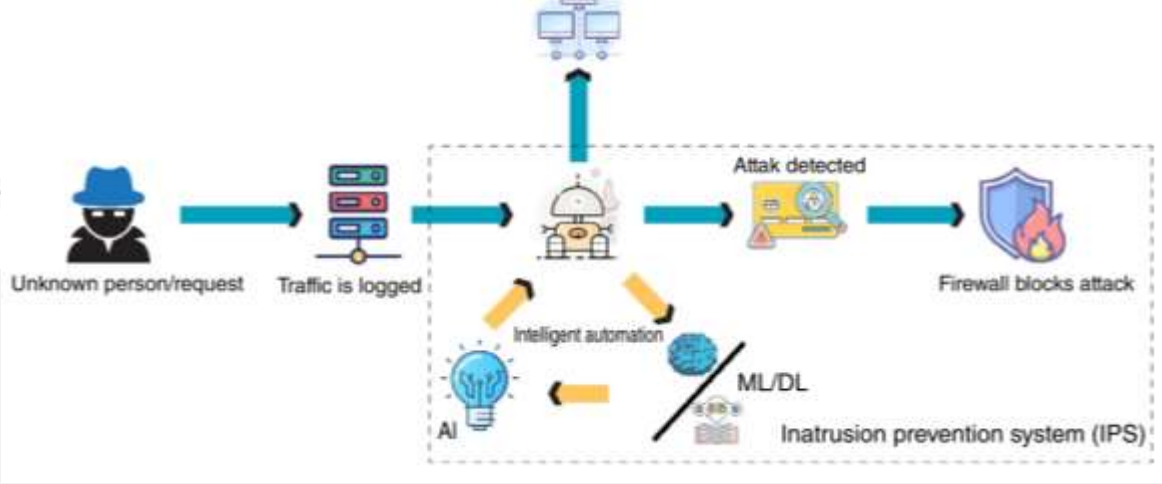
IS ARTIFICIAL INTELLIGENCE ABOVE THE LAW A FORENSIC PERSPECTIVE

New York Joint Cybersecurity Conference 2023

Kathy Braun

ARTIFICIAL INTELLIGENCE 101

1. Define automation, AI, machine learning, and deep learning.
2. Identify the major IA tools and technologies that are currently impacting the field of digital forensics.
3. Identify and explain IA-based frameworks for digital forensics.
4. Address the significant impacts of implementing these tools, technologies, and frameworks in digital forensics.
5. Discuss the challenges facing the integration of IA into digital forensics.



What have been the publicized legal disputes?

WHAT DO WE NEED TO UNDERSTAND

ALGORITHMS AND MORE ALGORITHMS

PROGRAMMING /INFORMATION SYSTEMS
WHAT DOES A FORENSIC ANALYST NEED TO UNDERSTAND

Machine Learning

Machine learning uses AI algorithms to learn from its experiences over time after an initial data input. Therefore, machine learning in SIEM

takes cybersecurity rules and data to help facilitate security analytics.

LLM Model

LLM is the means by which a computer system or algorithm can consume large amounts of data and ultimately can draw conclusions

Deep Learning Model

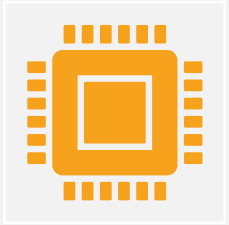
DL algorithm simplifies the dataset and verifies what it has learned.

Threshold logic

The combination of algorithms and mathematics. Artificial neural networks are used for solving artificial intelligence problems; to attempt to emulate human logic. It is the (backbone of "deep learning algorithms)

Computer Code

Series of steps that machines can execute. Code is composed in a high-level language that is then automatically translated into instructions that machines understand.



Algorithms

Series of steps for solving a problem, completing a task or performing a calculation. Algorithms are usually executed by computer programs, but the term can also apply to steps in mathematics for human problem solving.

Data Sets

A collection of related sets of information that is composed of separate elements but can be manipulated as a unit by a computer:

Unsupervised	Supervised
Unsupervised (baseline) learning is whether the data is linked to an outcome Can be taught using deep learning to detect previously unknown enterprise network threats.	Supervised machine learning, when applied to historical data is used to predict alert classification which is used in SIEM with the inclusion of Threat Intelligence, e.g., hash values, know bad Ip, known CVE vulnerability issues Supervised applications of machine learning can sort through clean, structured data that allows for clear rules and algorithms.

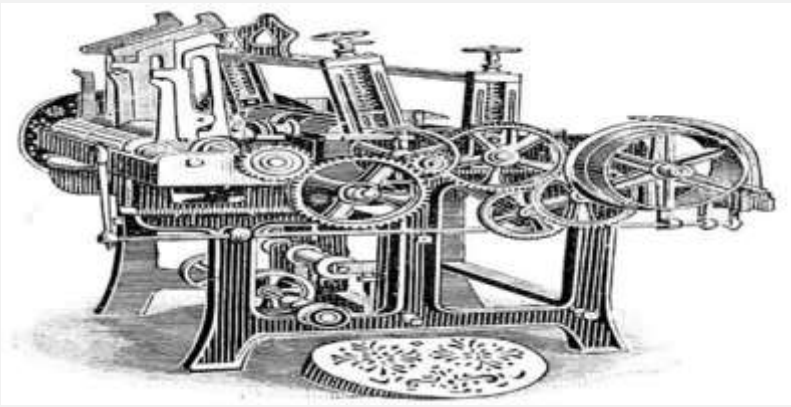
ARTIFICIAL INTELLIGENCE AND SIEM
USER AND ENTITY BEHAVIOR ANALYTICS (UEBA)

Monitors user behaviors, seeks out anomalies in those behaviors, and investigates security incidents

Prediction

ability to predict future data from previous patterns – previous breaches..

