IDC PERSPECTIVE

# A CISO's Guide to Artificial Intelligence

Frank Dickson, Christopher Kissel, Michelle Abraham, Philip Bues
Jennifer Glenn, Mike Jude, Ryan O'Leary, Grace Trinidad

## EXECUTIVE SNAPSHOT

## FIGURE 1

### Executive Snapshot: A CISO's Guide to Artificial Intelligence

As artificial intelligence (AI) is applied to the cybersecurity use cases, the goal is to create analytics platforms that capture and replicate the tactics, techniques, and procedures of the finest security professionals. Leveraging AI to drive security outcomes is top of mind for CISOs, but there are factors to consider.

#### Key Takeaways

- AI offers the potential to democratize the traditionally unstructured threat detection and remediation process by bringing the greatest expertise to all. The near-real-time automated detection and response techniques may be completed far more quickly by a machine than a human resource, ensuring that the security response is timely.
- Data is more important to successful AI implementations than analytics since data is the enabling infrastructure for security AI.
- AI is a force multiplier for the power of security professionals, having both positive and negative (or at least unintended) effects.

#### Recommended Actions

- No matter what AI is, focus on the benefit. Ultimately though, the offering needs to deliver value in *your* environment. You should force your vendor to demonstrate true value.
- Start with an outcome in mind. If the AI is to provide a benefit, it can be measured —objectively — with metrics. If the benefit cannot be objectively measured, you need to question if the "benefit" is real.
- Note that transparency is now more important than ever. You should expect vendors to provide detailed information on how and where system data is treated.
- Create decision trees. SOC analysts need lists of predictable outcomes from remediation actions taken.
- Demand low code, no code, or natural language processing (NLP).

Source: IDC, 2023

July 2023, IDC #US50933923

## SITUATION OVERVIEW

Microsoft's launch of Microsoft Security Copilot, which leverages the power of the generative artificial intelligence (AI) natural language model ChatGPT, has created a buzz around the topic of AI and its application to security. Google Cloud Security AI Workbench, AWS Titan/Bedrock, and IBM watsonx only served to intensify the buzz. We would all agree that greater analytics and automation are the future of security, but that future is much more than generative AI.

AI in security is hardly new. Analytics have been the foundation of cybersecurity since the very beginning. Machine learning (ML) and artificial intelligence found some of their early commercial applications in cybersecurity. Think about it. Products like McAfee Deep Defender, FireEye Malware Protection System, and Cisco Advanced Malware Protection were launched more than 10 years ago!

As vendors look to differentiate their offerings and appeal to sophisticated, knowledgeable, and astute cybersecurity buyers, marketers commonly promote buzzwords. Yet buzzwords are seldom defined and rarely have a shared meaning among practitioners in the industry. The result is confusion.

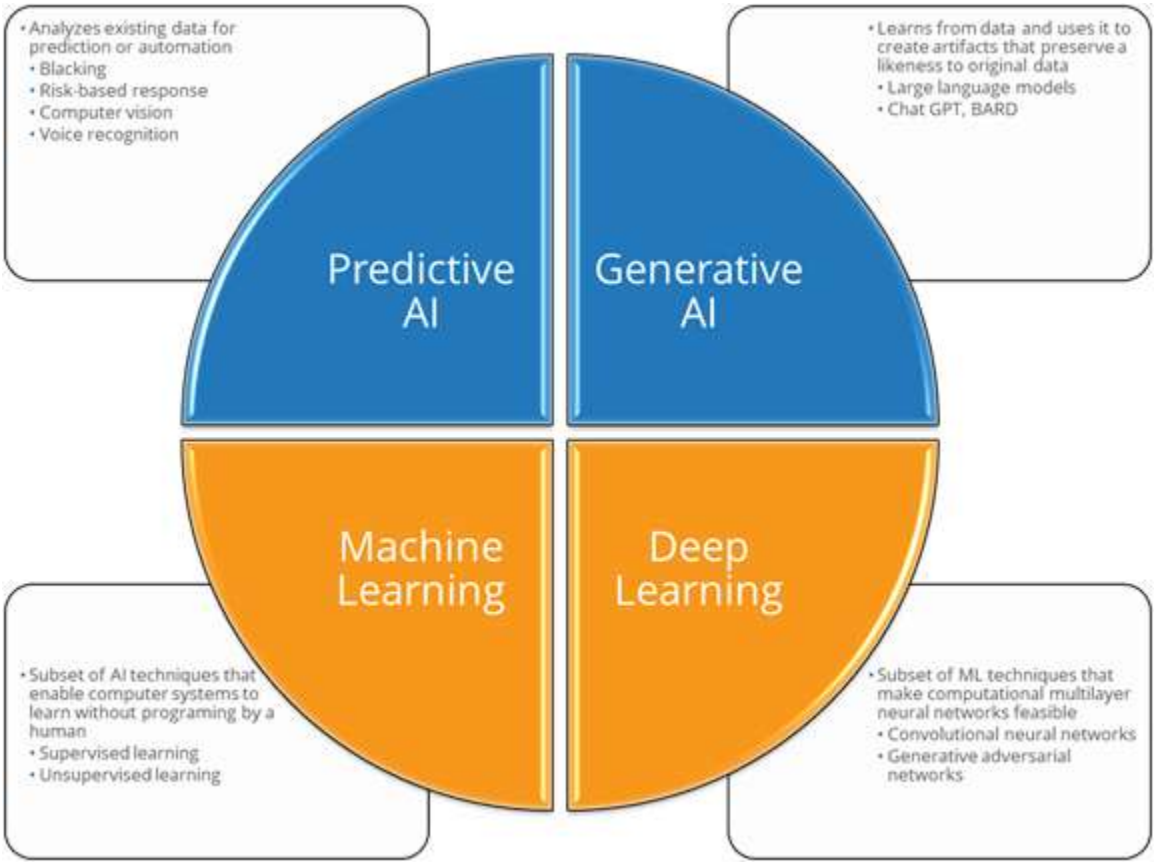
IDC does not want to contribute to the confusion. As a result, we present our formal definitions of artificial intelligence and machine learning pertaining to cybersecurity in this document, including illustrative examples, and provide a construct to frame the topic and guide thinking to help provide clarity.

