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 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) and natural language processing (NLP) — will only rise from here. So, it's crucial for organizations to combine AI 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 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 a requisite for cybersecurity.
- **Benefits & Dangers**
  The potentials and risks that AI brings to cybersecurity.
- **AI Cybersecurity Applications**
  Specific cases where AI capabilities enhance enterprise cyber resilience solution.
- **Best Practices**
  The best ways AI can enhance cybersecurity — given AI's current state of development.
- **Looking Ahead**
  How will AI 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.

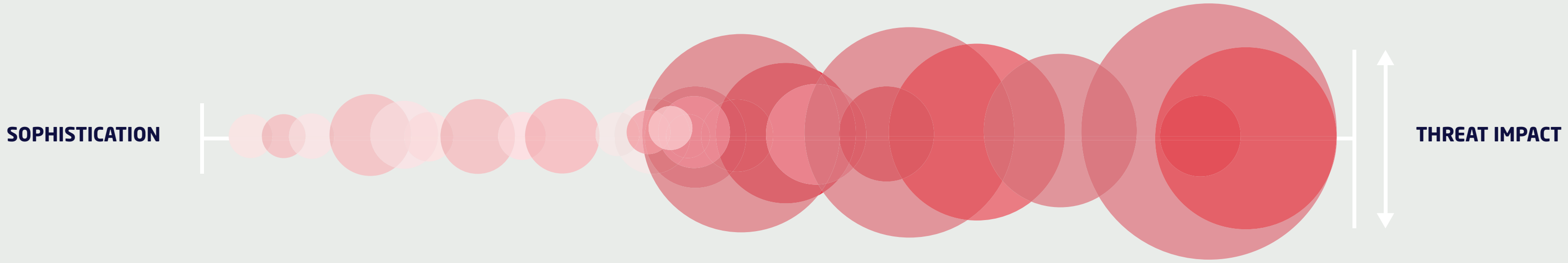
Communications channels also continue to expand. Email remains the primary channel, but collaboration applications like Microsoft Teams and Slack have become an indispensable part of employees' daily routines. Tools like SharePoint and OneDrive are also deeply embedded. The result is an expanded attack 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 vast amounts of data — much more than humans could make sense of.
- Is fast — it processes information far faster than any 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 plus support for multi-languages to attack regions not previously accessible.

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

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 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 AI 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 perfectly illustrates the company's AI philosophy. Mimecast's URL detection capability performs a single crucial task — identifying malicious URLs — but it is certainly not a single function or product. It combines dozens of scanning layers 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. AI algorithms kick in when needed; using AI to 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 AI algorithms, solely? Ultimately, the power of the solution is not in any one discrete function. It's the combination of all those functions, whether AI or not.

When Mimecast does decide to deploy AI in a product, it does so with a distinct advantage: The more training data algorithms are based 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 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 Functions

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

Defending against business email compromise and empowering employees with information at the point of risk

Even a business's savviest users can be 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 algorithms and NLP to effectively detect targeted email threats, empower users with information and limit attackers' information gathering capabilities. As with most of Mimecast's AI, this solution continuously learns and it's effectiveness improves with each thwarted attack.

How it works

- Uses Social Graphing technology, powered by machine learning, to map communications and build understanding typical communication patterns that serve as a benchmark for detecting anomalous behaviors.
- Text extracted from the email body is analyzed by NLP to determine the severity based on risky outcome categories, message characteristics or rules. Messages can have different policy levels applied to either reject or allow for administrative review.
- 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.
- Detects and disarms trackers embedded in employees' emails, stopping inadvertent disclosure of information to attackers.

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

Attackers aren't the only threat to your organization. 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 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 Microsoft 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 AI 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.

Categorizing and triaging suspicious emails

How it works

When users report suspicious emails to Mimecast's SOC, Mimecast uses AI 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 AI models.

Categorizing websites

How it works

Supervised learning categorizes websites as malicious or inappropriate, which tells Mimecast's inspection engines to block access to those sites.

Identifying "not safe for work" images

How it works

Not every image received via email is inappropriate – the challenge is how to identify the good from 'not safe for work'? Deep learning and computer vision algorithms work in concert to detect inappropriate images in emails, focusing on pornography, helping to maintain a safe and professional work enviornment. Being able to control this type of content whether it be received or sent is critical to brand reputation as an employee's email address by extension represents the organization.

QR Code detection

How it works

QR codes have become a common tool for sharing information quickly and conveniently, however, they can also pose a security risk. QR codes can contain links to malware, inappropriate sites, or other harmful content. Mimecast can not only detect QR codes through our deep learning and computer vision algorithms, but the link residing behind the QR code is resolved and passed to Mimecast's URL detection capability to identify high risk URLs. As described in the illustrative URL example, a combination of technologies including machine learning, is utilized to determine if the QR code links are malicious which results in the rejection of the email, helping to protect against this style of phishing attacks.

Malware and zero-day protection

How it works

Mimecast leverages AI within our inspection engines and offers protection from even previously unknown threats – APTs, zero-day attacks and ransomware. The machine learning algorithms incorporated in the various file inspection capabilities, extract features from existing malware samples or families enabling them to predict future malware based on shared similar features. Mimecast's sandbox uses machine learning and behavior detection technologies, which ensures that only files that require further analysis are sent for inspection improving analysis response times. Files sent to the sandbox are analyzed by advanced machine learning algorithms in addition to decoys, anti-evasion techniques, anti-exploit, and aggressive behavior analysis resulting in efficient malware detection. Malware detection technology that incorporates machine learning algorithms are more effective than signature-based systems, because of the enhanced detection rates for new malware variants.

Best Practices

The key is to incorporate AI capabilities into a multi-layered 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 first where a lot of data exists. For example, AI was first used in cybersecurity to identify anomalies in user behavior or to detect network traffic, as you have access to a large data set, that may suggest an intrusion.

> " 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**

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 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®