Clustering



With the right configurations machine learning can perform basic remediation tasks.

Threat Intelligence Feeds

Clustering allows ML to identify unknown values and group them together based on detected similarities for forensic investigation.

Incident Response ML can provide recommendations based on previous incident response efforts to facilitate future efforts.

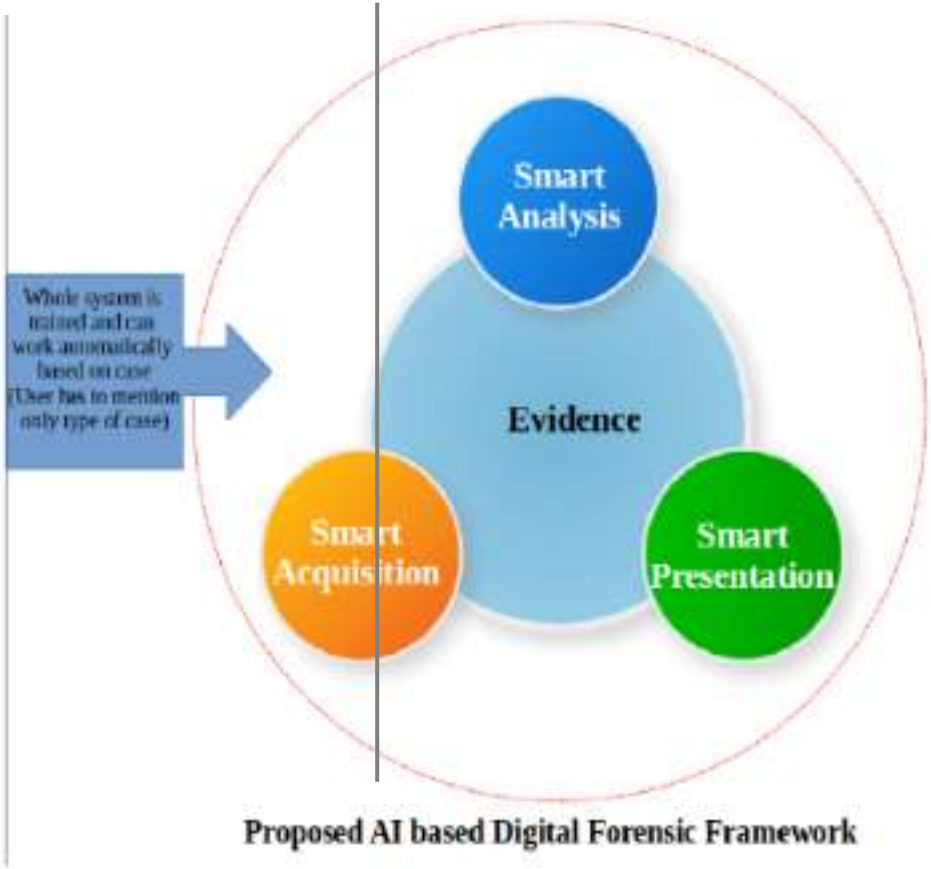
- <u>Profiling</u>: Grouping users or activity with similar characteristics to create a baseline of normal activity.
- <u>Social network analysis</u>: Identifying communities or groups within social networks based on connections and interactions between individuals
- <u>Anomaly detection:</u> Detecting abnormal behavior, such as network intrusions or suspicious bank transactions

The idea is that “playbooks” can be created based on use cases aligned with “training data sets”

FRAMEWORKS

Non-AI VERSUS PROPOSED AI FORENSIC FRAMEWORK

Non-AI Digital Forensic Framework



Proposed AI based Digital Forensic Framework

TECHNOLOGY
WHAT ARE WE DEALING WITH

THE FBI ADMINISTRATION PREDICTS THAT AI INCREASES TOOLS AT THE BAD GUY'S DISPOSAL

- Bad guys using Artificial Intelligence will increase the current level of threats and improve their malware.
- Artificial Intelligence will Create Rapid Fire of malware using the DarkWeb for AI type of malware.

Homeland Security News Wire

FraudGPT uses include writing malicious code, creating undetectable malware and hacking tools, writing phishing pages and scam content, and finding security vulnerabilities. Subscriptions start at US$200 a month through to US$1,700 for an annual license.

