Abnormal

AUTHOR: FC, Ethical Hacker

The Rise, Use, and Future of Malicious AI:

A Hacker's Insight

Introduction

As with any technological advancement, AI has become a double-edged sword.

This whitepaper aims to shed light on the growing trend of artificial intelligence (AI) usage by cyber threat actors, examining how AI technologies are being exploited for malicious purposes. It seeks to provide insights into the methods employed by cybercriminals, the impacts on cybersecurity, and strategies to mitigate these threats.

As with any technological advancement, AI has become a double-edged sword. While AI offers tremendous benefits in various sectors, its potential for misuse poses significant risks. Understanding and addressing the malicious use of AI is crucial to maintaining robust cybersecurity defenses.

This paper will cover the evolution of AI in cybersecurity, the emergence of malicious uses of AI, specific case studies, the impact on the cybersecurity landscape, current defense mechanisms, and recommendations for mitigating risks.

.

Table of Contents

Understanding AI in Cybersecurity

AI has been part of scientific and technological study for decades, as well as a theme of fascination in popular culture. While the latter arguably helps generate public interest and awareness, it also causes confusion by what we actually mean by AI.

At its core, AI refers to the simulation of human intelligence processes by machines—particularly computer systems. These processes include:

- Learning, where systems acquire information and rules for using it effectively.
- Reasoning, where systems apply rules to reach approximate or definite conclusions.
- Self-correction, where systems improve over time by refining their approach based on new information and experiences.

AI encompasses various technologies—including machine learning and natural language processing—each contributing to the overall goal of creating intelligent systems that can perform tasks that have historically required human intelligence.

But like any tool, even ones that are built for the good of the world, AI can be used by bad people to do bad things.

But like any tool, even ones that are built for the good of the world, AI can be used by bad people to do bad things. Historically, technological advancements have often been mixed blessings. For instance, while the internet has revolutionized communication and access to information, it has also facilitated cybercrime and the spread of dis- and misinformation.

Similarly, AI has vast potential for positive impact, such as improving healthcare, enhancing education, and optimizing logistics. Unfortunately, it also presents new opportunities for malicious use. One of the first ways we're already seeing this is in cybercriminals exploiting AI to develop more sophisticated malware, conduct large-scale phishing attacks, and manipulate public opinion through deepfakes and automated disinformation campaigns.

The Evolution of AI in Cybersecurity

Every day, new tools for defense and offense are created and utilized on each side of the cybersecurity realm. On the defense side, AI-powered systems can automate routine security tasks—such as monitoring network traffic, analyzing logs, and detecting malware. These tools can adapt to evolving threats by learning from new data, thus improving their effectiveness over time.

Offensive tools, however, are also becoming more advanced, with cybercriminals attempting to employ AI to develop sophisticated malware, phishing schemes, and even automated attack vectors that can bypass conventional security measures.

Technology transfer and knowledge spillover have always been at the forefront of both attack and defense, which makes cybersecurity challenging, exciting, fast-moving and frustrating—all at the same time. Innovations in one area can quickly influence the other, leading to rapid advancements in techniques and tools. For example, a breakthrough in AI for threat detection can be swiftly adapted to enhance offensive strategies, and vice versa. While some tools may fall out of favor quickly due to advancements in countermeasures, others are rapidly adopted and become standard in both cybersecurity and cyber attack arsenals.

