Published on *The National Law Review* https://natlawreview.com

Treasury Department Warns Financial Institutions to Prepare for AI-Age Fraud — AI: The Washington Report

Article By:

David G. Adams

Bruce D. Sokler

1. Pursuant to President Biden's October 2023 AI executive order, the US Department of Treasury (Treasury) released a report on cybersecurity risks in the financial services sector in March 2024.
2. While recognizing the benefits that AI-based cybersecurity tools provide, the report cautions financial institutions to be aware of both the special vulnerabilities of such AI-based tools and the novel capacities that AI grants to threat actors — defined as individuals or groups that intentionally harm digital devices or systems — seeking to carry out targeted cyberattacks against financial institutions.
3. To address these risks, Treasury recommends that financial institutions implement risk management procedures in line with the principles contained within "existing laws, regulations, and supervisory guidance."
4. Treasury also recommends that the industry and regulators work to create a common AI lexicon, expand the National Institute of Standards and Technology's AI Risk Management Framework ("NIST AI RMF") to more explicitly address the financial sector, support further research on algorithmic explainability, and address gaps in human capital.

On March 27, 2024, the US Department of the Treasury (Treasury) released a report entitled *Managing Artificial Intelligence-Specific Cybersecurity Risks in the Financial Services Sector*. Published in accordance with President Joe Biden's October 2023 AI executive order,[1] the report concerns "the current state of artificial intelligence (AI)–related cybersecurity and fraud risks in financial services, including an overview of current AI use cases, trends of threats and risks, best-practice recommendations, and challenges and opportunities."

In producing this report, Treasury conducted dozens of interviews with financial institutions of various sizes and market positions. This report considers both financial institutions' use of AI to detect fraud and the deployment of AI by threat actors seeking to commit fraud.

Financial Institutions' Use of AI in Fraud Detection

As the report acknowledges, financial institutions have been utilizing AI-powered fraud detection tools

“for more than a decade.” But recent advances in AI technology have led many financial institutions to either incorporate AI into existing threat detection tools or adopt new AI-based systems outright. “AI-driven tools are replacing or augmenting the legacy, signature-based threat detection cybersecurity approach of many financial institutions,” notes the report.

According to the financial institutions interviewed by Treasury, these AI-based cybersecurity tools have the “potential to significantly improve the quality and cost efficiencies of their cybersecurity and anti-fraud management functions” and “can also help institutions employ more proactive cybersecurity and fraud-prevention postures.”

Despite these potential benefits, the Treasury report expressed concern that smaller financial institutions’ relative lack of relevant expertise and data may lead these actors to unduly rely on third-party AI fraud detection tools, potentially to their detriment. “While smaller institutions may be able to access these tools through vendors,” noted the report, “internal development offers advantages in oversight and control of the development, testing, transparency, and governance of models and access to sufficient data monitoring for model risk management evaluation purposes.”

The report forecasts that “the resource requirements of AI systems will generally increase institutions’ direct and indirect reliance on third-party IT infrastructure and data. As a result, financial institutions should appropriately consider how to assess and manage the risks of an extended supply chain, including potentially heightened risks with data and data processing of a wide array of vendors, data brokers, and infrastructure providers.”

Furthermore, the report notes that in deploying AI fraud detection tools, financial firms are opening themselves up to a unique set of cybersecurity challenges. As compared with traditional fraud-detection solutions, AI tools present novel vulnerabilities “because of the dependency of an AI system on the data used to train and test it.” The report lists four such vulnerabilities that financial institutions should consider when implementing AI-based cybersecurity tools.

1. **Data Poisoning:** Corrupting an AI model’s training data to “impair the training process or gain a desired output of a model.”
2. **Data Leakage During Inference:** Securing confidential information from a model during its training process.
3. **Evasion:** Gaining a desired output from a model through strategically querying.
4. **Model Extraction:** Stealing an AI model wholesale by “iteratively querying the model.”

Threat Actors’ Use of AI to Commit Fraud

The second topic addressed by the Treasury report is threat actors’ use of AI to carry out targeted cyberattacks against financial institutions. In interviewing various financial institutions, Treasury found that market participants are concerned that with increasing access to AI, and especially generative AI tools, bad faith actors can more easily commit financial fraud.