### Defining Artificial Intelligence

Artificial intelligence comprises a grouping of machine-based technologies that perceive and synthesize data to infer information and insight to create systems that learn, reason, adapt, and self-correct. Example tasks include security intelligence, speech recognition, computer vision, translation between (natural) languages, as well as other mappings of inputs. Techniques such as machine learning that enable computer systems to learn without programming by a human or deep learning that make the computational multilayer neural networks feasible (see Figure 2).

## FIGURE 2

### Defining Artificial Intelligence




Source: IDC, 2022

## *Artificial Intelligence as Applied to Security*

As the term is applied to the narrow use case of cybersecurity, we define artificial intelligence as providing advisory, enhanced service, and semiautonomous cybersecurity defense functionality based on a range of structured and unstructured data, including logs, device telemetry, network packet headers, and other available information. Simply put, AI is the application of applied statistics to solve cybersecurity problems. The goal is to create analytics platforms that capture and replicate the tactics, techniques, and procedures of the finest security professionals; democratize the traditionally unstructured threat detection and remediation process; or complete a range of near-real-time automated detection and response techniques that theoretically can be replicated, but by the time the security professional completed the task, it would be far too late. The development approach typically begins with the mundane and remedial and gradually graduates to increasingly complex use cases. Using large amounts of structured and unstructured data, content analytics, information discovery, and analysis, as well as numerous other infrastructure technologies, AI-enabled security platforms use deep contextual data processing to answer questions, provide recommendations and direction, hypothesize, and formulate possible answers based on available evidence. The models are trained through the ingestion of vast amounts of content and automatically adapt and learn from their mistakes

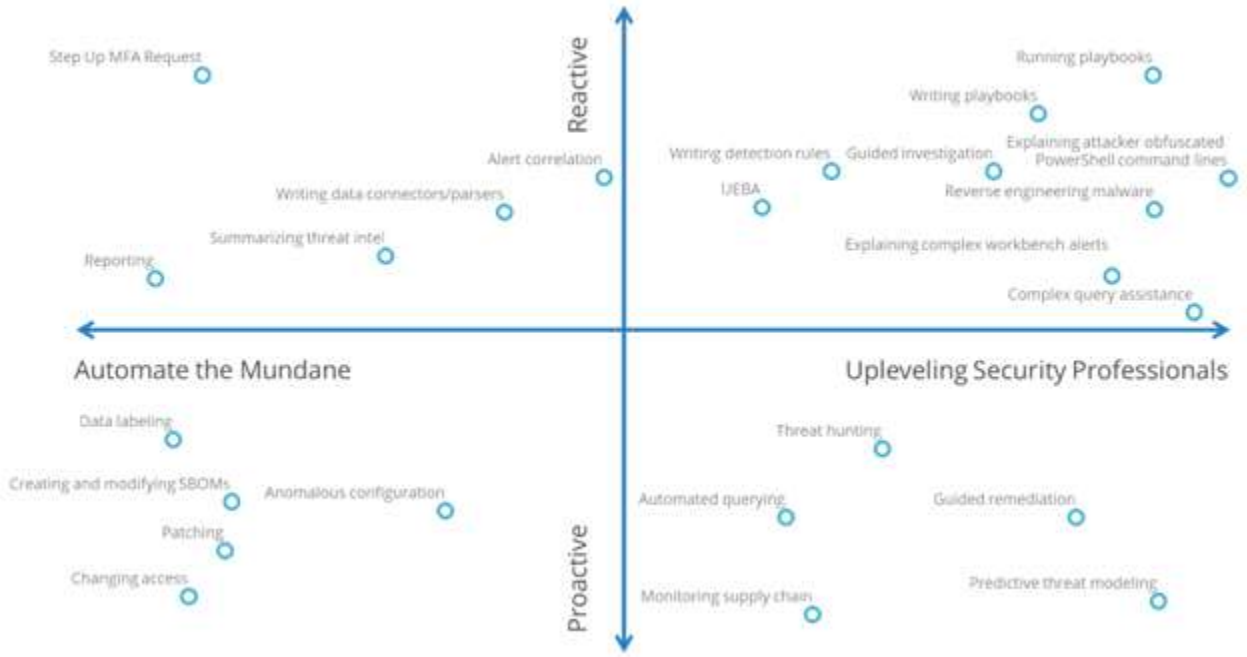
and failures to augment the capabilities of or enhance the efficiency of an organization's most precious cybersecurity assets — its scarce cybersecurity professionals. The technology components included range from unstructured information handling and knowledge extraction to machine learning/deep learning, including supervised and unsupervised learning to hypotheses generation to question answering. Specifically, these tools are used to build smarter applications that have the potential to learn and improve over time.