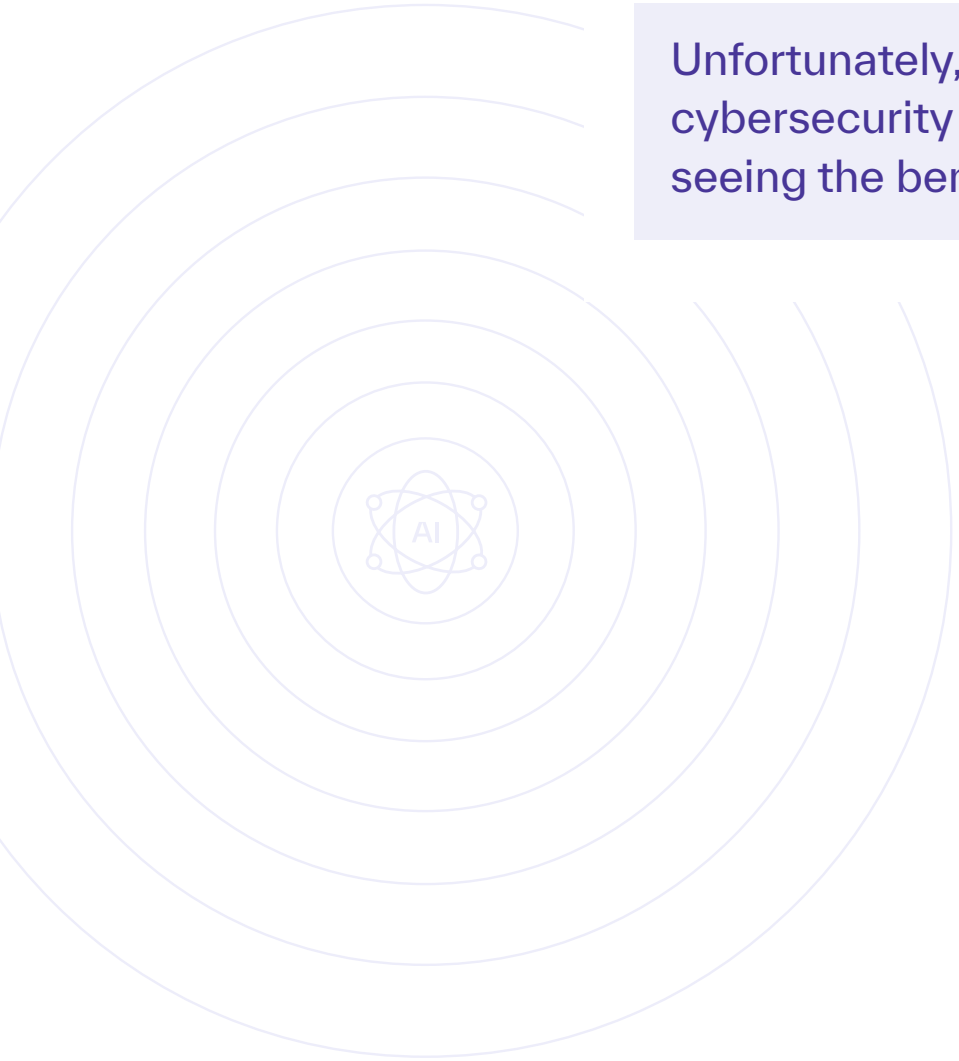
Because of its unique ability to harvest and learn from vast datasets and spot patterns in data or behavior quicker than a human could, AI can play a critical role in identifying unknown threats, reducing response times, and automating routine security tasks. This is vital in cybersecurity, where the volume of data and the speed at which threats evolve can overwhelm traditional defense mechanisms. AI systems can process and analyze data at scale, identifying subtle indicators of compromise that might be missed by human analysts. By continuously learning from new data, AI systems can adapt to emerging threats, ensuring that defenses remain robust against the latest attack vectors

Because of its unique ability to harvest and learn from vast datasets and spot patterns in data or behavior quicker than a human could, AI can play a critical role in identifying unknown threats, reducing response times, and automating routine security tasks.

As such, AI can be incredibly effective at detecting anomalies and pattern matching in areas such as email security and network traffic analysis. Historically, it has taken a skilled and experienced analyst to look and spot such patterns of behavior, but it can now be done in seconds with AI-based tooling. The significant reduction in time it takes to respond to a potential security incident can minimize the damage caused by breaches, whilst also reducing the workload of skilled workers to focus on more serious matters.

The more that AI can be leveraged to take over mundane tasks in order to identify low level threats, the better for the organization. This allows for more strategic allocation of human resources and as a result, the organization can maintain a stronger defense against cyber threats whilst optimizing their operational efficiency.

Unfortunately, it is not only cybersecurity professionals seeing the benefits of AI.



The Emergence of Malicious AI

It didn't take long for the criminals to start using AI to make money and cause mayhem. The rise of malicious uses of AI refers to the use of artificial intelligence technologies by threat actors to conduct cyberattacks, enhance the effectiveness of their methods, and evade detection. This includes AI-driven malware, automated phishing, and the creation of deepfakes. We will delve into this subject later in this paper.