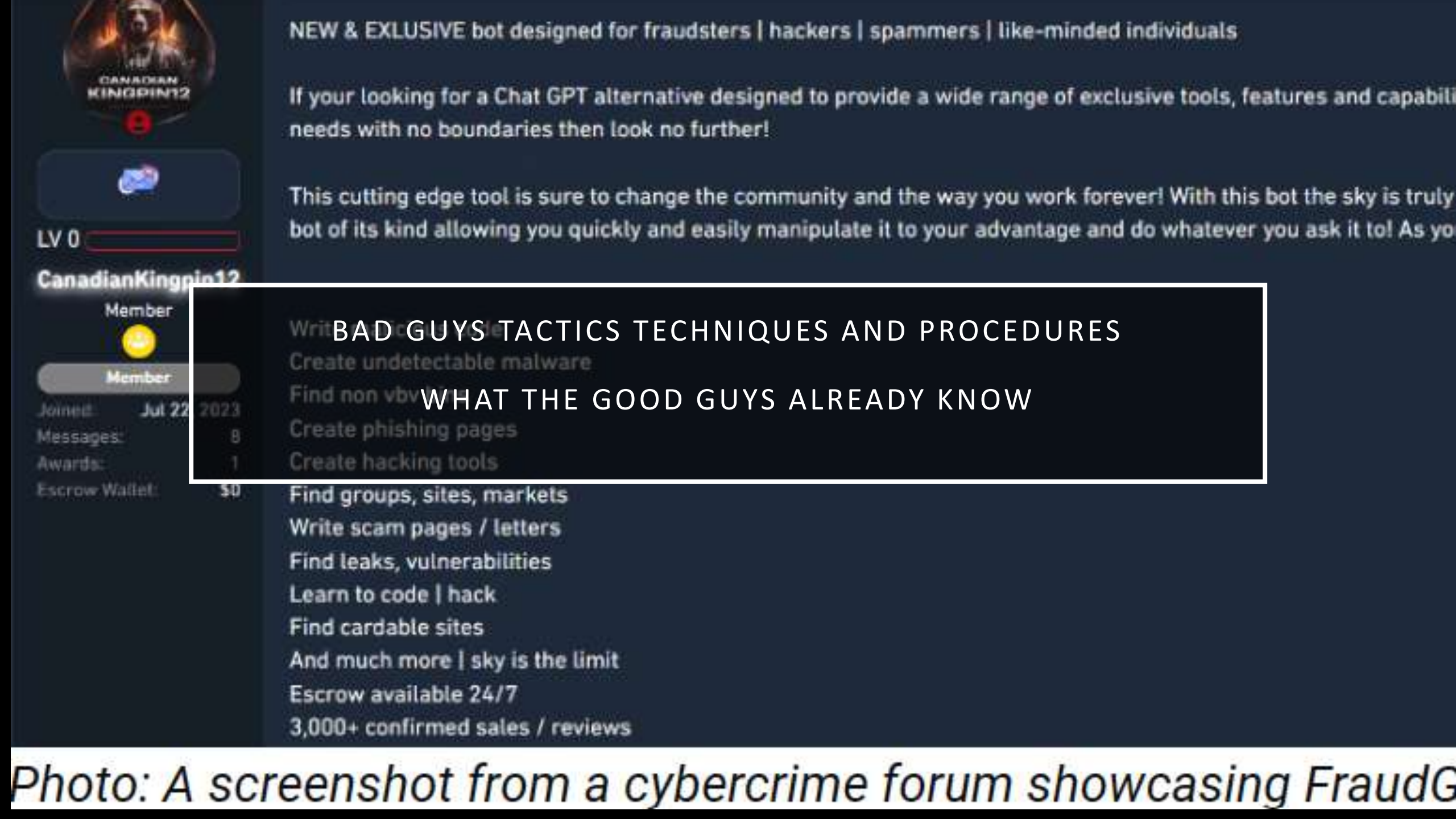
The Federal Trade Commission (FTC) is already enforcing "algorithmic disgorgement," where it forces tech firms to kill problematic algorithms along with any ill-gotten data that they used to train them.

Your Bard conversations are someone else's Google results
You may want to watch what you discuss with Google Bard or any other AI chatbot.
Written by Maria Diaz, Staff Writer on Sept. 27, 2023

Users on X, formerly Twitter, shared screenshots showing links to conversations with Bard that are showing up in Google Search results.

Haha 😂 Google started to index share conversation URLs of Bard 😹 don't share any personal info with Bard in conversation, it will get indexed and may be someone will arrive on that conversation from search and see your info 😳
Also Bard's conversation URLs are ranking as… pic.twitter.com/SKGXJD9KEJ

— Gagan Ghotra (@gaganghotra_) September 26, 2023
This appears to be unintentional, as Google only intends to allow users to create a public link to share conversations with others. Anyone with this link can see the conversation, but Google appears to be indexing them.

