Managing Cybersecurity Risks Arising from AI — New Guidance from the NYDFS

October 21, 2024

On October 16, 2024, the New York Department of Financial Services (the "NYDFS") issued an Industry Letter providing guidance on assessing cybersecurity risks associated with the use of AI (the "Guidance") under the existing 23 NYCRR Part 500 ("Part 500" or "Cybersecurity Regulation") framework. The Guidance applies to entities that are covered by Part 500 (i.e., entities with a license under the New York Banking Law, Insurance Law or Financial Services Law), but it provides valuable direction to all companies for managing the new cybersecurity risks associated with AI.

The NYDFS makes clear that the Guidance does not impose any new requirements beyond those already contained in the Cybersecurity Regulation. Instead, the Guidance is meant to explain how covered entities should use the Part 500 framework to address cybersecurity risks associated with AI and build controls to mitigate such risks. It also encourages companies to explore the potential cybersecurity benefits from integrating AI into cybersecurity tools (e.g., reviewing security logs and alerts, analyzing behavior, detecting anomalies, and predicting potential security threats). Entities that are covered by Part 500, especially those that have deployed AI in significant ways, should review the Guidance carefully, along with their current cybersecurity policies and controls, to see if any enhancements are appropriate.

In this Debevoise Data Strategy and Security blogpost, we summarize the key takeaways from the Guidance and discuss certain practical considerations that may be helpful to companies in assessing their cybersecurity measures for managing AI-related risks.

A. Cybersecurity-Related AI Risks

The Guidance divides cybersecurity-related AI risks into two categories: (1) risks caused by threat actors' use of AI; and (2) risks caused by companies' use of (or reliance on) AI.

Risks from Threat Actors' Use of AI

- **AI-Enabled Social Engineering:** This category covers increasingly sophisticated deepfakes that are being used effectively in social engineering attacks, including realistic and interactive audio, video, and text-based messages that target specific individuals via email (phishing), telephone (vishing), text (smishing), videoconferencing, and online postings. These attacks are designed to trick unsuspecting employees into sharing sensitive information or access credentials, or into transferring funds to accounts controlled by the attackers.

- **AI-Enhanced Cybersecurity Attacks:** AI allows threat actors to amplify the potency, scale, and speed of existing types of cyberattacks, and to quickly identify and exploit security vulnerabilities. AI can also accelerate the development of new malware and change existing ransomware so it can bypass defensive security controls. It also lowers the barrier to entry for new attackers, who can use AI to quickly learn how to launch successful attacks.

Risks from Companies' Use of AI

- **Exposure or Theft of Vast Amounts of Nonpublic Information:** Use of AI by companies will often involve the collection and processing of large volumes of nonpublic information, providing more opportunities for attackers and creating more data, devices, and locations for companies to protect. Additionally, the data used in AI applications sometimes contains biometric data, such as faceprints or fingerprints, which a threat actor can leverage to bypass MFA and gain access to additional information systems.

- **Increased Vulnerabilities Due to Third-Party, Vendor, and Other Supply Chain Dependencies:** Cyber-related AI risks are further compounded by companies' heavy reliance on third-party service providers (who are vulnerable to cyberattacks) to provide them with AI tools and/or the data used to train and operate them.

B. Controls and Measures That Mitigate AI-Related Threats

After listing several cybersecurity-related AI risks, the Guidance notes that Part 500 requires covered entities to assess risks and implement minimum cybersecurity standards designed to mitigate cybersecurity threats relevant to their businesses—including those posed by AI—and provides six examples of controls from Part 500 that will help companies reduce their AI-related cybersecurity risks.

Risk Assessments and Risk-Based Programs, Policies, Procedures, and Plans

- **Ensure risk assessments account for AI-specific risks.** The Guidance emphasizes the importance of incorporating AI-specific considerations in a covered entity's risk assessments and reminds covered entities to assess AI-related risks across their entire ecosystems, including for third-party service providers. When new AI risks are identified, covered entities should assess whether to update cybersecurity policies and procedures to mitigate those risks.

- **Conduct incident response planning, testing, and training to account for AI-related incidents.** The Guidance also signals the NYDFS's expectation that covered entities be prepared for cybersecurity incidents that can result from their use of AI, that such preparedness be tested, and that relevant personnel be appropriately informed about AI-related cybersecurity risks, including boards and senior leadership.

Third-Party Service Provider and Vendor Management

The NYDFS "strongly" recommends that due diligence of third-party service providers should include diligence on the AI-related risks they pose to themselves and to the covered entities. The Guidance reminds covered entities that they are required to impose minimum cybersecurity safeguards and mandatory cybersecurity incident-notification obligations on third-party service providers and encourages covered entities to consider obtaining additional representations and warranties relating to the secure use of the covered entities' nonpublic information.

