mimecast

White Paper

AI and Cybersecurity:

The Promise and Truth of the AI Security Revolution

Overview

When it comes to artificial intelligence and machine learning, information technology professionals are inundated with hype; indeed, Gartner's Hype Cycle for Artificial Intelligence shows a large majority of AI capabilities still climbing toward the "Peak of Inflated Expectations."[1] Unfortunately, the cybersecurity industry finds itself in the same position. Security providers that depend primarily on AI/ML capabilities often pitch their solutions as panaceas, despite the many obstacles and challenges that remain on the path to achieving the high expectations that have been set for this technology.

That said, it would be foolhardy for cybersecurity professionals to disregard AI's potential. With the modern threat landscape becoming more complex daily and the application of AI growing more advanced, AI and its associated disciplines are fast becoming essential cybersecurity tools. And the need for AI in cybersecurity — or more particularly in the near term, machine learning (ML) — will only rise from here. So, it's crucial for organizations to combine AI/ML with more traditional cybersecurity approaches to keep risk from cyber threats as low as possible. But how is a security professional to tell reality from hype and make the sound choices needed to protect their organizations' communications, people, and data?

This paper aims to answer that question by explaining why AI/ML is crucial to cybersecurity, how it fits in, its best use cases, and where it goes from here.

This Paper will Explore:

- **The Landscape**
  How evolving business technologies and cybercriminals' approaches to breaching them make AI/ML a requisite for cybersecurity.
- **Benefits & Dangers**
  The potentials and risks that AI/ML brings to cybersecurity.
- **Best Practices**
  The best ways AI can enhance cybersecurity — given AI/ML's current state of development.
- **AI Cybersecurity Applications**
  Specific cases where AI/ML capabilities enhance enterprise cyber resilience solutions.
- **Looking Ahead**
  How will AI/ML capabilities evolve in cybersecurity?

1. Gartner Identifies Four Trends Driving Near-Term Artificial Intelligence Innovation, Gartner Inc.

The Landscape

Rising Security Pressure as Businesses Consolidate on Fewer Platforms

The reality cybersecurity professionals face is that everything is connected everywhere all at once. Hybrid work environments have proven difficult for organizations to roll back and created a permanent world of distributed employees, systems, devices, and data.

At the same time, changes to technologies and work patterns continue. Generative AI has exploded onto the scene, democratizing access to an incredibly powerful technology whose potential is being explored by everyone from non-technical employees to the most sophisticated threat actors.

Communications channels also continue to expand. Email remains the primary channel, but collaboration applications like Microsoft Teams and Slack have become an indispensable part of employees' daily routines. Tools like SharePoint and OneDrive are also deeply embedded. The result is an expanded attack surface that's more targeted than ever.

Compounding that risk is continued homogenization, with hundreds of millions of users around the world using the same tools. The potential payoff from accessing mass volumes of data has made providers like Microsoft and Google irresistible targets, and the impact on victims should not be underestimated. The average cost of a data breach rose in 2023 to $4.45 million, the highest level seen in the 19-year history of IBM's Cost of a Data Breach report.[2]

> "The consolidation of digital tools is one of the biggest factors influencing the threat landscape today. The mass move to platforms such as Microsoft 365 has provided threat actors with the focus to develop more sophisticated attacks, to the extent that some phishing-as-a-service groups only sell Microsoft-themed phish kits now. Simultaneously, the efficacy of human defenses has been degraded by the abuse of legitimate services and use of compromised accounts to launch attacks from platforms users know and trust."
>
> **Dr. Kiri Addison – Senior Manager, Product Management, Mimecast**

There is no way to avoid the threat. Organizations are in for a fight.
Mimecast's most recent State of Email Security Report reveals that 76% of companies are preparing for the fallout of an email-borne attack in the coming year and 97% have faced at least one phishing attack in the past 12 months.

2. Cost of a Data Breach 2023, IBM

The Benefits and Dangers
of AI in Cybersecurity

With attacks growing in speed, scale, and complexity every day, there is a myth emerging that AI is a panacea for keeping communications, people, and data secure. The hope that many have for AI is understandable — it represents a lifeline for IT and security teams that face limited resources, increasing complexity, and growing risk. But the truth is more subtle. There's no question that AI is essential to a modern cyber defense strategy; but like every new security innovation, it is just a tool — albeit an extremely powerful one.

