JFX, LLC
Cybersecurity Consulting

JFX Cybersecurity Briefing: Guidelines for Secure AI Usage

October 2024

AI tools offer many efficiency benefits, but also introduce potential risks such as leakage of confidential data and providing false (“hallucinated”) or inappropriate information. These guidelines are intended to help small companies with their initial use of AI tools, enabling experimentation while supporting risk protection and management oversight.

Disclose usage of AI tools

To assist learning and supervision, use of AI tools should be disclosed verbally and as a note in documents or code documentation. Suggested examples, “Grammarly was used to review a draft and to suggest edits, most of which were adopted.”; “Copilot drafted an initial response to the topic question, which was reviewed for accuracy and edited. No company or client information was provided in the prompt.”

Use paid versions of AI tools

Although service terms change, free versions of AI tools have a higher chance of using user provided information in model training, which may leak to other users. A safety recommendation is to use paid versions of AI tools, which are more likely to restrict access to user provided information. Monitor Terms of Use and FAQ statements to see if and how user provided data is retained, shared or used for training.

Don’t provide confidential information (especially PII)

To avoid the risk of confidential data leakage, do not provide confidential data to an AI prompt, especially of PII covered by data privacy laws.

Verify accuracy of results

Generative AI tools create responses based on relationships identified in training data. Most AI tools do not explain how they created their responses, which means results can be based on information incorrectly ‘learned’ in training, spurious data relationships, or other AI “hallucinations” - true sounding statements that are untrue. As Einstein said, all AI generated information should be verified for correctness and appropriateness.

Perform Supervisory Responsibilities

Especially for legally binding and regulatory supervised actions, the person using an AI tool is responsible for its results, as is the supervisor with oversight for that person: the same supervisory responsibility as for any other activity. “AI said it so I assumed it was true” is not an acceptable defense, nor is “I didn’t know they were using AI”.

Data Retention

To meet data retention requirements, companies may need to archive certain AI interactions, e.g., if writing regulated communications. If archiving is not performed by the system then users should manually save and archive copies of their AI sessions.

Common Procedures

As AI tools expand beyond experimentation to become part of normal operations, organizations should document how AI tools are used for certain tasks. Companies with a technology focus may have a software development lifecycle to guide this process, other firms can just focus on making sure there are approved common procedures so results don’t vary depending on who does the work.

Summary

AI tools offer many benefits to improve the efficiency of existing tasks and to develop new functions. While smaller firms may get the biggest competitive benefits from AI, they also have fewer resources to evaluate and manage AI’s risks. These guidelines should give companies a lightweight approach to supervising and using AI tools so they can experiment while managing the most common risks,

- Disclose when and how AI is being used
- Avoid entering confidential information; use paid tools
- Verify any facts provided by an AI response
- Employees and supervisors are responsible for work produced using AI
- Save copies of AI interactions for critical or regulated topics
- Once in production, set common procedures to get consistent results

As AI tools and their use may change quickly, update these as appropriate for your firm.