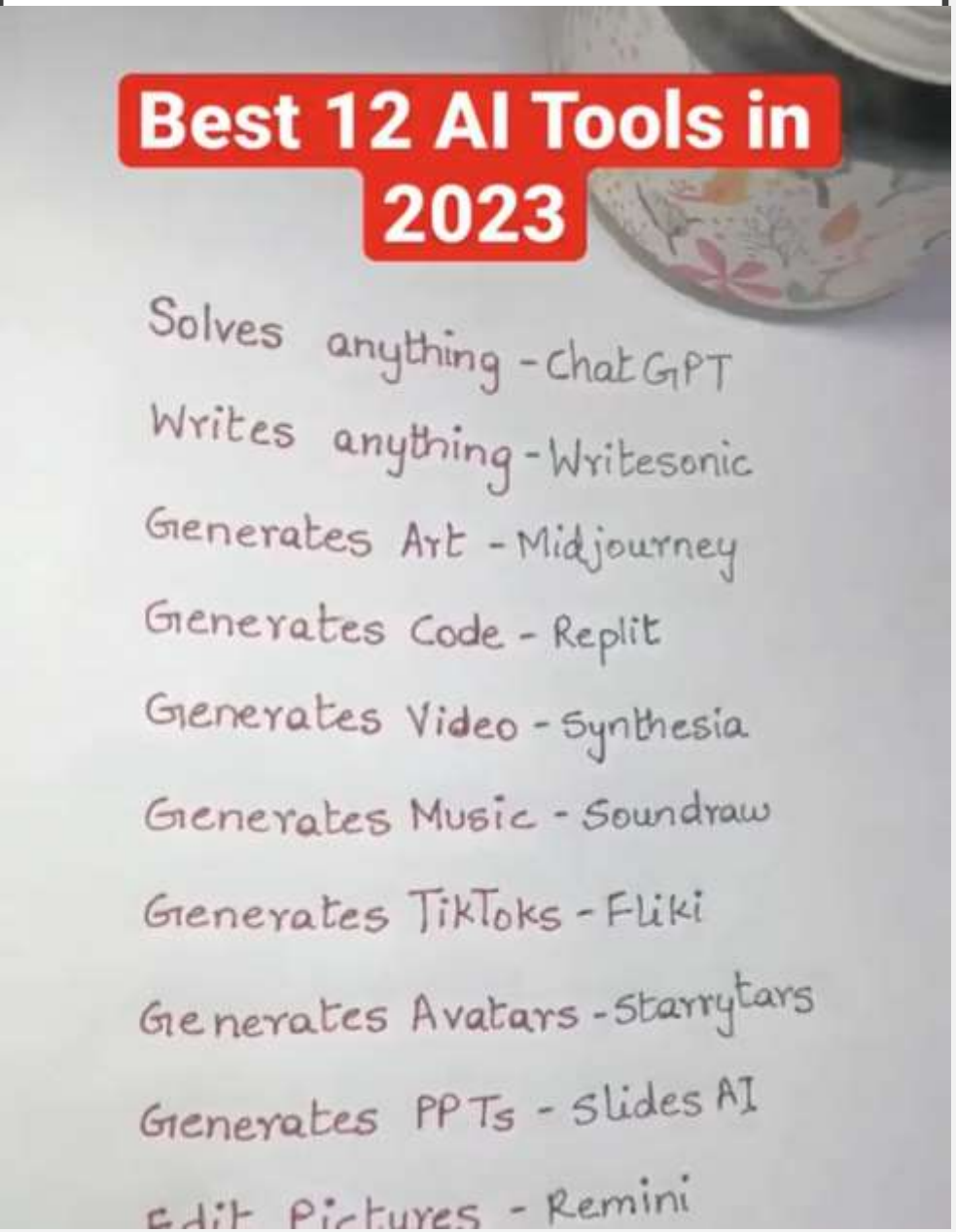


BAD GUYS TACTICS TECHNIQUES AND PROCEDURES

WHAT THE GOOD GUYS ALREADY KNOW

Photo: A screenshot from a cybercrime forum showcasing FraudG

ARTIFICIAL INTELLIGENCE SUPPLY CHAIN OPPORTUNITY



Facebook Flooded with Ads and Pages for Fake ChatGPT, Google Bard and other AI services, Tricking Users into downloading Malware - Check Point Blog

The ROI on these attacks is just too sweet for professional adversaries to resist." supply chain attacks have increased by 742% over the last three years.

The iconic Target breach of late 2013 was a supply chain breach. The attackers got into Target using credentials stolen from its HVAC provider

The 2018 breach of Ticketmaster was another supply chain breach. A Ticketmaster software was breached, and software was modified and weaponized. This was automatically downloaded to Ticketmaster.

MOVE it is a managed file transfer (MFT) software for secure data transfer within teams, departments, and companies. It encrypts files and employs secure File Transfer Protocols.

VARIATION ON A THEME
CREATING MALICIOUS SOFTWARE, SUCH AS RANSOMWARE, FAKE WEBSITES, USING SOCIAL MEDIA SITES, AND PHISHING CAMPAIGNS

BAD Guys What they can do

WormGPT DarkWeb Sales

Its primary purpose: generating convincing business email compromise (BEC) attacks, which use fake, personalized messages to access sensitive accounts.

Like ChatGPT, the tool is easy to use, meaning cybercriminals with limited experience could find success with it.

New capability?

- AI can quickly scrape the internet for personal details about a target to develop a tailored scam or carry out identity theft.
- AI can also assist in developing and deploying malware, including pinpointing vulnerabilities in software before they can be patched.
- Its primary uses are generating sophisticated phishing and business email attacks and writing malicious code.

Side Note: *"Supply chain attacks purposefully target the smaller organizations first because they're less likely to have a robust cybersecurity setup, and they can use those companies to get to the bigger fish,"*

Automated accounts known as social bots (fake human accounts) are a key part of propaganda campaigns that attempt to simulate and influence human behavior to produce content and interact on social media (trick people).

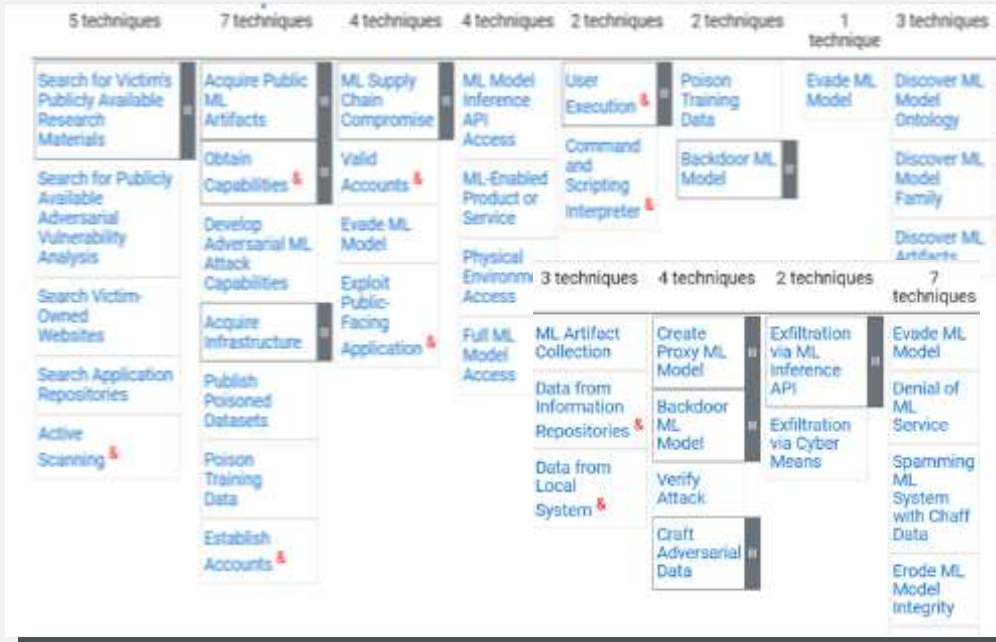
Deepfake or using algorithms to block content as examples.