Benefits

Despite the complex algorithms under the hood, the benefits of AI are straightforward and easy to articulate. AI:

- Can process huge data volumes — much more than humans could make sense of.
- Is fast — it processes information far faster than a human.
- Makes "smarter" decisions over time; it can learn, as long as there is a data science team in place monitoring and retraining the models as needed.
- Can simplify and/or automate some tasks

Applied to cybersecurity, where data volumes are large and a few extra minutes of response time can mean the difference between blocked attack and disastrous breach, those benefits can create tremendous enterprise value.

The impact of Generative AI

Generative AI is far from a new topic of discussion, but the introduction of Chat GPT in late 2022 unquestionably changed the conversation. With broad-based access to such a powerful tool available to anyone with an internet connection, experimentation exploded, along with the fears about what generative AI might ultimately be able to do.

The cybersecurity industry is no exception; and there is, unfortunately, little comfort to be taken from the idea that Chat GPT has safeguards built in. There are already variants for sale on the dark web that are fine-tuned to support malicious purposes.[3] While there are still many unknowns about how these technologies will be used, some early trends are taking shape.

Amongst malicious actors, one of the primary use cases appears to be creating more realistic phishing emails at scale – no more spelling or grammatical errors, no more typos and tell-tale stilted language.

Assessing how broadly generative AI is being used in this area isn't straightforward, as the jury is still out on whether humans or machines can consistently detect an AI-generated email, phishing or otherwise. However, it's clear that the number of phishing attacks is going up[4], with 98% being delivered through email.[5] It feels safe to assume that some portion of that increase can be credited to generative AI.

3. https://cybersecuritynews.com/black-hat-ai-tools-xxxgpt-and-wolf-gpt/ I 4. https://www.zscaler.com/blogs/security-research/2023-phishing-report-reveals-47-2-surge-phishing-attacks-last-year
5. Verizon Data Breach Report 2023

The good news is that AI-generated malicious emails are just that – malicious emails. The same principles and techniques for blocking them still apply. Beyond mass phishing email generation, the following are some areas of potential risk.

- Creating polymorphic malware
- Helping lower-skilled attackers execute more advanced attacks
- Reconnaissance on systems and people
- Identification of system vulnerabilities

For those working to keep attackers at bay, generative AI is also thought to hold great potential, with use cases like the ones below coming to the forefront:

- Researching new tactics, techniques, and procedures
- Suggesting remediation actions
- Summarizing security reports
- Deciphering code
- Creating payloads for pen testing
- Providing new tools that can augment an analyst's work, like Microsoft's Security CoPilot

It goes without saying that this is a space to watch.

A Note About AI vs. ML

Though lay people use the terms interchangeably, experts usually consider ML a subset of AI — and some define them as separate fields. Both process large volumes of information far faster than human brains ever could and react to that input in a way that goes beyond "if then" logic.

- AI systems make decisions based on circumstances, in many cases attempting to replicate or surpass the skill with which human intelligence would respond to the same tasks. Generative AI is technology capable of generating text, images, synthetic data, or other media, using generative models.

- ML is basically the fusion of computer science and statistics, focusing on large-scale data processing and analysis. ML systems use iterative statistical analysis to make decisions about how to approach and make sense of data, and what to do about it. They can improve themselves over time as they observe more data and results.

Challenges/Dangers

Effectively incorporating AI capabilities into an organization's cyber defenses requires being aware of, and working around, the challenges and dangers AI comes with. AI:

- Needs huge volumes of high-quality data to work well; many organizations just don't have enough, or good enough, data.
- Can generate too many false positives.
- Is hard to tune for the needs of individual customers, especially smaller organizations without extensive data.
- Can lack transparency in how it makes decisions — a serious challenge for enterprise security teams that need to understand these decisions in order to make their own subsequent choices.
- Is vulnerable to "poisoning" of data sets, leading to incorrectly trained models which will produce incorrect results.
- Relies on models that may degrade over time if they aren't properly maintained by data science experts to keep up with changes to the characteristics of the data being analyzed.
- Uses models that can be reverse-engineered and mimicked.
- Can be used by cybercriminals to craft more sophisticated attacks with less effort.

