defend or defend–attack situations in cybersecurity. For certain domains, the ability to consciously disable AI/ML actions or disregard recommendations is an enabler of AI/ML operationalization for cybersecurity. In such cases, it is important to have the ability to disable or alter specific system aspects without necessarily turning everything off while, at the same time, comprehending any repercussions.

While understanding and trust may grow on a societal level to eventually allow AI/ML to make response decisions, humans must always have a way to veto those decisions, particularly when preplanned fail-safes fail. However, in many other situations, having AI/ML run closed loop will be fine (perhaps even preferable)—but not always. Clear categorization is required to determine when a human should be in the loop versus when not. For example, safety favors a human in the loop, while limitations in scaling humans' ability to arbitrate favors automation.

Dueling security AI systems is an area ripe for long-term research, as society will eventually need to confront the full potential of AI. Google recently announced that AlphaGo Zero learned how to beat AlphaGo without

human training. Although clearly constrained to a well-structured (though exceedingly large) universe, the trend line from computers beating humans to computers beating other computers will steepen, not flatten. As AI/ML systems gain expertise in conducting, or helping conduct, cyberspace operations, there will come a time when AI/ML will face AI/ML. Learning how to recognize the situation, establishing how to off-ramp or escape the situation, and determining how and when (or even if) to invoke human expertise are all fields of research that must be explored—if for no other reason than knowing bad actors will be using AI/ML to help them achieve their own objectives.

All industry sectors together have a common interest in managing the cybersecurity workforce as it grows and changes its skill mix, driven by the ever-increasing presence of AI/ML. There is historical precedent for workforce evolution in the automotive industry. At the industry's beginning, little effort was required to learn how to maintain and operate an automobile. AI/ML usage and developing trust will not require extensive grounding in the theory and fundamentals of AI; driving a modern car does not require the operator to know the intricacies of the ignition system. However, the AI/ML industry must become better at maintaining and retaining skilled labor to design, build, operate, maintain, and defend AI/ML systems.

The supporting partners for the operationalization of AI/ML in cybersecurity are governments, industry, academia, and the consuming public. At the core, industry partnerships with academia will be the strongest way to bring the research-driven AI/ML capabilities to operational use in cybersecurity. At the government level, research funding for academia and incentives for industry to participate with academia are required. Industry, in the form of consortia, can facilitate workshops and the creation of standards, as cross-company bodies playing a specific role in terms of articulating problems can help illuminate the risk and share best practices. Standards organizations and consortiums like the IEEE have a role in establishing common business practices. Such standardization must rise above the tendency to seek a seal of approval or a checklist of minimal behaviors that are assessed once and then forgotten.

We make the following recommendations on operationalization.

1. *Demonstrate the compelling case that AI/ML systems* embedded within cyberspace operations *make operations better along multiple dimensions*, e.g., speed to patch, remediating a malicious event (or events), increasing up-time in systems of interest, decreasing the number of incidents and the number of false positives, increasing action–reaction–counteraction time cycles, and decreasing unintended consequences of cybersecurity operations decisions and actions.
2. *Retain human and organizational responsibility for decisions* made by the organization's humans and systems. Disclaiming responsibility for organizational actions (or inactions) because the AI/ML influenced or made a decision is unwise and will contribute to public and regulatory backlash.
3. To gain the trust of those humans responsible for cybersecurity operations, *AI/ML systems and their makers must prepare to be transparent about the processes* by which their systems are trained and tested, evolve (at both the operating system/application levels and the data processing/recommendation levels), make decisions, receive and process feedback for improvement, and provide indicators and warnings of being under attack (fast and overt as well as slow and subtle data poisoning).
4. Rigorous academic and industry review of *thought leadership on AI/ML topics in cybersecurity* is needed to address the lack of vetting and openness of practitioner influence. Interdisciplinary review may be applied prior to publication, getting the correct information out. First, publications may result in inaccuracies. Industry should be asked to *fund interdisciplinary policy chairs at leading universities to connect research from industry and academia*. We are at the dawn of publications in this field, and a serious shortage of interdisciplinary AI policy scholars exists.
5. *The evolution of the workforce must be supported by encouraging university curricula in AI/ML, with specific coverage of security*, such that future designers and operators gain a mutual understanding of the limitations and risks.

