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.


Computer Security  Network Security  AI Security

Orders of magnitude increase

1970  1990  2010
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
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?
Exploitation of AI Services:

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

OWASP Machine Learning Top 10


<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poisoning</td>
<td>Modifying the ML model through deceptive training inputs.</td>
</tr>
<tr>
<td>Evasion</td>
<td>Making illegitimate inputs appear legitimate.</td>
</tr>
<tr>
<td>White Box</td>
<td>Training inputs and/or model parameters are known.</td>
</tr>
<tr>
<td>Black Box</td>
<td>Model is hidden, but inputs and outputs are visible.</td>
</tr>
</tbody>
</table>
DEFCON does it again…

…but why is this important?

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

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

CISA and NSA Red + Blue Team Top 10:
* [https://www.cisa.gov/news-events/cybersecurity-advisories/aa23-278a](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:

<table>
<thead>
<tr>
<th>Tools and Techniques</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social Media</td>
</tr>
<tr>
<td>Internet Archive</td>
</tr>
<tr>
<td>Builtwith</td>
</tr>
<tr>
<td>Robots.txt</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
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](www.shodan.io)

Internet Archive: [https://web.archive.org](https://web.archive.org)

Builtwith: [www.builtwith.com](www.builtwith.com)

Robots.txt: [www.domain.com/robots.txt](www.domain.com/robots.txt)

Google: [www.google.com](www.google.com)
Shodan:

156.20.66.147

Country: South Africa

Web Technologies:
- AS1 IN ONE SID PACK
- FEEDBACK
- FEAT AWESOME
- JOURNAL
- MINEAL
- PHP
- WORDPRESS

Vulnerabilities:
- CVE-2008-8509: Remotely executable backdoor in auth-gssapi in OpenSSH through 5.6 could be used by remote attackers to detect existence of users on a target system when GSS is in use. NOTE: The Discover documentation "No understand that the OpenSSH developers do not want to file such a vulnerability report in "vulnerable" as a vulnerability."
- CVE-2007-5681: The process, open function in sshd-server in OpenSSH before 5.6 does not properly prevent write operations in read-only mode, which allows attackers to create zero-length files.
# Google Search Operators:

<table>
<thead>
<tr>
<th>Search operator</th>
<th>What It does</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>+ &quot; &quot;</td>
<td>Search for results that mention a word or phrase.</td>
<td>&quot;Steve Jobs&quot;</td>
</tr>
<tr>
<td>OR</td>
<td>Search for results related to X or Y.</td>
<td>jobs OR gates</td>
</tr>
<tr>
<td>]</td>
<td>Same as OR:</td>
<td>jobs</td>
</tr>
<tr>
<td>AND</td>
<td>Search for results related to X and Y.</td>
<td>jobs AND gates</td>
</tr>
<tr>
<td>-</td>
<td>Search for results that don't mention a word or phrase.</td>
<td>jobs -apple</td>
</tr>
<tr>
<td>*</td>
<td>Wildcard matching any word or phrase.</td>
<td>steve * apple</td>
</tr>
<tr>
<td>( )</td>
<td>Group multiple searches.</td>
<td>(iPad OR iphone) apple</td>
</tr>
<tr>
<td>define:</td>
<td>Search for the definition of a word or phrase.</td>
<td>define:entrepreneur</td>
</tr>
<tr>
<td>cache:</td>
<td>Find the most recent cache of a webpage.</td>
<td>cache:apple.com</td>
</tr>
<tr>
<td>filetype:</td>
<td>Search for particular types of files (e.g., PDF).</td>
<td>apple filetype:pdf</td>
</tr>
</tbody>
</table>

| ext:            | Same as `filetype:`, search for a file with a specific extension. | 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

Donald J. Trump
@realDonaldTrump

The United States of America was not involved in the catastrophic accident during final launch preparations for the Safir SLV Launch at Semnan Launch Site One in Iran. I wish Iran best wishes and good luck in determining what happened at Site One.

9:44 AM · Aug 30, 2019
Questions?
mark.breunig@cisa.dhs.gov
(907) 795-5673