Behavior in a production environment also can't be known with certainty until it is deployed — a potentially dangerous situation depending on the application that requires extreme care and oversight.

Together, these benefits and challenges paint a complicated picture. By rapidly analyzing data to accelerate detection and response, AI algorithms can add a valuable layer of defense to existing security infrastructure, helping companies stave off malicious actors, secure their data, and stay compliant with regulations. But doing so requires organizations to implement AI-capable tools with sufficient rigor to counter the new challenges and dangers that AI introduces.

How can organizations enjoy the benefits of AI in cybersecurity while keeping these dangers at bay? That is the focus of the next section, Best Practices in AI Cybersecurity.

6. AI for Cybersecurity Shimmers with Promise, but Challenges Abound," Dark Reading

Best Practices

The key is to incorporate AI/ML capabilities into a multilayered defense strategy. That means deploying it in security solutions that can take advantage of AI's strengths, and then combining those with other security solutions to backstop against its weaknesses. The result should be a broad and layered cyber defense system that combines the latest in machine intelligence with the best of rules-based and other types of security controls, all arbitrated — when necessary — by the brains of human security operations center (SOC) analysts.

AI is adept at recognizing and neutralizing common threats at scale and can do so with greater accuracy than human beings. But to stop truly dangerous attacks, organizations need a comprehensive security architecture that deploys AI-powered filtering designed by data science experts who know how to navigate the gray area between clear threats on one hand and emails or links that are legitimate and crucial to business operations on the other. Since no detection model will be perfect, this should also be supported by feedback loops to help quickly identify where machine learning models are not quite hitting the mark.

In practice, this means deploying AI/ML first where a lot of data exists. AI was first used in cybersecurity to identify anomalies in user behavior or to detect network traffic that may suggest an intrusion. Large networks generate tremendous volumes of data, as do individual users' email exchanges, which can number in the hundreds per day. This is the underlying principle of Mimecast's social graphing capability where anomalous exchanges between users are flagged to recipients to alert them to potential threats.

> "In the realm of AI and machine learning, the profound synergy of human expertise and algorithmic prowess propels innovation to new heights. The symbiotic 'human-in-the-loop' support is not just a framework; it's the linchpin that delivers unparalleled precision and adaptability. As algorithms navigate the data landscape, human insight brings contextual understanding, ethical discernment, and a nuanced touch that algorithms alone cannot emulate. This dynamic collaboration crafts a future where the convergence of artificial intelligence and human intuition becomes the catalyst for groundbreaking advancements, ensuring a harmony that not only surpasses automated capabilities but also resonates with the essence of our shared human experience."
>
> **Vedant Ruparelia - Senior Applied Machine Learning Scientist, Mimecast**

> "Artificial intelligence and machine learning technologies aren't inherently superior. The effectiveness of machine learning relies heavily on the quality of data it's trained on. The principle of 'garbage in, garbage out' holds true in this context: poor data can lead to flawed machine learning outcomes. Moreover, the human intelligence guiding its development plays a crucial role. Incorrect decisions during this phase can result in biased or inaccurate results. It's surprisingly easy to fall into these pitfalls in machine learning."
>
> **Navya Vats - Senior Data Scientist, Mimecast**

AI/ML should also be exploited where complexity is present, but speed is crucial. This can mean, for example, rapidly deciding whether the URL on which a user just clicked is safe or malicious, or whether the email containing sensitive information that a user just sent is safe to let through. Thinking more broadly, another compelling area where speed is necessary is in proactive threat hunting. Rather than responding to every potential attack the same way, security analysts can combine AI/ML algorithms with proven cybersecurity tools as part of a multilayered system to detect and respond to threats in the most effective and efficient way possible, on a case-by-case basis. Not only does this make them more performant, but it also reduces the time and cost drain of managing false positives, giving teams the confidence to separate genuine attacks from benign activity.

Finally, that last point suggests an important way in which all these AI/ML cybersecurity applications can create value for organizations: Addressing the acute skills shortage in cybersecurity by optimizing the work of existing SOC analysts and helping to raise their productivity.