The use of AI for malicious purposes has evolved alongside advancements in AI technology. Early instances involved simple automation, but today's threat actors leverage sophisticated AI tools to create more complex and harder-to-detect attacks. While it is important to resist AI hyperbole, we also need to acknowledge that it is actively being used by criminals.

The increasing accessibility of AI technologies and frameworks combined with the knowledge transfer has led to an explosion of malicious tools and AI models. While commercial AI tools available to the public (like ChatGPT and CoPilot) have built-in safety systems and controls in place, cybercriminals are now creating their own versions, such as FraudGPT, PoisonGPT, VirusGPT and EvilGPT—each name inspired by their niche intended use.

As the dark web becomes flooded with new malicious tools and open source AI models are being de-censored, criminals can utilize them for malicious activities. The backbone behind any AI model is the dataset, and whilst commercial ones do not allow you to ingest your own data into them, the criminalized versions do. This makes them not only more capable of creating attacks, but also in defining their target data.

Methods and Techniques of Malicious AI

We have recently seen a huge uptick in the progress of deepfake technology, which uses AI to create convincing fake videos or audio recordings.

Gone are the days of disowned Nigerian Princes trying to offload their wealth to you. While they worked for awhile, those old-school phishing emails were so identifiable due to lazy pretexting and poor spelling and grammar.

Now, though, AI can be used to craft convincing phishing emails by analyzing and mimicking communication patterns. It can also target individuals more accurately by using data-driven insights. AI can be used to translate phishing communications into different languages, while adding urgency and persuasion at the click of a button. This is by far the biggest area of growth and what has been driving the mass media frenzy of headlines about malicious uses of AI.

We have recently seen a huge uptick in the progress of deepfake technology, which uses AI to create convincing fake videos or audio recordings. When done correctly, it can be used to impersonate individuals, spread disinformation, and facilitate social engineering attacks. With recent advancements, it is now possible to fully clone a face from a single image and create a realistic deepfake video within minutes—and just 20 seconds of audio will enable a very good clone of a voice. Realtime deepfakes are now being used in Zoom calls and other virtual meetings, making social engineering even harder to spot.

.

The Future is Now: AI-Powered Attack Types

We're no longer talking about science fiction; AI is here to stay.

Here are a few example attacks that have made recent headlines.

01

AI-Enhanced Malware

In 1980, the first malware that replicated itself and moved through a network was released. Often called worms, this malware was known as the Morris Worm after its creator Robert Morris. Since that time, malware has quickly advanced and grown. It therefore was an inevitability that the rise of AI would usher in AI-enhanced malware. In March 2024, we saw the release of what has been dubbed Morris Worm 2.0.

Public generative AI tools are a powerhouse to users who may wish to create code without learning the intricacies of coding.

The Morris Worm 2.0 targets generative AI systems. Like traditional worms, it steals data and deploys malware, but it does so by manipulating the AI prompts to bypass security measures and replicating itself across different platforms. It was created by researchers as a proof of concept, but the creation of this tool is an example of the kind of advanced malware research and development that we are likely to see—both by criminals and legitimate security researchers. Within the next few years malware will have advanced techniques built in that will allow it to recognize the system it is in and morph itself to defend against, or even avoid, current detection systems.