Note that the use cases being addressed by machine learning (in cybersecurity applications) are often very different than other technologies within the broader domain of artificial intelligence (refer back to Figure 2). IDC limits the definition of machine learning in cybersecurity applications to pattern recognition in large sets of structured data. Based on our simple description of AI as the application of applied statistics to solve cybersecurity problems, machine learning illuminates patterns. For example, machine learning is used to create algorithms based on large sets of malicious and benign executables that produce reliable, repeatable judgments regarding the potential maliciousness of a file. Some forms of machine learning are typically applied to repeatable and known use cases/problems, while more complex forms of machine learning, such as neural networks and other AI technologies, are applied to more complex and unstructured problems.

Both artificial intelligence and machine learning are being applied to cybersecurity use cases to address fundamentally the same problem: the cybersecurity workforce shortage by targeting the workforce shortage, upskilling SOC analysts, or detection augmenting capabilities by providing higher-fidelity, higher-speed detection in myriad ways, such as identifying malware, spotting phishing, analyzing target web pages for credential prompts, and so forth. In theory, one is looking to make security professionals more effective, while the other is looking to improve efficiency. Figure 3 provides illustrative examples. The reality is that the scale of machine learning has grown to such a point that problems solved can exceed what a single person can pattern match in a lifetime or even a thousand lifetimes, but the principle is still valid.

## FIGURE 3

### Illustrative Examples of the Practical Application of AI to Cybersecurity




Source: IDC, 2023

Machine learning makes security professionals more effective as it provides the ability to analyze structured data at a terabyte scale and identify patterns that indicate maliciousness. Eric Chien, a Symantec fellow, described cybersecurity machine learning as:

> *Feeding large amounts of data about both malicious and legitimate files into an algorithm. The algorithm outputs a "classifier" that can then be used to look at a new file it has never seen before and determine if that file, or that URL, or that situation on an endpoint, whatever the particular in question may be, is good or bad. Previously, writing a classifier was always work done by human analysts, but machine learning allows for it to be done in an automatic way without a human needing to write the program. The machine becomes the analyst.*

Essentially, machine learning allows cybersecurity professionals to find the malicious "needle" in a haystack of data, which humans are much less likely to see in a timely manner unless they stumble upon it.

Other implementations of artificial intelligence, in contrast, are being applied to make security professionals more efficient. To illustrate, in large enterprises and managed security services providers, the typical ratio of level 1 to level 2 SOC analysts is 3:1, an approximation made by IDC in its conversations with clients. The goal is to change that ratio to 5:1 in the coming years. The only way to get there is to make level 1 SOC analysts more efficient and augment the abilities of the more experienced level 1 SOC analysts to allow them to perform at level 2.

The recent wave of chatbots from a variety of security vendors would be an example of how AI is helping uplevel SOC analysts — "Please explain this alert to me," "Please explain this PowerShell script to me," "Help me search for any cases of PowerShell being launched from Microsoft Word documents," and so forth.

For a more sophisticated example, event sequencing is a core attribute of most detection and response offerings. Traditionally, security information and event management (SIEM) systems were very good at detecting rule-based violations. However, the detection of anomalies can be problematic if the type or veracity of an attack is not understood. A detection and response offering like a SIEM or an endpoint detection and response (EDR) can process a series of notable events (alerts) and correlate with risk modifiers to reduce the investigations (grouping a number of alerts into a single "incident") and give SOC analysts increasingly comprehensive visibility into the type of attack that is occurring. In addition, security platforms automate event sequencing to group multiple alerts into a single "incident," leveraging AI and complementary expert rule sets and allowing a security analyst to address alerts at scale. We would be remiss if we did not mention that the standard for event sequencing is the MITRE ATT@CK (adversarial tactics, techniques, and common knowledge) framework, a guideline for classifying and describing cyberattacks and intrusions. The framework consists of 14 tactics categories consisting of "technical objectives" of an adversary. Examples include privilege escalation and command and control.