Best Ways to Use AI/ML in Cybersecurity

- As part of a multilayered defense strategy.
- When it provides a measurable advantage over other cybersecurity technologies, such as identifying and responding to common threats at scale.
- Where data volumes are high, so algorithms can easily learn the organization's "standards" and, therefore, make anomalies stand out.
- When speed is of the essence, such as scanning URLs between the time a user clicks and the page renders or detecting intrusions before bad actors can exfiltrate data.
- To assist human SOC analysts and raise their productivity, reducing pressure from the cybersecurity skills shortage.

While AI algorithms allow quick assessment of thousands of emails and URLs each day, the true test of cyber defenses comes down to making the right choice when things fall into the gray area between a clear threat and a benign deviation from the norm.

AI Cybersecurity Applications

The Mimecast Approach: Multilayered, AI-Powered Security

Cybersecurity has always been about making high-stakes decisions. What should be let through? What should be blocked? What risks should be taken? AI/ML doesn't change these questions, but used wisely, it can help answer them faster, at scale, and with greater efficiency.

Mimecast — a leader in email security for 20 years — has always been at the forefront when it comes to new technologies and strategies for defending against a relentless set of adversaries. That includes AI, which we incorporate across every layer of our solutions, wherever the technology is the right choice to help us maximize our customers' defenses, neutralize more threats, and take pressure off their security teams. But AI is not a panacea. Our detection stack applies the right inspections at the right time, with ML algorithms working alongside proven technologies that we have continuously improved over the course of nearly 20 years.
We combine dozens of different approaches, augmented by AI, to yield the industry-leading security efficacy for which Mimecast is known.

> "AI and machine learning techniques are not magical black boxes that deliver security out the other side. Ultimately, they are tools to help us solve problems. The key to their safe and effective use in cybersecurity is the correct application of these tools informed by knowledgeable individuals and good quality data."
>
> **Robin Moore – Principal Product Manager for AI and Machine Learning, Mimecast**

> "Large language models and natural language processing techniques have evolved greatly in the last year, and this has supported the creation of more sophisticated attacks by cybercriminals. AI is needed to effectively counter and stay ahead of threat actors who would misuse these technologies."
>
> **Guhan Sukumaran – Senior Manager, Data Science and Machine Learning, Mimecast**

An Illustrative Case: Malicious URL Detection

Mimecast URL detection is a perfect illustration of the company's AI/ML philosophy. Mimecast's URL detection capability performs a single crucial task — identifying malicious URLs — but it is not a single function or product. It combines dozens of scanners working together to detect high-risk URLs as effectively and efficiently as possible.

Some of these are simple lookups that catch simple threats. Others are rules-based algorithms that catch more complex but common attacks. ML algorithms kick in when needed. Crucially, they render rapid decisions in cases where other technologies cannot give a deterministic value about whether, based on what is known so far, the risk is low enough to render the page or whether a next-level scan is needed.

Why doesn't Mimecast use ML algorithms, solely? Ultimately, the power of the solution is not in any one discrete function. It's the combination of all those functions, whether ML or not.

When Mimecast does decide to deploy AI/ML in a product, it does so with a distinct advantage: The more data algorithms have to train upon, the better they learn. Mimecast protects more than 42,000 customers globally and inspects 1.7 billion emails every day — all of which fuels machine intelligence that helps deliver the world-class email security efficacy needed to keep our customers secure. Mimecast's data science team uses that data to build AI/ML models into Mimecast solutions, then continually monitors the efficacy of those models in the context of the ever-evolving threat landscape and decides when to retrain our models so they maintain peak performance.

Selected Mimecast AI/ML Functions

Here is a rundown of just some of the ways in which AI/ML algorithms empower Mimecast's industry-leading security efficacy.

Limiting the intelligence attackers can gather and empowering employees with information at the point of risk

Even a business's savviest users can get tricked by a malicious email, especially with attackers using advanced technologies to gather information about employees and target them with convincing spear-phishing attacks. Mimecast uses AI/ML algorithms to limit attackers' information gathering capabilities, effectively detect targeted email threats, and empower users with information. And this solution keeps learning and becoming more effective with each thwarted attack.