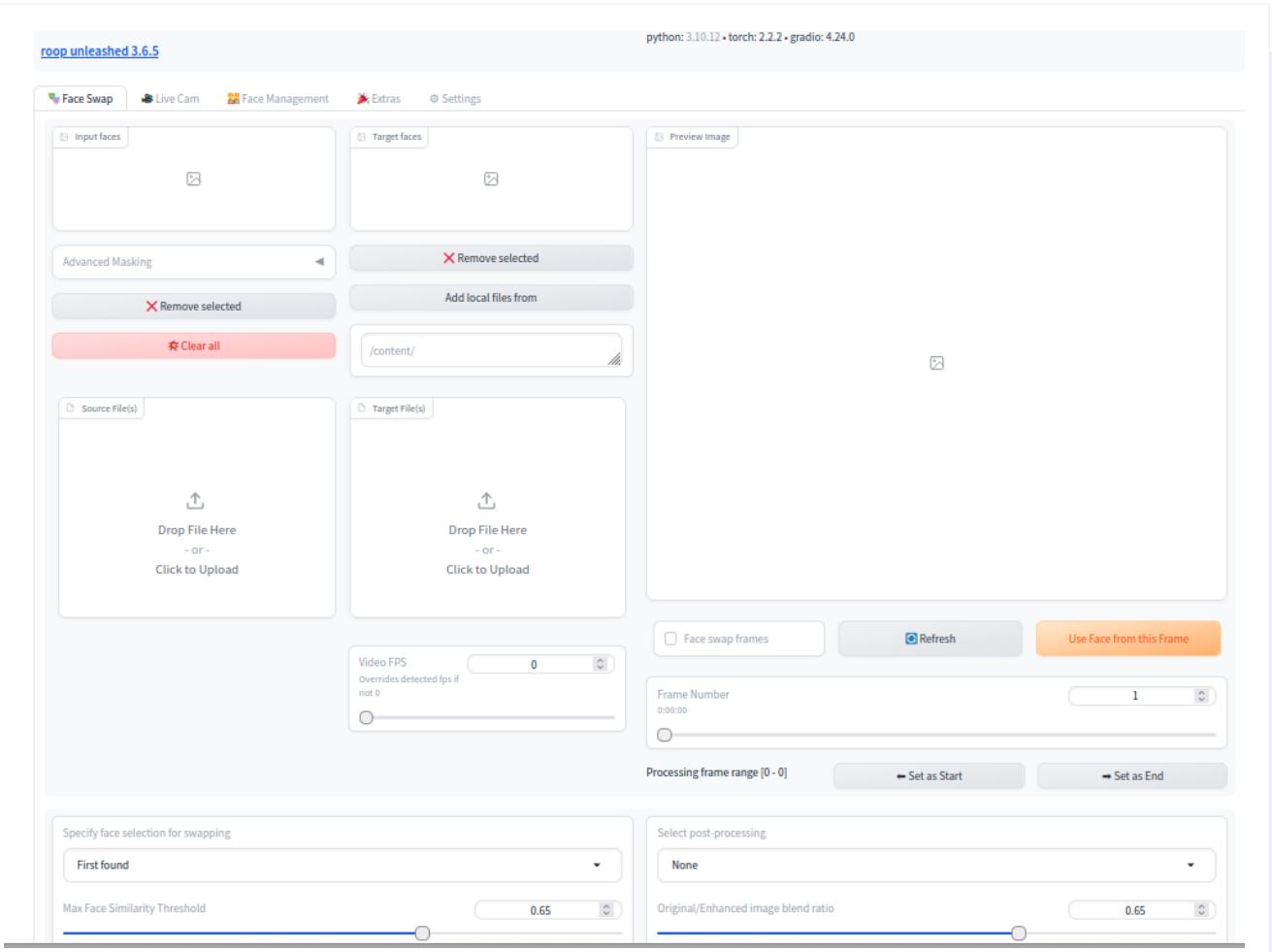
Whilst the use of Morris Worm 2.0 is just a proof of concept used in a lab environment, public generative AI tools are a powerhouse to users who may wish to create code without learning the intricacies of coding. Thankfully, safeguards are put in place in public systems to prevent malicious code being generated, but criminals are finding ways around those safeguards to help them code malicious software quicker and easier.

02

Deepfake Tools

In addition to malware, the introduction of generative AI tools has led to a substantial rise of impersonation attacks. These tools have enabled criminals to use digital twins and face swapping technologies, adding far more sophistication to their more traditional scamming techniques.

One such tool that has given this uncensored power to the masses is Roop-Unleashed. Based on the popular, safer tool Roop, the unleashed version has no safety guards in place and can be used on both images and video to perform face replacements easily.



Here you can see the interface for Roop-Unleashed running on a laptop. The user needs no knowledge of the science behind the face swapping, instead simply needing to upload single or multiple source images and a target image or video. Within minutes, even a standard laptop can create incredibly realistic images without censorship.

The release of similar tools that are even easier to use has resulted in an explosion of deepfakes for criminal use. In February 2024, a finance worker in Hong Kong was tricked into paying $26 million to fraudsters posing as the multinational firm's chief financial officer. Meanwhile, they're also being used to further political agendas, with deepfakes being used to imitate US President Joe Biden, Florida Governor Ron DeSantis, and other political officials.