Summary

AI/ML will become one of the key components of next-generation security, enabling elevated degrees of cybersecurity. At the same time, AI/ML can become a threat used by attackers. In this trend paper, we addressed six different dimensions related to the intersection of AI/ML with cybersecurity: legal and policy

issues; human factors; data; hardware; software and algorithms; and operationalization. As noted earlier, these recommendations are intended for industry (I), academia (A), government (G), and standardization bodies (S). In addition to specific recommendations within each of these six dimensions, we make the following five cross-cutting recommendations, indexed by dimensions (1–5) and to whom they are targeted (I, A, G, or S):

- The future needs of cybersecurity will require an interplay of advances in technology (hardware, software, data), legal and human factors, and mathematically verified trust (1, 2, 3, 4, and 5) (I, A, and G).
- It will require concerted business efforts to establish market-accepted products, certified by established regulatory authorities (1, 2, 4, and 5) (I and G).
- AI/ML-fueled cybersecurity must be based on standardized and audited operations if humans are to trust AI/ML (1 and 5) (I and S).
- Regulators will need to protect research and operations and establish internationally recognized cooperative organizations (1 and 2) (S and G).
- Data, models, and fault warehouses will be essential for tracking progress and documenting threats, defenses, and solutions (3, 4, and 5) (S, I, and A).

Our recommendations can be applied at different time horizons. Operationalization takes the least time and could be accomplished in under two years. This is similarly true for data and software. Legal and policy issues take longer, up to five years. Hardware, e.g., new processor architectures, typically takes more than five years to materialize. It will be essential to continue evaluating and advancing contributions of AI/ML to cybersecurity through focused efforts of governments, industry, and academia.

Afterword: Background, Motivation, and Overview

The IEEE has a rich and distinguished heritage dating back to the American Institute of Electrical Engineers, founded in 1884, and the Institute of Radio Engineers, founded in 1912. Notable early presidents of the IEEE and its founding organizations were engineers and practitioners, including Alexander Graham Bell, Charles Proteus Steinmetz, Robert H. Marriott, William R. Hewlett, and Ivan Getting. Over the decades, IEEE membership has fundamentally changed, with those working in industry increasingly outnumbered by academics. And this trend continues, with the number of IEEE Members who identify industry as their employer continuing to decline. Since 2000, the percentage of IEEE Members from industry has fallen from roughly 60% to 39%. Our content has diminishing relevance to industry because it is progressively more academic in nature. Our career development efforts are not optimally aligned with emerging industry needs.

Over the past several years, the IEEE leadership has taken great strides to engage with industry and mounted a concerted effort to provide technical professionals with the tools and information they need to excel. We have aggressively engaged with industry to understand its needs along with those of Members who work in industry and so bring forth products and services of value and importance. In 2015, we met with over 175 industry leaders from 45 companies in China, Germany, Japan, and Silicon Valley in the United States. In 2016, we met with over 270 leaders from 70 companies in Canada, China, India, Israel, Japan, Singapore, South Africa, South Korea, Taiwan, the United Kingdom, the United States, and Uruguay. These discussions provided important insights into industry needs. One recurring theme heard from a wide variety of different industries was the importance of technology trend papers and road maps. As a result of this input, we responded by chartering two trend papers in 2016, one on 5G and a second on smart cities. These two trend papers were delivered in the third and fourth quarters of 2017, respectively. While the content of these trend papers was valuable, the more than 12-month delivery time was contrary to industry's need for rapid and relevant information. To more quickly deliver contemporary and relevant trend papers, we considered an alternative model.

In partnership with Syntegrity, a group having a long-standing relationship with the IEEE, we conceived the idea of bringing together a group of experts in a technology vertical and using the Syntegration process to rapidly develop a technology trend paper. After careful consideration of the technology landscape and those areas with the greatest interest and impact, we chose the intersection of AI and ML as applied to the broad field of cybersecurity. In this context, cybersecurity encompasses the financial services, critical infrastructure

(e.g., smart grid and SCADA [supervisory control and data acquisition]), and defense sectors. Syntegrity combined insights from geometry, neurology, and cybernetics with advanced mathematical models and social technologies in the Syntegration process, which enables group interaction to consolidate thinking and ultimately formulate solutions in dramatically compressed time frames.

On 6–8 October 2017, we convened 19 experts from the AI, ML, and cybersecurity sectors in Philadelphia, Pennsylvania, United States, for a two-and-a-half-day collaborative session focused on the following complex question:

> *Given the rapid evolution of AI/ML technologies and the enormous challenges we all face with respect to cybersecurity, what is needed from AI/ML, where can it be best applied, and what must be done over the next ten years?*

During the first day, the group, as a whole, identified challenges associated with the question, proposed multiple topics for discussion that could potentially address the question, and then collectively prioritized six specific topics the group believes must be addressed to answer the question. Over the remaining two days, the group conducted iterative and focused discussions regarding each of the six topics to reach a more refined understanding of the challenges and identify the most viable solutions. By the end of the confluence, the group produced a draft of this trend paper that will be shared with the greater community to address the challenges associated with the question.