## *User Behavioral Analytics — Somewhere Between Machine Learning and Artificial Intelligence*

A mesh point exists between the tier 1 analyst role that we assign to machine learning and the tier 2-3 analyst role that we place as in the realm of artificial intelligence (refer back to Figure 3). An apt description of that mesh point is user behavioral analytics (UBA).

UBA is an applied set of AI technologies across multiple domains, such as endpoint detection and response, threat analytics, and SIEM. User and entity behavioral analytics (UEBA) is an analytic set that establishes individual relationships for each individual user and entity in the network. The application of UEBA establishes a statistical baseline that is both a basis for deviant behavior and can also be a "golden image" of what a user/entity is and what it should be doing.

The statistical profile of individual entities is vitally important in a network for several reasons. In the first place, a zero-day threat is the type of malware that infiltrates the network and goes undetected until it detonates. If a new malware design evades a cybersecurity perimeter, UEBA may find evidence of the malware as it tries to beacon comply-to-connect (C2C) servers or exhibits other anomalies. Another difficult threat to defend against is an insider threat. An insider threat can generally occur in three different instances: a careless insider that has created an exposure, a deliberate malicious insider that is exfiltrating data, and a miscreant that has created a privilege escalation and, for all intents and purposes, looks like an insider. What is common in all three instances is that the end user or the device is now conducting communications in a secure tunnel. Basically, an end user who is using legitimate network protocols and initiates inbound/outbound communications tools such as intrusion detection systems (IDS)/intrusion prevention systems (IPS), next-generation firewalls, and SIEM is difficult to differentiate this nefarious activity from legitimate activities on the network. The last use case involves the Internet of Things (IoT). As billions of devices interact with a larger network, machine-to-machine (M2M) onboarding occurs without roles assigned by administrators and without agents. Individual and peer group analytics may be the only way to find anomalies at all.

However, acknowledging that UEBA is not what it will be in 5 or 10 years' time, a security team should be aware of a few things that are problematic. The companies that are strong at UBA will intimate that their platforms get stronger the more data that you feed them. Obviously, that means there is a nascent stage in deployment where a UBA platform does require other tools while it self-tunes. Second, there is a fear that without proper controls, repeated bad behavior begins to look like good or non-anomalous behavior. Last, for the same reasons that sandboxes, and deception are used to mimic an authentic network, it may be possible in the future for a sophisticated hacker to undermine the platform by adding so much meaningless data that as the analytics engine is analyzing data streams, the true intention of the adversary remains hidden (roughly akin to a buffering error attack).

Note that IDC does make a distinction between user behavioral analytics and user and entity behavioral analytics. UEBA is a superset category that examines devices, instances, and identities from a "black box" perspective, establishing "normal" and looking for significant deviation from normal. Many describe UBA as anomaly detection, which is appropriate. As a result, UBA tends to be more ML centric than AI centric. UEBA removes the black box and applies identity context to the problem; UEBA operates under the premise that activity that is not anomalous may yet not still be allowed. For example, an identity vendor in introducing its new UBA had wonderfully successful product demonstrations in its first two beta customer interactions. At the third product demonstration, nothing happened. The platform had now determined that downloading personally identifiable information and emailing it to a Gmail account was no longer anomalous for that use. Identity context and proper policy creation prevent the problem. As a result, UEBA tends to be more AI centric and less ML centric.

## It's Not About the AI; It's About the Data

The hype and conversation focus on AI. Why not? The possibilities of AI inspire the imagination, illuminating the possible. However, the key to enabling outcomes in security is not about the AI; it is about the data. Many children are inspired by the power and girth of locomotives. The potential of the locomotive, though, relies on the boring and tedious process of laying the tracks and the enabling infrastructure. Likewise, data is the enabling infrastructure for security AI. Three characteristics are deterministic of success:

- Data framework structures
- Data management
- Data curation

### *Data Framework Structures*

As we look to unlock the potential of artificial intelligence to unlock the potential and promise of extended detection and response (XDR), for example, creating frameworks and structures is critical. The most basic definition of XDR is:

- The collection of telemetry from multiple security tools
- The application of analytics to the collected and homogenized data to arrive at a detection of maliciousness
- The response and remediation of that maliciousness

As we look to apply analytics to the collected and homogenized data to detect maliciousness, AI needs structure to be able to look at the data at scale. Telemetry optimized for a point use case, such as the perimeter-centric defense of network perimeters of a firewall, is of little use if you cannot relate it with other data sets, such as identity, and if it is not framed in a way to achieve an end goal. As we

discussed the value of event sequencing as a core attribute of most detection and response offerings, much of the value was unlocked by application of the MITRE ATT@CK framework. Not only does the framework provide structure to the task of threat detection by mapping to the cyber kill chain, but it also creates a manner in which different tools from different vendors can structure data and prepare it for analysis.