Whilst criminals are using deepfake faces, they are also generating fake audio to create realistic fake kidnap ransom demands. Many parents have received disturbing phone calls of their children's voices begging for their lives and demanding money be paid. In reality the child is fine, and audio has been harvested from online social media accounts in order to create the clone. But for a scared parent who only wants their child returned safely, this can result in thousands (or even millions) of dollars lost.

In February 2024, a finance worker in Hong Kong was tricked into paying

$26 million

to fraudsters posing as the multinational firm's chief financial officer.

03

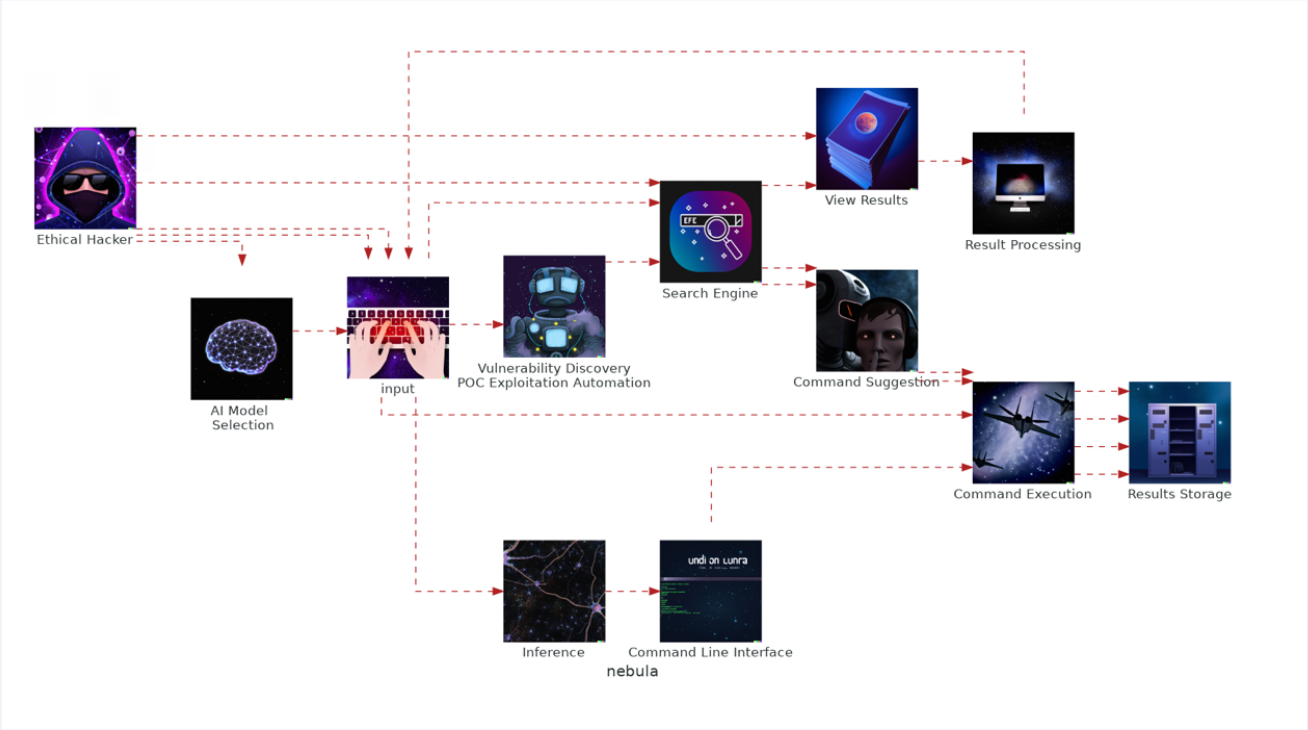
AI-Powered Hacking Tools

Successful hacking requires a complex series of events, and the average criminal is often not a professional hacker. Instead, they typically use playbooks of commands to run; if things go wrong, they tend to move on to the next target as they are unable to continue. This is why we have seen a rise in cybercrime-as-a-service or CaaS, with groups that sell ready-to-use access to systems—eliminating the hard work and increasing chances of success for those who buy their programs. These CaaS entities are themselves often buying access from professional hackers who do not have the drive, risk appetite, or intention of causing harm to their targets. It's win-win-win for all three sides—and a loss for the victims.