Data Poisoning: With knowledge of the training data, an attacker can inject malicious samples or manipulate existing data points to introduce biases. This can lead to incorrect predictions or unauthorized access to sensitive information.

Model Extraction: Attackers may attempt to extract the underlying model to reverse-engineer proprietary algorithms. This can lead to intellectual property theft or the creation of malicious replicas.

System Disruption: White-Box attacks can also aim to disrupt the normal functioning of the AI system by exploiting vulnerabilities in the source code or the model's architecture.

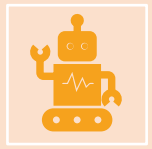
MITRE Tactics Techniques Procedures



COUNTERMEASURES

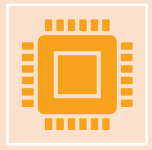
Repeatable, Documented, and Verifiable Process

UEBA AND INCIDENT RESPONSE



AI Improvements: Been around for 8-9 years or more…

Threat Hunting: UEBA analyzes user behavior patterns and detect anomalies that may indicate compromised accounts or insider threats.



As **SIEM's that use UEBA now coined AI** uses pattern recognition, aggregation/correlation, basic data analysis which falls under the category of machine learning, which falls under the category of artificial intelligence.

Machine learning (ML)

Predictive Analytics: AI algorithms can analyze historical data and identify trends, to address potential cyber threats.

Automated Security Research: Threat intelligence including vulnerability data to perform basic remediation.

- Block an IP
- Quarantine a known bad file
- Create an Alert and align with a rating such as very high, high, moderate, or low

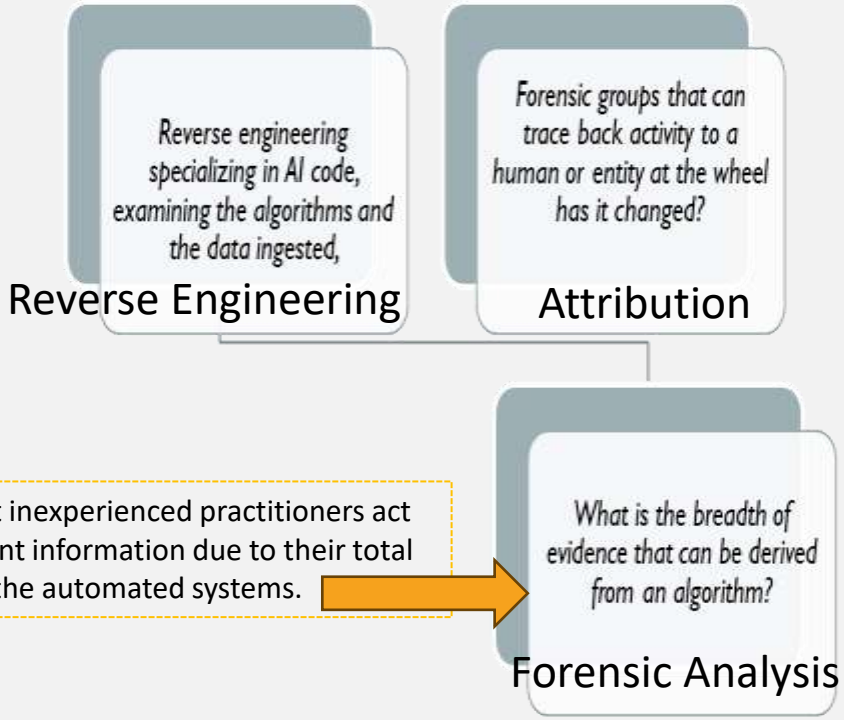
INCIDENT RESPONSE NEXT GENERATION

Intrusion Detection Systems (IDS): AI-powered *Application that can analyze network traffic and detect anomalies that may indicate unauthorized access attempts or suspicious behavior.*

User Entity Behavior Analytics (UEBA)*: A category of software that helps security teams identify and respond to insider threats.*

Threat Intelligence and Monitoring*: The collection and analysis, of threat intelligence for indicators of compromise.*

Malware Detection and Prevention*: AI algorithms can analyze file attributes, behavior patterns, and network traffic to identify and mitigate potential malware threats such as zero-day.*



The risk that inexperienced practitioners act on insufficient information due to their total reliance on the automated systems.