“Concerns identified by financial institutions,” notes the report, “are mostly related to lowering the barrier to entry for attackers, increasing the sophistication and automation of attacks, and decreasing time-to-exploit. Generative AI can help existing threat actors develop and pilot more sophisticated malware, giving them complex attack capabilities previously available only to the most well-resourced actors.”

The report details four primary ways that cyberthreat actors can utilize AI against institutions with

sensitive data, financial or otherwise.

1. **Social Engineering:** Utilizing generative AI to facilitate “targeted phishing, business email compromise, and other fraud. Using generative AI systems, threat actors can more realistically misrepresent themselves as reflecting a variety of backgrounds, languages, statuses, and genders.”
2. **Malware/Code Generation:** Threat actors could use generative AI to rapidly develop malware code, such as “a false copy of a financial institution’s website entirely to harvest customers’ credentials.”
3. **Vulnerability Discovery:** Utilizing AI-based tools usually deployed for cyber defense, threat actors could discover vulnerabilities in a financial institution’s IT network.
4. **Disinformation:** Threat actors could pair a targeted cyberattack on a financial institution’s IT network with an AI-generated disinformation campaign to increase the attack’s efficacy.

AI Risk Management by Financial Institutions

Through their interviews with financial institutions, Treasury found that “existing risk management frameworks may not be adequate to cover emerging AI technologies,” and as such, “financial institutions appear to be moving slowly in adopting expansive use of emerging AI technologies.” To address this situation, the Treasury report provides recommendations and guidance to financial institutions seeking to responsibly adopt AI-based systems.

In Treasury’s view, technological advances “do not render existing risk management and compliance requirements or expectations inapplicable… Although existing laws, regulations, and supervisory guidance may not expressly address AI, the principles contained therein can help promote safe, sound, and fair implementation of AI.” As such, Treasury recommends that financial institutions “identify, monitor, and control risks arising from AI use as they would for the use of any other technology.”

One document that the report recommends that financial institutions consult in drafting their AI risk management strategies is the National Institute of Standards and Technology’s AI Risk Management Framework, “a resource to the organizations designing, developing, deploying, or using AI systems to help manage the many risks of AI and promote trustworthy and responsible development and use of AI systems.”

In line with the recommendations of the NIST AI RMF, the Treasury report suggests that financial institutions adopt AI tools in accordance with their enterprise risk tolerance. “The use case for AI systems should account for the risk tolerance associated with current Generative AI shortcomings,” asserts the report. “If a higher level of explainability is appropriate for a use case, Generative AI may not currently be a viable option. If a use case is intended to have anti-bias assurances, it may be appropriate to train AI models only on data that is prepared with anti-bias standards.”

Conclusion

The report ends by identifying 10 “next steps that can be taken by Treasury along with other agencies, regulators, and the private sector to address immediate AI-related cybersecurity and fraud risks for financial institutions.” These next steps include creating a common AI lexicon, expanding the NIST AI RMF to more explicitly address the financial sector, supporting research on algorithmic explainability, addressing gaps in human capital, and more.

The report and its next steps implicitly suggest that while Treasury does not oppose the financial sector's adoption of AI-based tools, the agency is attempting to ensure that market participants are aware of the attendant risks of such adoption and implement risk management procedures to minimize such risks.

This goal aligns with those of other financial services regulators, including the US Securities and Exchange Commission (SEC), which recently proposed rules to govern the use of AI and other predictive analytics technologies by broker-dealers and investment advisers. The SEC proposal met with substantial criticism due, in part, to a requirement that broker-dealers and advisers be able to fully audit, and eliminate, any conflicts of interest related to their use of AI technology. Meeting this requirement could, for example, substantially restrict the use of "black box" technologies for trading or other applications, including the use of technology based on large language models, which often lack perfect explainability.

Interestingly, the Treasury report stops short of requiring perfect explainability in all situations. It instead recommends that financial institutions "establish best practices for using Generative AI without explainability . . . [which] could include practices like ensuring good data hygiene for the data used to train the models and using the systems only when explainability is not necessary."

[1] The "Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence" states that "Within 150 days of the date of this order, the Secretary of the Treasury shall issue a public report on best practices for financial institutions to manage AI-specific cybersecurity risks." Please reference our "Timeline of Biden's AI Executive Order" for more information on the specific provisions of the AI EO.

Raj Gambhir contributed to this article.

National Law Review, Volume XIV, Number 95