

Generative AI and OSINT Threats and Trends to Hawaii Critical Infrastructure

JUNE 27 2024



CISA AI Focus

National AI Initiative (NAII) Act of 2020:

Coordinated complementary AI R&D, demonstration activities among FCEB, DOD, IC.

AI in Government Act of 2020:

Established the AI Center of Excellence within GSA.



EO 13859: Maintaining American Leadership in AI:

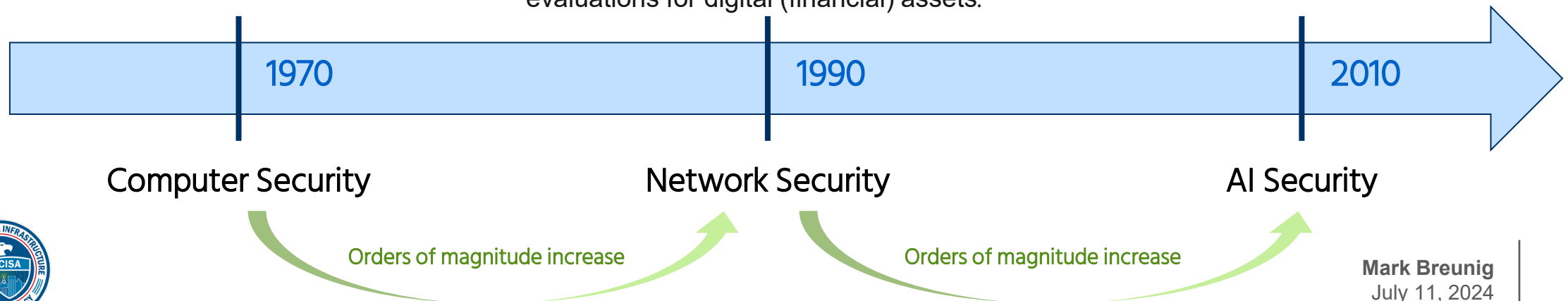
Established federal principles and strategies to strengthen the nation's capabilities in AI.

EO 13960: Promoting the Use of Trustworthy AI in the Federal Gov't:

Required Agencies to inventory and share AI use cases.

EO 14067: Federal Policy on Ensuring Responsible Development of Digital Assets:

Requires systemic risk evaluations for digital (financial) assets.



What is Generative AI?

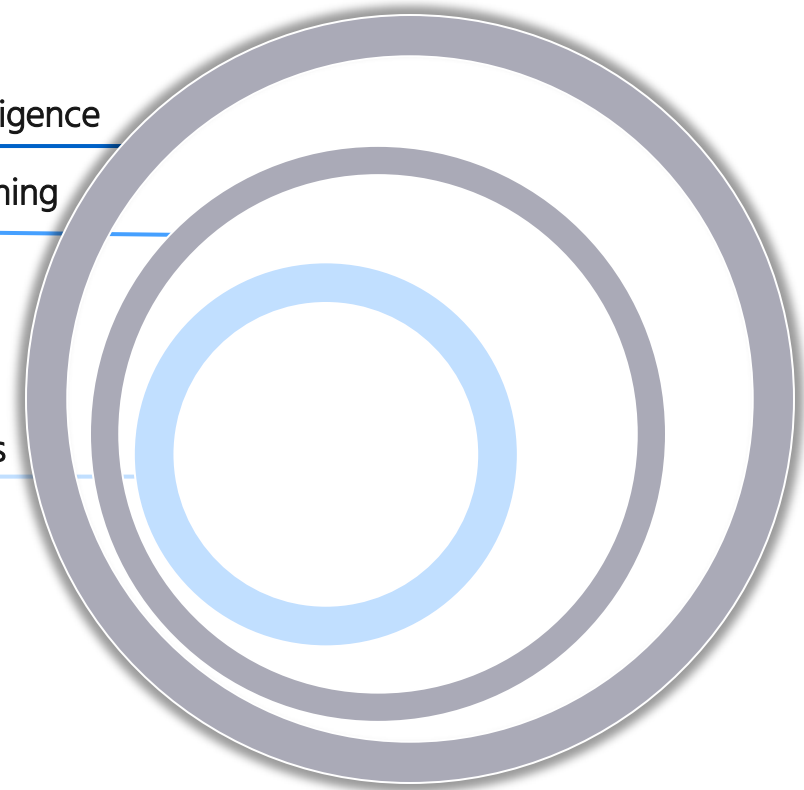
GPT = Generative Pretrained Transformer

- Very large text data sets are used for training to build a predictive capacity
 - GPT-3 (OpenAI) was trained on approximately 45 TB of text data (equivalent to a quarter of the entire Library of Congress)
- Has, until recently, required billions of dollars in funding
- Hosting is now offered via Amazon, Google, Microsoft, and many others

Artificial Intelligence

Machine Learning

Large Language Models



Subsets of AI



How Can AI Be Abused?