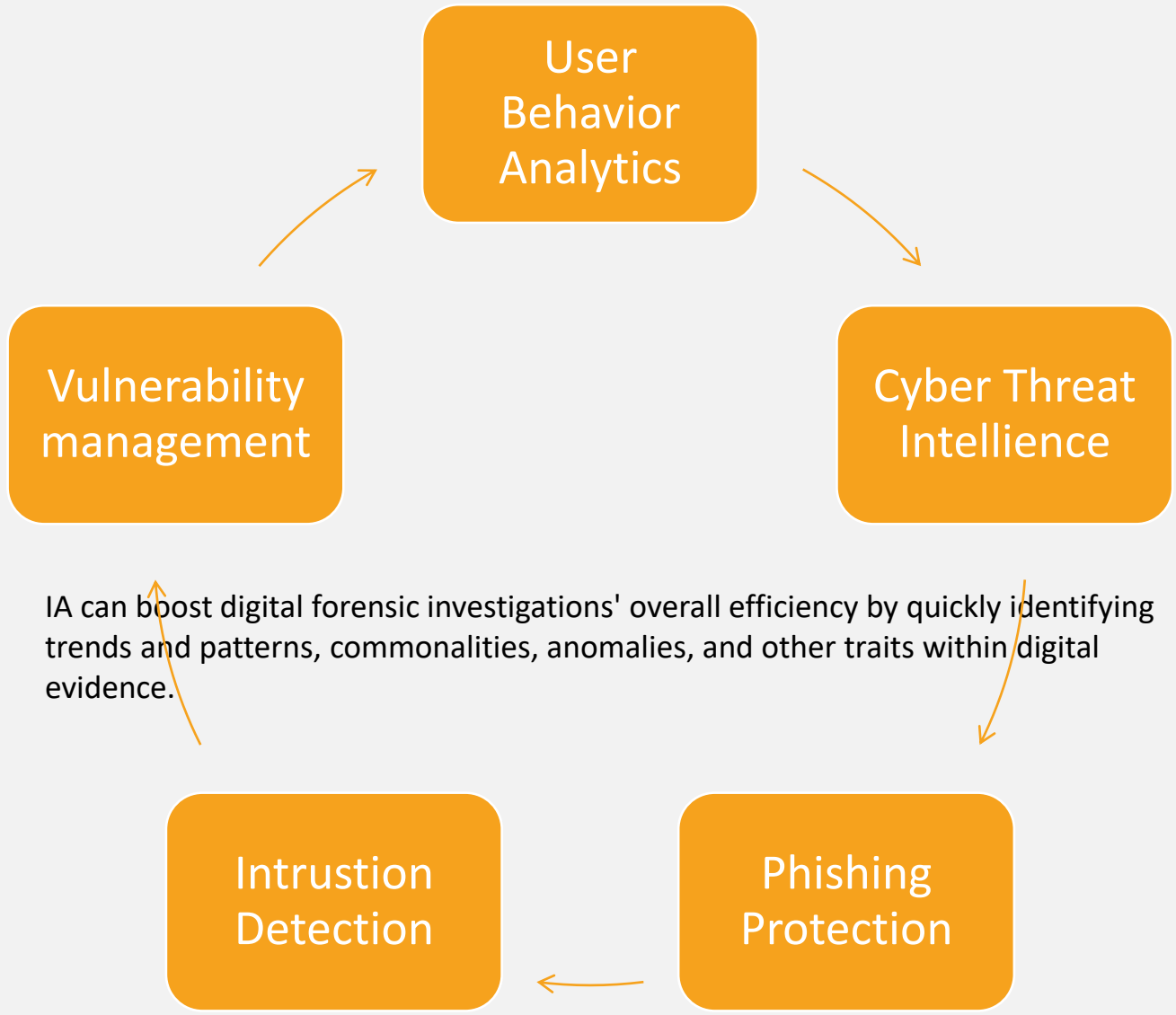
Multimedia Content, Copyright Enfringement and Crime

AI algorithms can analyze images and videos to detect objects, faces, locations. assists investigators in identifying illegal content, such as explicit images or videos involving minors, ***comparing them against known databases****.*

INCIDENT RESPONSE ENDS AND FORENSIC BEGINS

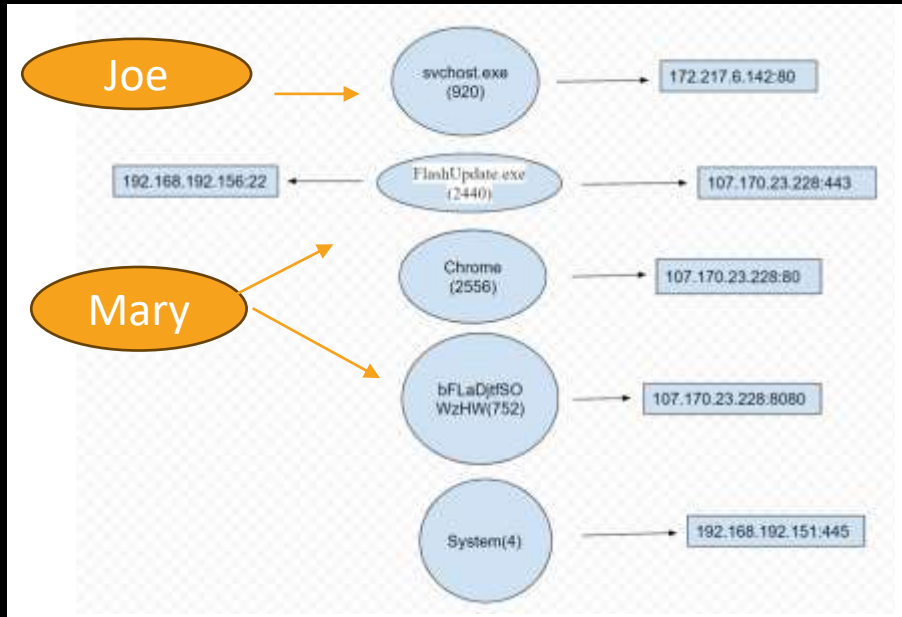
Incident Response: The overarching process that an organization will follow in order to prepare for, detect, contain, and recover from a data breach.

Digital Forensics: A subset of forensic science that examines system data, user activity, and other pieces of digital evidence to determine if an attack is in progress, has already been determined as threat, and who, what, where, how, and why may be behind the activity.



IA can boost digital forensic investigations' overall efficiency by quickly identifying trends and patterns, commonalities, anomalies, and other traits within digital evidence.

Mapping Digital Investigations



AI algorithms can analyze email communications and user activity to identify relevant conversations, attachments, and timestamps, enabling investigators to reconstruct **timelines and establish connections between individuals and events**.

Threat Hunting: AI-powered systems can assist security analysts in proactively searching for hidden threats by analyzing historical data, network traffic, and logs, significantly reducing the time required to identify potential risks.