## *Data Management*

Data has weight. Security data has a lot of weight. For example, a typical endpoint protection platform agent will produce 150-200MB of data a day. Movement, storage, and management of such data quickly create a problem of scale. Data retention policies thus can become quickly divisive topics.

In addition, only with AI can the increasing pools of telemetry be put to the very best use. ML has limits, but using AI to train for previously unseen patterns and lens on the data can (time-to-X) be reduced in a truly significant way.

Data weight has become a competitive differentiating tool. For example, the move by the infrastructure-as-a-service (IaaS) vendors to retain their own cloud logs at no or very low cost is significant, as SIEM is often priced based on the volume of data ingested, and the SIEM vendors cannot simply "eat" the cost of ingesting and storing voluminous cloud logs. Analysis needs to happen on the native format in a predictable manner. The entire business model of SIEM, XDR, and other analysis platforms thus is increasingly challenged and is changing based on the weight of data.

## *Data Curation*

In a world where every vendor has a different data structure, curating heterogeneous data sets to create data homogeneity to enable analysis is an extra step, a potentially ominous step depending on the calculus and scale required. As AI continues to promise simplicity in the face of the complexity of today's security environment, it will be helped by the homogeneity of data. In a world where every vendor has a different data structure, curating heterogeneous data sets to create data homogeneity to enable analysis is an inhibitor.

Restructuring data takes time and costs money. Thus large vendors with broad portfolios have the advantage as multiproduct but single platform offerings save time and cost due to having a larger percentage of multi-technology homogeneous data sets.

Overcoming the issue of data curation is the objective of many standards. For example, Structured Threat Information Expression (STIX) and Trusted Automated eXchange of Indicator Information (TAXII) were developed by MITRE as the U.S. Department of Homeland Security FFRDC. STIX is a common language for threat intelligence, so it can be shared and machine-read by any tool supporting it. TAXII is the application layer protocol designed to simplify the transmission of threat intel data. In 2015, STIX/TAXII development was moved to the OASIS international standard organization. Today, the work is free, open, and community driven.

We would be remiss if we did not mention Open Cybersecurity Schema Framework (OCSF) here and its significance to AI. Normalization of hybrid multicloud security telemetry is needed before any converged data is useful. The goal of OCSF is to simplify the exchange of data between the tools that ingest it, manage it, and enrich it because every organization has a cornucopia of solutions purchased over the past half dozen years. OCSF means a single format to make it easy for those getting started instead of writing data connectors to a lot of solutions. The real story here is one of simplicity, which is the holy grail of cybersecurity solutions.

The 18 founding members of the OCSF are AWS, Broadcom, Cloudflare, CrowdStrike, DTEX, IBM Security, IronNet, JupiterOne, Okta, Palo Alto Networks, Rapid7, Salesforce, Securonix, Splunk, Sumo Logic, Tanium, Trend Micro, and Zscaler. Of the groups, AWS has the strongest voice. OCSF is not being driven by a single security vendor, looking to create a standard that provides unfair advantage to it. AWS' goal is to not necessarily profit from security but to make the cloud more secure, so it limits objections to cloud adoption, thereby monetizing its efforts by selling more cloud. A somewhat impartial perspective is relevant.

AWS also has the market might to dictate standards. A cloud security vendor only needs to comply with the standard if it wants to sell solutions that work in AWS, which is clearly all of them. Very few customers are comfortable with a solution that cannot work in AWS.

The OCSF road map follows best practices, and some concepts are still in beta. This reflects the diversity in thought of its members and recognition that our enemies are always innovating, and we need to be as well. The OCSF game plan follows these best practices for consideration, acceptance, and participation — much like a community: open source, lightweight governance, and extensibility.

## Concerns and Considerations in Leveraging AI

According to Spiderman, "With great power comes great responsibility." The concern is valid for AI. AI is a force multiplier for the power of security professionals, having both positive and negative (or at least unintended) effects. There are considerations that should be examined.

### *Present Limits to AI*

In reality, artificial intelligence is not particularly smart. Please see our simple description of AI as the application of applied statistics to solve cybersecurity problems. Machine learning illuminates patterns; AI applies patterns. As a result, machine learning and AI are great at recognizing patterns in data lakes and responding at a scale that no human observer could, and it can compare static conditions over time, looking for things like configuration drift or deviation from a golden state that happen slowly, almost imperceptibly. AI is also able to do repetitive tasks more precisely and many magnitudes faster than humans can.

However, AI really leverages existing knowledge but cannot create new constructs; essentially, it does not "think." For example, IDC does not feel that a general-purpose AI engine can write a software development kit (SDK) from scratch. Intelligent coding solutions leveraging large language models (LLMs) based upon GPT-3–trained models using humanly written code (e.g., GitHub Copilot) can generate code for developers. In this sense, these tools improve developer productivity by removing mundane aspects of coding, but they cannot replace a developer. In addition, there is a focus on creating domain-specific LLMs trained in subject matter specific to the topic at hand, as well as potentially customer data. Even after all of that, ChatGPT and its competitor, Google Bard, still make glaring errors, often misunderstanding simple prompts or generating outcomes that make no sense.