Bad Data Sets:





- WormGPT
 - Malicious counterpart of ChatGPT
 - Focused on writing phishing emails and malicious code
 - Built on GPTJ (similar to GPT-3) and operates on approximately 6 *billion* parameters with a vocabulary of 50,257 words
- FraudGPT
 - Intended to help automate malicious tasks
- PoisonGPT, Evil-GPT, XXXGPT, WolfGPT... what's next?



How Can AI Be Abused? (cont.)

Exploitation of AI Services:

- Prompt injection and Data Poisoning
 - (Input Manipulation Attack)
- Social Engineering... of AI?
 - (Membership Inference Attack)
- Abuse of Popularity

Name		Description
Poisoning		Modifying the ML model through deceptive training inputs.
Evasion		Making illegitimate inputs appear legitimate.
White Box		Training inputs and/or model parameters are known.
Black Box		Model is hidden, but inputs and outputs are visible.

OWASP Machine Learning Top 10

- <https://owasp.org/www-project-machine-learning-security-top-10/>



Generative Red Teaming

DEFCON does it again...

...but why is this important?

Reconnaissance 2 techniques	Resource Development 6 techniques	Initial Access 1 technique	ML Model Access 4 techniques	Execution 1 technique	Persistence 2 techniques	Defense Evasion 1 technique	Discovery 3 techniques	Collection 1 technique	ML Attack Staging 5 techniques
Search for Victim's Publicly Available Research Materials	Acquire Public ML Artifacts	ML Supply Chain Compromise	ML Model Inference API Access	User Execution: Unsafe ML Artifacts	Poison Training Data	Evade ML Model	Discover ML Model Ontology	ML Artifact Collection	Train Proxy ML Model
Search for Publicly Available Adversarial Vulnerability Analysis	Obtain Capabilities: Adversarial ML Attack Implementations		ML-Enabled Product or Service		Poison ML Model		Discover ML Model Family		Replicate ML Model
	Develop Capabilities: Adversarial ML Attack Implementations		Physical Environment Access				Discover ML Artifacts		Poison ML Model
	Acquire Infrastructure:		Full ML Model Access						Verify Attack
									Craft Adversarial Data

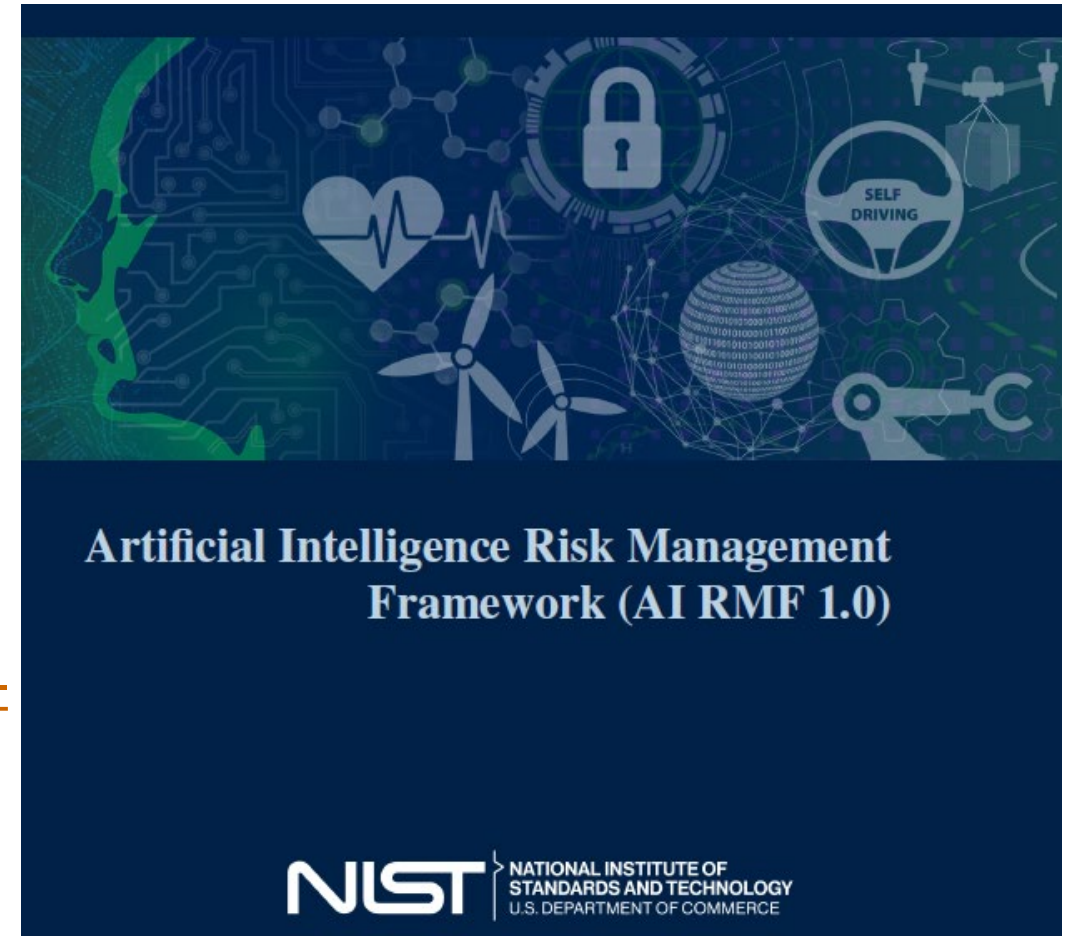
MITRE's Adversarial Threat Landscape for Artificial-Intelligence Systems (ATLAS)