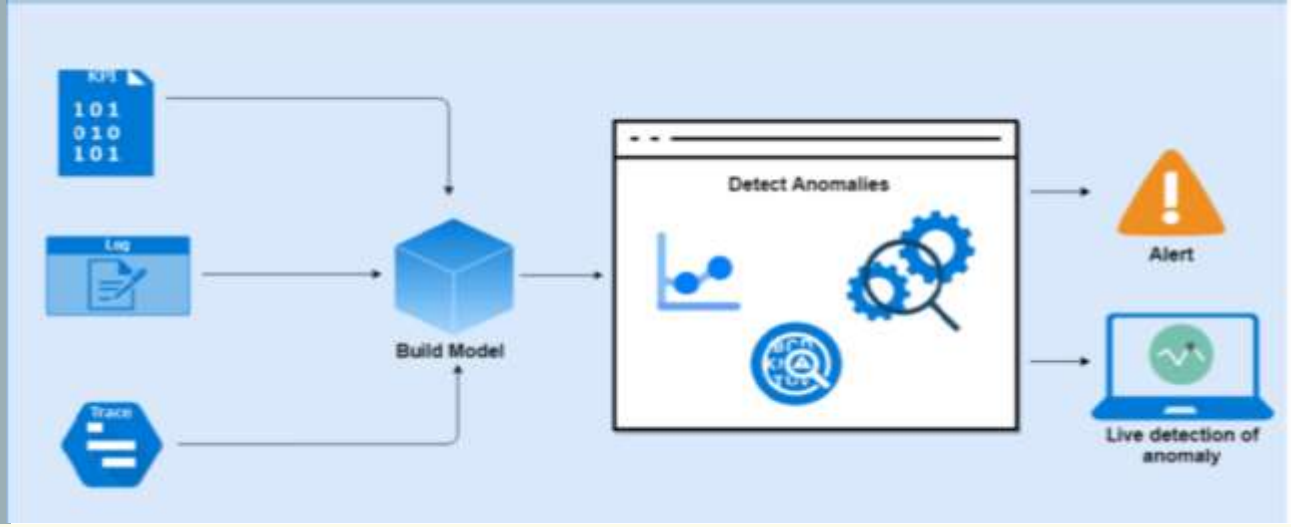
CHALLENGES
GAME OF CAT AND MOUSE
AUTOMATION ENABLED DIGITAL FORENSICS

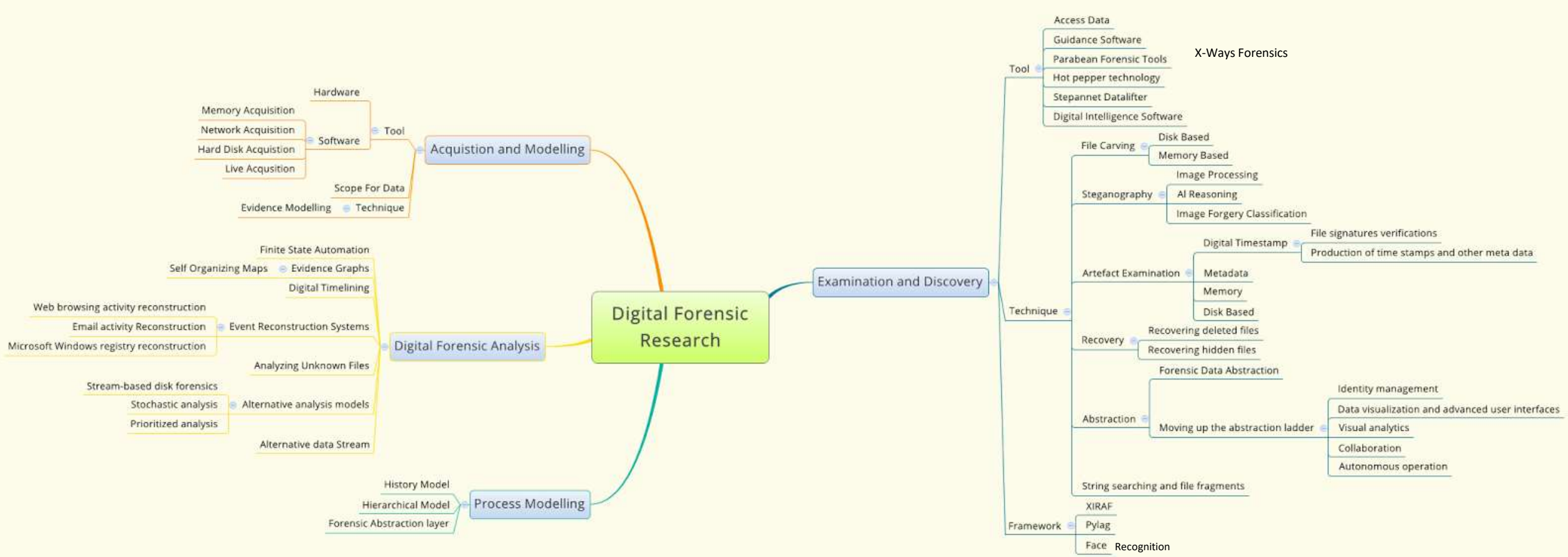
AI can analyze file headers and signatures to locate concealed or deleted files, providing valuable leads for forensic investigations.

We could always do this!

FORENSIC INVESTIGATION HUMAN HANDLING

AI models can be powerful tools, but they should be used in conjunction with human expertise and oversight for conducting advanced threat investigations.





RULE OF LAW
ARTIFICIAL INTELLIGENCE
NO PROOF ABOVE THE LAW

INDIVIDUAL CULPABILITY

WHAT ARE THEIR DUTIES (AUDITABLE), SPECIFICALLY AS CEO OR AS A BOARD MEMBER

Several existing regulations and guidelines provide a foundation for addressing ethical considerations in the field of AI-driven cybersecurity.

- General Data Protection Regulation (GDPR): The GDPR, implemented in the European Union, establishes principles for the lawful and fair processing of personal data.