### *Hallucinations and the Role of the Analyst*

The term "hallucination" in AI is simply reminding the user that the output is only as reliable as the human that built the model and the accuracy of the data in any corpus — not perfect. Every day, more and more data is created. IDC forecasts that 221,178EB of data will be created in 2026. The problem is that if you throw in everything (time between packets arrivals/departures, headers from session data, and handshake protocols), there is an insane amount of pattern recognition occurring, but little of it is useful.

Creating value requires context, but context can be challenging. Take the "orange" for example. Orange can be a color, a flavor, a fruit growing on a specific tree, a staple of nutrition as a source of Vitamin C, the national color of the Netherlands, or a massive global telecom provider. The intention of the word "orange" is easily interpreted by humans almost immediately when used in context. Machine learning must create the means to understand the context, infer the meaning and, last but still the hardest, what to do with the information it has acquired.

AI/ML will not now or likely ever be fully trustworthy, not requiring review by a human expert. Just as planes have autopilot to augment some operations, you still have licensed, trained human pilots in the cockpit for takeoff, landing, and other unforeseen mechanical/electrical failures. While AI/ML training and certification programs are being delivered, AI/ML may face challenges ramping up with not the technology per se but with the ways in which humans will use it to enrich data and make decisions. These best practices will be created and debated over time.

Relying on AI to guide analysts and report on security events will only take off if the models are trustworthy. The data used to train the model must be accurate, or the decisions made with AI will not go the way desired. The terms confabulation or hallucination are being used to describe when a model is wrong because the models are trained to give some answer instead of saying I don't know when they do not have an answer. Customers and AI suppliers should understand what data underlies each decision so they could figure out if training went wrong and how the model needs to be retrained. The industry needs to avoid AI bias, which occurs if the training data set chosen is not diverse enough.

A vendor needs to protect the data used to train the model because if it is breached, the model will not be trained correctly; for example, it could be trained to ignore malicious behavior instead of flagging it. A customer needs to have guardrails to ensure that it is keeping proprietary data out of public models. Vendors will also need to check for drift with their models; they will need to be tuned in and updated with new information. An AI model is not something that can be set up and forgotten.

## *Data Security and Privacy Risks*

Managing and controlling the integrity and privacy of growing data stores has been a challenge for security practitioners for some time now. Now, the same data security and privacy challenges that are currently vexing an overburdened and underfunded security team are happening at an exponential rate. Sensitive or confidential information may be uploaded into an AI model with little to no thought about the long-term implications. Individual data sets that are innocuous when standing alone may be combined, thus creating personally identifiable information (PII). Frankly, it's another (powerful) tool that requires security awareness training for users to operate appropriately.

Further, there are compliance issues to consider. Data privacy regulations like the European Union's General Data Protection Regulation (GDPR) threaten large fines and regulatory actions against enterprises that violate them. Data privacy regulations attempt to ensure that personal data is not processed without the consent of the data subjects.

The risk for organizations is in the processing of data that they do not have consent to process or do not have any other legal basis to process. Vendors that are experimenting with generative AI must be cognizant of these issues. If generative AI, as some have announced, is being put into products without working with data classification and other enterprise rules and then being trained on the entire data corpus of an organization, then end users are being exposed to massive risk. It is possible, if not probable, that organizations have data that they do not have the right or consent to process.

The concern exists not only for existing but also for new sensitive information being created from otherwise nonsensitive information. In February 2020, IDC wrote about an issue that Google had concerning the Illinois Biometric Information Privacy Act (BIPA) (see *Move Over CCPA, Illinois' BIPA May Have the Bigger Privacy Impact,* IDC #IcUS46024620, February 2020). The 2008 law prevents organizations from obtaining or possessing an individual's biometric data unless the data subject is informed in writing that the data will be collected, the specific purpose and length of time it will be collected for, receives a written release to collect the data, and publishes publicly available retention schedules for the data. On February 6, a class action lawsuit was filed against Google under BIPA for unlawfully collecting, storing, and using biometrics in the form of photos uploaded to Google Photos. Specifically, the lawsuits call out "face templates" that Google creates using sophisticated facial recognition technology that creates detailed geometric maps of the face. BIPA provides a comprehensive definition of "biometric identifier," in which it specifically says that photographs are not biometric identifiers. However, BIPA does say that a scan of hand or face geometry is a biometric identifier. Essentially, AI created sensitive data from non-sensitive information.