With the rise of AI-enhanced hacking tools, it has become even easier for lesser skilled criminals to start dabbling into the hacking side themselves. As with all tools, AI systems developed legally and legitimately can unfortunately be subverted by criminals for malicious purposes—through no fault of the developers.

Take, for example, the AI-powered tool Nebula by BerylliumSec, which is effectively an assistant for hackers. Originally created to aid ethical hackers trying to do good, tools such as this can unfortunately be used by criminals as well.

Rather than a hacker (either legal or criminal) having to learn complex and disjointed commands they can interact with the computer using natural language.



As you can see in this image provided by the Nebula creators, the platform does the heavy lifting of commands and execution. The key here is that the user types what they want to do in a natural language and the tool infers how to perform those actions.

Fortunately, many of these tools still require a level of technical skill and knowledge, rendering them out of reach for many petty criminals. However, with the pace of advancement in AI-powered tools, it will not be long before the barrier to entry will be lowered to the level of common criminals.

Impact of AI on the Cybersecurity Landscape

Organizations should invest in robust AI-based security solutions, continuously monitor and leverage threat intelligence, and conduct regular training and awareness programs for employees.

AI has had a significant impact in enhancing the sophistication of cyber attacks, enabling them to adapt and evolve in ways that make them increasingly difficult to predict and counter. Unfortunately, traditional security measures often fall short in detecting and responding to these AI-driven threats—highlighting the need for AI-enhanced defenses.

The economic and social impacts of such attacks can be profound, affecting individual organizations as well as the broader economy and society. To ensure we keep pace with the criminals, it is important that we leverage AI for defense. Integrating AI with existing security frameworks can lower the burden on organizations. Further, if we automate responses, while also collaborating and sharing amongst our organizations and cybersecurity professionals, we have a chance of staying ahead of cybercriminals.

Organizations should invest in robust AI-based security solutions, continuously monitor and leverage threat intelligence, and conduct regular training and awareness programs for employees. Additionally, governments and regulatory bodies should work together with cybersecurity professionals to establish guidelines to address the use of AI in cyberattacks and promote AI for defense.

The Role of AI in Future Cybersecurity

Cybersecurity has always been an arms race, a cat and mouse game in the digital realm waged by people and technology on both sides. As the future speeds forward, AI will play an increasingly critical role in both offensive and defensive cybersecurity measures, highlighting the need for balanced and ethical AI development.

As a result, AI-based tools will become so commonplace that—much like the car, TV, Internet and cell phones—we will soon forget how we lived without them. But we must not forget that cybercriminals will always perform the least amount of effort to obtain their goals. You are (for now, at least) much more likely to be hit by an AI-generated email with a phishing link embedded than a sophisticated AI-based attack system that is playing 4-dimensional chess with your GSOC. For this reason, it is imperative that we use AI to defend the front door while we continue to secure the perimeter and block access from every angle.

Conclusion

The rise of the malicious use of AI presents significant challenges to cybersecurity. Understanding the methods, impacts, and defense mechanisms is crucial for mitigating these threats.

To protect themselves and their employees, organizations must proactively adopt AI-driven defense measures, collaborate on threat intelligence, and continuously educate their workforce to stay ahead of the malicious use of AI. In addition, because AI has become so prolific, governments and industry leaders must attempt to regulate this new technology by developing robust ethical guidelines and regulatory frameworks to mitigate its risks—while maximizing its benefits.

It must be remembered that whilst the rise of AI is a force multiplier for threat actors, it is also a force multiplier for those defending against those attacks. Together, with the right tools, we can stay safe from these attacks.

About the Author

FC is the author of the bestselling book How I Rob Banks, and advises organizations on how to create secure, safe environments. As an ethical hacker for the last three decades, FC has helped thousands of banks, governments and other organizations advance their security.

He has shared his expertise in mainstream media, including the BBC and ITV, as well as popular industry podcasts such as Darknet Diaries. He has also been featured in printed media around the world, educating people about cybersecurity from a hacker's perspective.