References

[1] B. D. Johnson. (2017, Mar.). A widening attack plain. [Online]. Available: http://threatcasting.com/wp-content/uploads/2017/03/A-Widening-Attack-Plain.pdf

[2] BBC News, Volkswagen: The scandal explained. [Online]. Available: http://www.bbc.com/news/business-34324772

[3] http://fortune.com/2017/10/25/reaper-botnet-mirai-iot-ddos/

[4] A. M. Matwyshyn, "CYBER!," *2017 BYU L. Rev.*, vol. 101, 2018.

[5] See 80 FR 208, 65956. [Online]. Available: https://www.copyright.gov/fedreg/2015/80fr65944.pdf

[6] http://www.wassenaar.org/

[7] http://www.marketwatch.com/story/volkswagen-diesel-emissions-fixes-approved-2017-10-23

[8] https://www.csoonline.com/article/3122460/techology-business/over-6000-vulnerabilities-went-unassigned-by-mitres-cve-project-in-2015.html

[9] https://www.csoonline.com/article/3146046/security/security-products-are-among-the-most-vulnerability-riddled-software-products.html

[10] https://www.wired.com/2017/05/accidental-kill-switch-slowed-fridays-massive-ransomware-attack/

[11] https://www.nytimes.com/2017/09/18/world/europe/stanislav-petrov-nuclear-war-dead.html?_r=2

[12] See, e.g., http://psc.dss.ucdavis.edu/sommerb/sommerdemo/sampling/intro.htm

[13] https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist

[14] https://www.darpa.mil/program/cyber-grand-challenge

[15] IEEE Cybersecurity Initiative. Available: https://cybersecurity.ieee.org/

[16] IEEE Ethically Aligned Design (EAD), Version 2, *A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems*.

[17] See, e.g., ISO 29147 and 30111.

[18] https://www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm

[19] https://christian-rossow.de/publications/sandprint-raid2016.pdf

[20] https://www.rand.org/pubs/research_reports/RR1751.html

[21] https://www.wired.com/2016/12/botnet-broke-internet-isnt-going-away/

[22] J. Mirkovic, S. Dietrich, D. Dittrich, and P. Riher, *Internet Denial of Service: Attack and Defense Mechanisms*. London: Pearson, 2004.

[23] https://www.arm.com/products/security-on-arm/trustzone
[24] S. Borkar and A. Chien, “The future of microprocessors,” *Communications ACM*, vol. 54, no. 5, pp. 67–77, May 2011.
[25] R. Gioiosa, D.J. Kerbyson, and A. Hoisie, "Quantifying the energy cost of data movement in scientific applications," in *Proc. Energy Efficient Supercomputing Workshop*, 2014, pp 11–20.

Other material used in preparation for the paper includes the following:

[1] ISACA. (2017, Feb.). State of cyber security 2017: Current trends in workforce development. [Online]. Available: http://www.isaca.org/Knowledge-Center/Research/Documents/state-of-cybersecurity-2017-part-2_res_eng_0517.pdf
[2] ISACA. (2017, June). “State of cyber security 2017: Current trends in the threat landscape.” Available: http://www.isaca.org/Knowledge-Center/Research/Documents/state-of-cybersecurity-2017-part-2_res_eng_0517.pdf
[3] *The New Dogs of War: The Future of Weaponized Artificial Intelligence*. Available: http://threatcasting.com/wp-content/uploads/2017/09/ThreatcastingWest2017.pdf
[4] https://www.rsaconference.com/writable/presentations/file_upload/spo1-t11_combatting-advanced-cybersecurity-threats-with-ai-and-machine-learning_copy1.pdf
[5] https://www.rsaconference.com/writable/presentations/file_upload/exp-t11-advances-in-cloud-scale-machine-learning-for-cyber-defense.pdf
[6] https://www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide/
[7] https://www.dhs.gov/sites/default/files/publications/CSD-DHS-Cybersecurity-Roadmap.pdf
[8] A. Epishkina and Sergey Zapechnikov, “A Syllabus on Data Mining and Machine Learning with Applications to Cybersecurity.” Proceedings of the 2016 Third International Conference on Digital Information Processing, Data Mining, and Wireless Communications (DIPDMWC), Moscow, 2016, pp. 194-199.
[9] D. Zhu, H. Jin, Y. Yang, D. Wu, and W. Chen, “DeepFlow: Deep learning-based malware detection by mining Android application for abnormal usage of sensitive data,” 2017 IEEE Symposium on Computers and Communications (ISCC), Heraklion, 2017, pp. 438-443.
[10] J. B. Fraley and J. Cannady, “The Promise of Machine Learning in Cybersecurity.” SoutheastCon 2017, Charlotte, NC, 2017, pp. 1-6.
[11] A. Tuor, S. Kaplan, B. Hutchinson, N. Nichols, S. Robinson, “Deep Learning for Unsupervised Insider Threat Detection in Structured Cybersecurity Data Streams.” Proceedings of AI for Cyber Security Workshop at AAAI 2017.
[12] K. Alrawashdeh and C. Purdy, “Toward an Online Anomaly Intrusion Detection System Based on Deep Learning.” 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA), Anaheim, CA, 2016, pp. 195-200.