### To Presort with Labels or Without Labels Creates a Question of Privacy

Previously, we presented a quote from Eric Chien, and labels are adjacent to classifiers. A label is a way of saying that running algorithms in a data set should have natural parameters. Most companies will try to hedge bets by matching entities to predicted behaviors from like machines or peer groups; this being the concept of labeling. There is a reasonable argument that can be made in support of this approach. If you allow an AI/ML engine to run in a siloed environment, such as a microsegmentation group, there can be a high-fidelity result that is recursive to that group. There is also safety in using anonymized data. The problem, of course, is if a vendor is using truly anonymized data, it is difficult to know where to place labels.

It should be noted that network detection and response (NDR) vendors insist that their AI models have always run autonomously without labels. The legacy technology basis for Darktrace products is that each entity has a "pattern of life." The idea of a "pattern of life" is that all telemetry coming to and from an entity has value. Trying not to sound superfluous, the behaviors of an entity provide its own version of truth. Darktrace argues that there is too much to be learned in the sum of all sessions; a siloed approach limits the potential learning. Labels limit not only create a limit on what could be listed as anomalies but also restrict what can be established as good lists.

## *Input/Content Manipulation/Bias*

AI, and the output thereof, is a product of the input and the prompts that it is given. Prompts asking essentially the same question but with different levels of craftsmanship can generate vastly different responses. Further, the AI learns from everything that is input into it. The AI learns from the inputs regardless of whether they are truthful, biased, or nefarious. This includes sensitive and confidential information. AI models are trained to learn — not necessarily to forget. If sensitive or confidential data is fed to the model, it's out there and available until the model can be retrained.

Further, it is possible to brute force influence the response that the AI gives to others. In the early days of the internet, marketers were obsessed with search engine optimization. They were flooding the internet with content around their solution to move from page 12 to page 1 on the search results. There may be a resurgence of this tactic to ensure that ChatGPT or others suggest their solution over others.

Generative AI also shares all the problems of AI that have come before it. It is a window into the bias of the people who created it. Overwhelmingly, men who are software engineers are creating these tools.

Even OpenAI's CEO Sam Altman has admitted that ChatGPT struggles with bias. There are handfuls of examples of gender bias in responses that ChatGPT and other generative AI tools have given to users, not to mention the outright racism that has appeared in many different outputs of generative AI after a very short time accessing the open internet. This bias perpetuates itself and impacts the world around it as generative AI gains in popularity and importance.

### *Efficacy/Seed Set*

If the data used to train an AI model is not high quality, high volume, and mapped closely to the problem of interest, the model will likely be close to useless for the purpose of aiding your business. High-quality data is data that is error free and especially free from skewness. Skewed data, or data that leans heavily in a direction that is not reflective of a normal or normalized population, will hold little insight. Currently, available generative AI resources attack error and skewness through volume. Trained on a large corpus of text data, OpenAI's ChatGPT and Google's Bard are the products of over 175 billion parameters — and growing — at the time of this writing. While volume is critical to establishing a working AI model, volume also smooths data errors that can lead to skewness. Although perhaps not an immediate concern, especially for those merely exploring potential uses of generative AI, a close review of the data points used to inform an AI model will become more necessary as the AI and automation market develops. Standardized data collection does not exist for many industries, which makes the application or integration of AI algorithms more difficult. Organizations selecting an AI model should have a robust understanding of the data being created and collected within their own organizations as they evaluate AI and automation models for use in their companies.

## ADVICE FOR THE TECHNOLOGY BUYER

This document is the first in a series of documents that IDC looks to publish on security in AI. The objective has been to provide a primer for IDC's thinking on the topic. Future documents will be specific to vertical markets and technology types, providing very specific and relevant guidance. There is some general guidance that IDC would recommend that CISOs, CIOs, and security professionals consider before shifting to domain-specific treatments:

- **Focus on the benefit, no matter what AI is.** The buzzwords might help initially as you "tin cup" the C-suite to finding funding for security investment, perhaps there is no denying. Ultimately, though, the offering needs to deliver value in YOUR environment. You should force your vendor to demonstrate how the offering delivers fast time to true value. Too often, vendors offer "vanity metrics" about the features of the offering and push the onus for determining how the attributes of the offering create value in your environment onto you, the CISO.
- **Start with an outcome in mind.** The syndicated research community is as guilty as any when we insist on assigning categories to technologies, but consumers of security technologies have to keep their objectives that need to be accomplished:
  - Reduce incident triage time.
  - Cut false positives.
  - Reduce mean time to detect by 40%.

  Please remember. If the AI is to provide a benefit, it can be measured … objectively … with metrics. If the benefit cannot be objectively measured, you need to question if the "benefit" is real. In application, AI is typically blended in alongside five plus other techniques for detection in vendor stacks, broadening the great question of "how can I measure whether my security controls are effective?" AI is but one of many such layers (e.g., most vendors also use expert

rule engines in their EDR/XDR, "when I see X followed by Y"). Arguing about the additive effects of a single AI tool is counterproductive; measuring the effectiveness of the complete solution is the goal. Let the vendor worry about "how to cook the dinner." Hold the vendor accountable for the outcome.