Do you know what you need to ask third-party providers of AI services?



Shadow AI

- What is it?
- Do you have internal guidelines?
 - What framework are you using?
 - Are your guidelines formalized?
- Hello, it's NIST again!
 - AI Risk Management Framework 1.0
 - <https://www.nist.gov/itl/ai-risk-management-framework>
 - Playbook, Roadmap, and Crosswalk



APT Concerns

CISA Advisory on PRC Activity:

- <https://www.cisa.gov/news-events/cybersecurity-advisories/aa23-270a>

CISA and NSA Red + Blue Team Top 10:

- <https://www.cisa.gov/news-events/cybersecurity-advisories/aa23-278a>
- Newly tracked groups appearing regularly, tracked by multiple research groups
- Known groups with a history are not going away
- No significant findings in relation to APT use of AI (doesn't mean it isn't occurring)



What is OSINT:



Techniques:

Tools and Techniques	
Social Media	Scans.io / Shodan
Internet Archive	Google Searching
Builtwith	Maltego
Robots.txt	Networking Tools (Nmap, nslookup, dig, etc.)



OSINT Framework:



www.osintframework.com

Mix of passive and active tools

Ask: What information is important, and how can we recognize it?



More Tools:

Shodan: www.shodan.io

Internet Archive: <https://web.archive.org>

Builtwith: www.builtwith.com

Robots.txt: www.domain.com/robots.txt

Google: www.google.com



Google:

Google Advanced Search

google.com/advanced_search

Google Sign in

Advanced Search

Find pages with...

To do this in the search box

all these words:

this exact word or phrase:

any of these words:

none of these words:

numbers ranging from: to

Type the important words: tricolor rat terrier

Put exact words in quotes: "rat terrier"

Type OR between all the words you want: miniature OR standard

Put a minus sign just before words you don't want: -rodent, -"Jack Russell"

Put 2 periods between the numbers and add a unit of measure: 10..35 1b, \$300..\$500, 2010..2011

Then narrow your results by...

language: any language Find pages in the language you select.

region: any region Find pages published in a particular region.

last update: anytime Find pages updated within the time you specify.

site or domain: Search one site (like wikipedia.org) or limit your results to a domain like .edu, .org or .gov

terms appearing: anywhere in the page Search for terms in the whole page, page title, or web address, or links to the page you're looking for.

file type: any format Find pages in the format you prefer.

usage rights: not filtered by license Find pages you are free to use yourself.

Advanced Search



Google Search Operators:

Search operator	What it does	Example
" "	Search for results that mention a word or phrase.	"steve jobs"
OR	Search for results related to X or Y.	jobs OR gates
	Same as OR:	jobs gates
AND	Search for results related to X and Y.	jobs AND gates
-	Search for results that don't mention a word or phrase.	jobs -apple
*	Wildcard matching any word or phrase.	steve * apple
()	Group multiple searches.	(ipad OR iphone) apple
define:	Search for the definition of a word or phrase.	define:entrepreneur
cache:	Find the most recent cache of a webpage.	cache:apple.com
filetype:	Search for particular types of files (e.g., PDF).	apple filetype:pdf

ext:	Same as filetype:	apple ext:pdf
site:	Search for results from a particular website.	site:apple.com
related:	Search for sites related to a given domain.	related:apple.com
intitle:	Search for pages with a particular word in the title tag.	intitle:apple
allintitle:	Search for pages with multiple words in the title tag.	allintitle:apple iphone
inurl:	Search for pages with a particular word in the URL.	inurl:apple
allinurl:	Search for pages with multiple words in the URL.	allinurl:apple iphone
intext:	Search for pages with a particular word in their content.	intext:apple iphone
allintext:	Search for pages with multiple words in their content.	allintext:apple iphone



A real-life example:

The 2019 Iranian rocket explosion





Questions?

mark.breunig@cisa.dhs.gov
(907) 795-5673