His time as the Head of Offensive Research at Raytheon enables him to bring to bear his knowledge of how nation states and the intelligence community work with cyber weapons and how to defend against them. The decades he has spent legally breaking into organizations, both physically and digitally, has taken him around the globe in the fight against cybercrime.

Whilst FC is his memorable real name (it really is what's on his passport!), he is often better known by his hacker handle 'Freakyclown'. The name that was given to him by bullies as a schoolboy has now been turned into a positive alter ego for FC to educate the world on cybersecurity issues.

Born in the United Kingdom, FC and his wife Dr. Jessica Barker now live in Las Vegas, USA with their cat Bubble.

Abnormal

Abnormal Security provides the leading behavioral AI-based email security platform that leverages machine learning to stop sophisticated inbound email attacks and dangerous email platform attacks that evade traditional solutions. The anomaly detection engine leverages identity and context to analyze the risk of every cloud email event, preventing inbound email attacks, detecting compromised accounts, and remediating emails and messages—all while providing visibility into configuration drifts across your environment. You can deploy Abnormal in minutes with an API integration for Microsoft 365 or Google Workspace and experience the full value of the platform instantly, with additional protection available for Slack, Teams, and Zoom.

Interested in Stopping AI-Generated Email Attacks?

Request a Demo →

See Your ROI →

Appendices

Glossary of Terms

AI (Artificial Intelligence): The simulation of human intelligence processes by machines, particularly computer systems. These processes include learning, reasoning, and self-correction.

Anomaly Detection: The identification of unusual patterns that do not conform to expected behavior, often used in the context of detecting fraudulent behavior activity or network intrusions.

Botnet: A network of private computers infected with malicious software and controlled as a group without the owners' knowledge. Botnets are often used to send spam or to conduct distributed denial of service (DDoS) attacks.

Cybersecurity Framework: A set of standards, guidelines, and best practices to manage cybersecurity-related risk. Examples include the NIST Cybersecurity Framework and ISO/IEC 27001.

Deepfake: A technique that uses AI to create realistic yet fake images, videos, or audio recordings of individuals.

Deep Learning: A subset of machine learning that uses neural networks with many layers (deep neural networks) to analyze various factors of data.

Machine Learning (ML): A subset of AI that involves the use of statistical techniques to enable machines to improve at tasks with experience.

Phishing: A cyber attack method that uses disguised emails (or other types of messaging applications) as a weapon. The goal is to trick the recipient into believing that the message is something they want or need—a request from their bank, for instance, or a note from someone in their company—and to enter their credentials into a malicious webpage or attachment.

Predictive Analytics: A branch of analytics that uses historical data, statistical algorithms, and machine learning techniques to identify the likelihood of future outcomes.

Ransomware: A type of malicious software designed to block access to a computer system or data until a sum of money is paid.

Spear Phishing: A more targeted form of phishing, where the attacker customizes their message based on information they have gathered about their intended victim.

Threat Intelligence: Information about current or emerging threats to systems and networks, which can help organizations prepare for, identify, and mitigate potential attacks.

Zero-Day Vulnerability: A software security flaw that is known to the software vendor but does not yet have a patch in place to fix. Because they are unknown, they pose a higher risk to systems.

Additional Resources

Books:

Artificial Intelligence: A Guide for Thinking Humans - Melanie Mitchell
An accessible introduction to AI that covers its history, technology, and potential impacts.

Cybersecurity and Cyberwar: What Everyone Needs to Know - P.W. Singer and Allan Friedman
A thorough overview of cybersecurity concepts and their implications for both individuals and nations.

Deep Learning - Ian Goodfellow, Yoshua Bengio, and Aaron Courville
A comprehensive textbook on deep learning, explaining fundamental principles and modern techniques.

Websites:

AI4ALL (www.ai-4-all.org)
A nonprofit working to increase diversity and inclusion in AI education, research, development, and policy.

National Institute of Standards and Technology (NIST) Cybersecurity Framework (www.nist.gov/cyberframework)
Provides guidelines and best practices for managing cybersecurity risk.

OpenAI (www.openai.com)
A research organization dedicated to ensuring that artificial general intelligence benefits all of humanity.