How it works

- Detects and disarms trackers embedded in employees' emails, stopping inadvertent disclosure of information to attackers.
- Uses Social Graphing technology, powered by machine learning, to map communications and build social graphs that serve as a benchmark for detecting anomalous behaviors.
- When warranted, decides to add contextual warning banners to emails to alert employees. Banners are updated in real-time across devices when risk levels change.

Stopping outbound emails and sensitive data from falling into the wrong hands

Attackers aren't the only threat to email data. Employees can be a danger to themselves and the business when they send an email to the wrong address, whether intentionally or not. At best, this is an embarrassing mishap. But at worst, it can leak sensitive data, leading to dire consequences, including fines, reputational damage, and compliance violations. Mimecast Misaddressed Email Protection uses AI/ML to track users' communications, identify anomalies, and alert employees if they are about to send an email to a new or unrecognized address.

How it works

- Detects anomalous communications activity by using Social Graphing, powered by machine learning, to "understand" a user's typical patterns.
- Holds misaddressed emails at the gateway and alerts the sender of the potential problem.
- Gives the sender a second chance to confirm their email is going to the right place, helping to ensure email data is handled judiciously and in a compliant way.

Catching malicious emails disguised as legitimate messages from credible sources

Credential harvesting is on the rise, especially as more businesses use file-sharing services like OneDrive and SharePoint to collaborate remotely. Attackers reference these sites in malicious emails to bypass detection, luring victims to URLs where they may unwittingly share their business logins.

Mimecast's credential harvesting protection uses machine learning and advanced computer vision to check whether a URL is legitimate, with analyses so precise that they can notice if even a single pixel is off on a seemingly safe webpage.

How it works

- Computer vision algorithms detect anomalies in the branding, login, or payment form information that appears on screen.
- Depending on the level of suspicion and associated risk, users are either warned or blocked from the page. The ML algorithm learns from each analysis to become more precise with time – and far harder to fool with a sophisticated brand impersonation than the average user.

Optimizing Threat Detection

How it works

Mimecast's detection technologies are powered by detection engines that are optimized with the help of machine learning algorithms. These algorithms' decisions are analyzed and adjusted by Mimecast's SOC analysts and data science team, which continuously builds, monitors, and iteratively improves the models. Together these collaborative interactions provide a layered email security approach that captures the domain knowledge of our security experts, teaches ML algorithms with that knowledge, and applies those learnings to enhance protections for all customers.

Categorizing and triaging suspicious emails

How it works

When users report suspicious emails to Mimecast's SOC, Mimecast uses ML to automatically categorize, triage, and prioritize them for investigation. Each suspicious email's metadata is also automatically enriched with a risk score and any information about the user's past reporting behavior. This information helps analysts make a rapid decision about whether an email is malicious or not. The final decisions made by our analysts feed back into our ML models.

Categorizing websites

How it works

Supervised learning categorizes websites as malicious or inappropriate, which tells Mimecast's email and web security products to block access to those sites.

Identifying "not safe for work" images

How it works

Deep learning and computer vision algorithms work in concert to detect inappropriate images in emails.

Looking Ahead

How will AI capabilities evolve in cybersecurity?

Science fiction is rife with fully automated cyber battles between good and evil artificial intelligences. In fact, the author who coined the word "cyberspace" (William Gibson) used just such a plotline in the landmark trilogy of novels that brought cyberspace into mainstream language, beginning in 1984 with Neuromancer. The general industry sentiment is that such a scenario is still many years into the future; but as recent developments have shown, the AI tide may be turning much more quickly than many anticipated.

However, AI's importance should not be under or overstated. Today, AI/ML is a sophisticated and critical part of the multilayered cybersecurity solutions that organizations need to protect their communications, people, and data. Its deployment must be considered carefully and should be focused in areas that exploit its strengths while deemphasizing its limitations.

There's no question that the use of AI in cybersecurity will expand; and at present, there seem to be as many questions as answers. While not yet fully understood or realized, AI's potential is as vast as the imaginations and skill sets of those who are working to harness it. Only time will tell where and how far it will take us.

WORK PROTECTED.™

Advanced Email & Collaboration Security

mimecast®