Participants

Confluence Participants

David Brumley	For All Secure and Carnegie Mellon University
Robert K. Cunningham	MIT Lincoln Laboratory
Chris Dalton	HP Inc.
Erik DeBenedectis	Sandia National Laboratories
Flavia Dinca	Stockholm University, Sweden
William G. Dubyak	IBM Watson Group
Nigel Edwards	Hewlett Packard Enterprise
Rhett Hernandez	U.S. Department of Defense
Bill Horne	Intertrust Technologies

Brian David Johnson	Arizona State University
Aleksandar Mastilovic	University Novi Sad, Serbia
Andrea M. Matwyshyn	Northeastern University
Abraham (Avi) Mendelson	The Technion–Israel Institute of Technology
Dejan Milojicic‡	Hewlett Packard Enterprise
Katie Moussouris	Luta Security, Inc.
Adrian L. Shaw	ARM Ltd.
Barry Shoop‡	U.S. Military Academy, West Point
Trung Tran	Laboratory of Physical Sciences
Mike Walker	Microsoft Corporation

Technical Writers

Glenn Zorpette	IEEE, *IEEE Spectrum*
Clay Moody	U.S. Military Academy, West Point
Mike Lanham	U.S. Military Academy, West Point
Matt Sherburne	U.S. Military Academy, West Point
Daniel Hawthorne	U.S. Military Academy, West Point

Observers

Donna Hourican	IEEE
Providence More	IEEE

The participants brought expertise from a wide variety of sectors. David Brumley and his team from For All Secure won the 2017 DARPA Cyber Grand Challenge. Mike Walker was the DARPA program manager who developed and offered the DARPA Cyber Grand Challenge. Will Dubyak from the IBM Watson Group is applying Watson's NPL to cybersecurity. Brian David Johnson, previously Intel's futurist, has recently been applying future casting and threat casting to the area of cybersecurity. Rhett Hernandez served as the first commanding general of the U.S. Army Cyber Command. Dr. Robert K. Cunningham chairs the IEEE Cybersecurity Initiative and leads the Secure Resilient Systems and Technology Group at MIT's Lincoln Laboratory. Trung Tran has worked for Intel and HP and, more recently, works for the federal government on building the next generation of AI. Andrea Matwyshyn, a professor at Northeastern University, focuses on technology innovation and its legal implications, particularly corporate information security regulation and consumer privacy. Katie Moussouris is a computer security researcher who created the bug bounty program at Microsoft, was chief policy officer at HackerOne, and was named one of "10 Women in Information Security That Everyone Should Know." Erik DeBenedectis is a member of the technical staff at Sandia National Laboratories, leading a project to build a petaflops-scale supercomputer, and is also deputy project lead for the ASCI Red Storm supercomputer. Adrian Shaw is a security architect at ARM with experience in securing software-defined services to mitigate threats in the IoT. Abraham (Avi) Mendelson served at Intel and Microsoft prior to joining the Technion, where he focuses on operating systems, computer architecture, high-performance computing, and cloud computing. Chris Dalton is a distinguished technologist at HP Inc. and leads the Platform and Device Security Research Group within HP Labs. Nigel Edwards is a distinguished technologist at Hewlett Packard Labs, where he leads the Security Research Group. Bill Horne is a vice president at Intertrust Technologies, where he is general manager of the Secure Systems Division. Flavia Dinca is an information security Ph.D. degree student at Stockholm University, with a background in the social implications of technology and policy. Aleksandar Mastilovic is the EU Marie Curie Fellow at the University of Novi Sad, Serbia.

The writers focused on capturing the dialog and debate during the collaboration engagements. Glenn Zorpette is a senior technical editor for *IEEE Spectrum*. Clay Moody, Mike Lanham, Matt Sherburne, and Daniel Hawthorne are all U.S. Army Cyber Branch officers and faculty in the Department of Electrical Engineering and Computer Science at the U.S. Military Academy at West Point.

Dejan Milojicic is a Distinguished Technologist at Hewlett Packard Labs, past president of the IEEE Computer Society, and chair of the IEEE Industry Engagement Ad Hoc Committee. Barry Shoop is a professor and head of the Department of Electrical Engineering and Computer Science at the U.S. Military Academy, West Point, and served as 2016 IEEE president and chief executive officer.

‡ Project sponsors.