- **Note that transparency is now more important than ever.** In the last couple of years, the various detection and response vendors have often produced threat or risk scores. Buyers often cast a disparaging eye at these scores, understanding that they were only relative to each other. Our gut feeling is that the idea of "black box" analytics is officially over in an AI-generated world. In addition, an important assurance a cybersecurity provider can make is to demonstrate how the data it collects from the buyer's environment is treated. Customers should expect vendors to provide detailed information on how and where system data is sent to the cloud. Likewise, customers should expect clear statements on how customer data is transmitted to cloud-based AI models, whether such data is used for model training, and so forth.
- **Create decision trees.** Even when network conditions are stable and the cybersecurity vendor feels like it creates fantastic risk scores or prioritizes vulnerabilities, it comes down to the analyst in the SOC. That analyst should have a list of predictable outcomes stemming from remediation actions taken.
- **Demand low code, no code, or natural language processing (NLP).** There are still sophisticated SOC teams that are happy to develop code, but largely, drag and drop is king.

This is the first document in a series on how IDC defines AI in cybersecurity. Subsequent documents will focus on putting AI in action, dotted with how cybersecurity vendors have created products leveraging AI. Planned topics include:

- How generative AI and ChatGPT specifically are used to optimize security operations center (SOC) protocols.
- How AI is evolving to fashion better discovery and facilitate automation in cybersecurity discovery and response. In association with this, IDC will take a more granular view of how AI will be used to improve endpoint and extended detection and responses, managed detection and response (MDR), security information and event management, and threat intelligence.
- The role of governments, industry regulators, and what cyberinsurers favor is being played out dynamically, but these silent hands will shape what is possible in cybersecurity software and services.
- Detection and response are arguably the juiciest of cybersecurity use cases, but AI is likely disruptive in how it can be used to ensure heterogeneity in on premises and public cloud environments, improve cybersecurity posture management, demonstrate compliance adherence, augment physical security, and close gaps between networking and cybersecurity.
- Last, IDC will consider the many points of conflict stemming from generative AI. Quickly identified conflicts include enhanced end-user digital experience versus individual privacy: responsibly deployed AI versus AI deployed without caretaking; and open XDR versus proprietary platforms … there are others.

## LEARN MORE

### Related Research

- *Generative AI Especially Benefits AI Infrastructure OEMs, ODMs, and Cloud Service Providers* (IDC #US50595223, May 2023)
- *Is Cybersecurity Spending Recession Proof?* (IDC #US50546623, April 2023)
- *ChatGPT: Protecting Your Organization from Unintended Risks of Generative AI* (IDC #US50583423, April 2023)
- *Engendering Trust with Proactive Cybersecurity Using Continuous Risk-Based Posture Assessment* (IDC #US50456123, March 2023)
- *Six Quick Thoughts on How ChatGPT Affects Developers, Development, and the Developer Experience* (IDC #US50448123, March 2023)
- *Worldwide Threat Intelligence Market Shares, 2022: Providing Insight on the Gathering Threats Outside the Network Perimeter* (IDC #US49128022, March 2023)

### Synopsis

This IDC Perspective discusses how IDC defines AI in cybersecurity. Combinations of artificial intelligence (AI) and machine learning (ML) have influenced the cybersecurity landscape for the better of 15 years. What computers have always been able to do is make correlations to the bytes, files, hashes, and code that comprise a network. However, for all of the improvements in computing and years of refining algorithms, so much of operating the network and the cybersecurity software that protects the network are still manually intensive processes.

After all of this time, recent developments in generative AI and, more specifically, ChatGPT are seemingly addressing the cybersecurity manpower gap. IDC calls it "autonomizing the SOC." The process of realizing a fully autonomous SOC involves several intermediary steps, but the new efficiencies in evidence are:

- The availability of enriched data at the time of the incident investigation
- The ability to generate an instantaneous response based on the type of attack is increasingly automated
- The implementation of analytics to discover unmanaged devices in the network
- The development of natural language processing (NLP) that enables threat hunting and security querying at the speed of speech

"AI is improving SOC processes and empowering security analysts; the power comes at a critical time as organizations struggle with hybrid, multicloud complexity and a chronic workforce shortage," said Frank Dickson, group vice president, IDC Security and Trust Division. "However, the hard work is not about the AI at all, but creating and enabling the security data foundations that will allow AI to create outcomes. Regardless, the cause for overall optimism is real."

## About IDC

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications and consumer technology markets. IDC helps IT professionals, business executives, and the investment community make fact-based decisions on technology purchases and business strategy. More than 1,100 IDC analysts provide global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries worldwide. For 50 years, IDC has provided strategic insights to help our clients achieve their key business objectives. IDC is a subsidiary of IDG, the world's leading technology media, research, and events company.

## Global Headquarters

140 Kendrick Street
Building B
Needham, MA 02494
USA
508.872.8200
Twitter: @IDC
blogs.idc.com
www.idc.com

---