It emphasizes the need for transparency, data protection, and user consent, which are relevant in the context of AI-driven cybersecurity systems.

While not specific to AI, this framework includes principles applicable to the ethical use of AI technologies in cybersecurity.

- IEEE Ethically Aligned Design: The Institute of Electrical and Electronics Engineers (IEEE) has developed a comprehensive framework called Ethically Aligned Design, which encompasses various domains, including AI and cybersecurity.

This framework promotes the integration of ethical considerations into the design, development, and deployment of AI systems.

NIST Framework for Improving Critical Infrastructure Cybersecurity: The National Institute of Standards and Technology (NIST) provides a framework that guides organizations in managing and mitigating cybersecurity risks.

Verify systems are locked down /configuration management

Proactive Integrated Risk Management:

Considering the hefty costs of cyberattacks such as identity theft and data loss, organizations should strongly consider investing in improving the security, privacy, and confidentiality of their cyber infrastructure and information assets.

Privacy, Bias, Ethics, Trust and Cybersecurity

FTC To Hold Facebook CEO Mark Zuckerberg Liable For Any Future Privacy Violations

Data Governance: Organizations must enforce compliance, security, and privacy first. Data sources and their characteristics should be well-documented and available to stakeholders via transparent data handling and disclosure policies.

Algorithmic Transparency: Civil Rights and Algorithms

With respect to AI, The American Data Privacy and Protection Act (ADPPA) a Federal Regulation includes a provision, —Section 207: Civil Rights and Algorithms—under which covered entities or service providers "may not collect, process, or transfer covered data in a manner that discriminates in or otherwise makes unavailable the equal enjoyment of goods or services on the basis of race, color, national origin, sex, or disability."

- Covered data is defined as “information that identifies or is linked or reasonably linkable to one or more individuals, including derived data and unique identifiers = Potential Victim

CURRENT LANDSCAPE
EVIDENCE ADMISSIBILITY

WHAT IS DIGITAL FORENSICS

Digital forensics is the identification, extraction, interpretation and documentation of computer evidence which can be used in a court of law.

Privacy Cases

Over a two-week period in 2023 a number of federal class action lawsuits were filed in the US District Court for the Northern District of California against the developers of some of the most well-known generative AI products on the market, including OpenAI, Inc. (OpenAI) and Alphabet Inc./Google LLC (Google).

Right of Publicity and Facial Recognition Cases

In April 2023, a television personality filed a class action complaint against software developer.Claiming that AI-powered "Reface" application, which allows users to digitally "swap" their faces with celebrities and public figures in photos and videos, constitutes a violation of the common law right of publicity, protected by California's Right of Publicity Statute.

Suit against OpenAI LP (OpenAI) alleging that OpenAI stole private and personal information belonging to millions of people by collecting publicly-available data from the Internet to develop and train its generative AI tools including ChatGPT , and an AI speech generator.

The complaint also includes claims for violations of the Electronic Communications Privacy Act, the Computer Fraud and Abuse Act (CFAA), various state consumer protection statutes, and a number of common law claims.

Trademark Cases

(Getty), filed suit against Stability AI asserting claims of copyright and trademark infringement. Claims that AI "scraped" Getty's website for images and data used in the training of its image-generating modelwith the aim of establishing a competing product or service.

Claim purports Stability AI has provided false copyright information in violation of 17 U.S.C. § 1202(a).

https://www.linkedin.com/posts/paul-curley-a083442b_ai-related-do-claims-ugcPost-7108134838188077058-Hgab/?utm_source=share&utm_medium=member_ios

AI-related claims against directors and officers (D&O) can arise from a variety of circumstances where the leadership team is alleged to have failed in their duties or responsibilities concerning artificial intelligence. Here are some examples of AI-related claims that could lead to legal action against D&Os:

1. **Data Privacy Breach**: Directors and officers may be held accountable if their organization experiences a data breach involving AI systems due to inadequate cybersecurity measures, leading to the exposure of sensitive customer information.
2. **Algorithmic Bias and Discrimination**: If AI algorithms used by a company are found to discriminate against certain groups (e.g., in hiring, lending, or customer service), D&Os may face claims related to discrimination and bias.
3. **Regulatory Non-Compliance**: Directors and officers may be accused of failing to ensure that their AI systems comply with relevant laws and regulations, such as GDPR, HIPAA, or industry-specific standards, leading to regulatory fines and legal actions.
4. **Securities Fraud**: Misrepresentations or omissions regarding the capabilities or risks associated with AI technology in public disclosures can lead to securities fraud claims by shareholders or regulators.
5. **Intellectual Property Infringement**: If a company is accused of infringing on the intellectual property rights of others through its AI technology or engaging in misappropriation of trade secrets, D&Os could be named in lawsuits related to intellectual property infringement.
6. **Failure to Mitigate AI Risks**: Directors and officers may face claims if they are alleged to have neglected their duty to identify, assess, and mitigate risks associated with AI, such as AI system failures, ethical concerns, or potential societal impacts.
7. **Product Liability**: If AI-driven products or services malfunction or cause harm to users or third parties, D&Os might be held responsible for their role in the development, testing, and release of those products.
11. **Shareholder Derivative Lawsuits**: Shareholders may file derivative lawsuits against D&Os alleging that they failed to oversee AI-related risks, breached their fiduciary duties, or engaged in misconduct related to AI.

It's important to note that the specific circumstances and legal claims against directors and officers in AI-related cases can vary widely based on the industry, the nature of the AI application, and the jurisdiction in which the company operates. To mitigate these risks, D&Os should prioritize risk management, compliance, and ethical considerations when implementing AI technologies in their organizations. Adequate D&O insurance coverage is also essential to protect the personal assets of directors and officers in the event of legal claims.

THANK YOU

Kathy.braun@ljmcnyc.com