Access Controls

Following its 2021 MFA Letter, and enhancements to the MFA requirements in the recent Part 500 Amendment, the Guidance reinforces the NYDFS's focus on MFA as a critical measure to combat cyberattacks. The Guidance calls on covered entities to consider avoiding SMS text, voice, or video for MFA, and instead utilize forms of authentication that AI deepfakes cannot impersonate, such as digital-based certificates and physical security keys. Similarly, instead of using a traditional fingerprint or other biometric authentication system, the Guidance encourages covered entities to use liveness detection or texture analysis to verify that a print or other biometric factor comes from a live person. The Guidance also reminds covered entities of their obligations under Part 500 to limit privileged access to only those necessary for that job function, to limit the number of privileged users, and to disable access privileges that are no longer necessary.

Cybersecurity Training

Under Part 500, covered entities are required to provide annual cybersecurity awareness training, including on social engineering. The Guidance also outlines specific additional topics to be covered for each of the following training groups.

- **All personnel should be trained on:**
  - the risks posed by AI;
  - procedures adopted by the organization to mitigate risks related to AI;
  - how to respond to AI-enhanced social engineering attacks;
  - what to do when personnel receive unusual requests, such as a request for credentials, an urgent money transfer, or access to NPI;
  - the need to verify a requestor's identity and the legitimacy of the request when an employee receives an unexpected money transfer request by telephone, video, or email; and
  - circumstances in which human review and oversight must be included in verification procedures.
- **Cybersecurity personnel should be trained on:**
  - how threat actors are using AI in social engineering attacks;
  - how AI is being used to facilitate and enhance existing types of cyberattacks; and
  - how AI can be used to improve cybersecurity.
- **For covered entities deploying AI, relevant personnel should also be trained on:**
  - how to secure and defend AI systems from cybersecurity attacks;
  - how to design and develop AI systems securely; and
  - how to draft queries to avoid disclosing NPI (as applicable).

Monitoring

The Guidance states that covered entities that use AI-enabled products or services should consider monitoring for unusual queries to AI systems that might indicate an attempt to extract NPI, as well as blocking queries from personnel that might expose NPI to a public AI product or system.

Data Management and Minimization

Given the large amounts of data that are often used to operate AI systems, the Guidance further underscores the need for covered entities to implement the data minimization practices required by Part 500 to dispose of nonpublic information that is no longer necessary for business operations or other legitimate business purposes. The Guidance also reminds covered entities of their obligation (as of November 2025) to maintain data inventories and encourages triaging inventories of information systems that rely on AI.

C. Practical Considerations

The Guidance highlights the NYDFS's expectation that cybersecurity risk management should cover any new cybersecurity risks presented by AI, so covered entities should consider reviewing their cybersecurity policies and procedures with that in mind. In doing so, as the NYDFS notes in the Guidance, it is important to remember that there are many other risks associated with AI adoption, such as loss of intellectual property, privacy, bias, transparency, explainability, quality control, loss of skills, conflicts, antitrust, and overselling. Accordingly, companies will need to decide how best to integrate their general cybersecurity risk management programs with their general AI risk management programs to make sure AI-related cybersecurity risks are properly addressed and don't fall through the cracks between the two. Here are a few tips for navigating that challenge:

Significant AI Tools, Vendors, and Use Cases Should Be Assessed for Cybersecurity Risks. There is no single optimal process for assessing AI-related cybersecurity risk. This can be done as the AI component of a general cybersecurity risk management program or the cybersecurity part of a general AI risk management program. It can also be achieved as the AI and cybersecurity components of more general software and vendor risks management programs. Leveraging their existing resources, controls, and risk management functions, companies will have to decide on the best way to ensure that

AI-related cybersecurity risks have been adequately identified and addressed. That process will likely involve some experimentation and pilot programs.

AI Governance Committees Membership. One way to ensure that cybersecurity risks for AI projects are properly addressed is to have a cross-functional committee, with a cybersecurity representative, that approves AI use cases and tools and also assists in the design of AI pilot programs.

Consider Specific Training on Deepfakes. Combatting deepfakes is largely a training issue. Companies can implement policies requiring verification, but if an employee honestly believes that they are being told to do something by the CEO, they are likely to do it, even if that action contravenes company protocols. Training should make clear, ideally with actual examples, that audio and video deepfakes can be extremely convincing, but any request to employees made by audio or video could be fraudulent, especially if it has the following characteristics: (a) it is unusual, (b) it involves the transfer of large sums of money or highly sensitive information, (c) it includes a requirement to keep the request confidential or not to follow normal protocols, (d) it has an element of urgency, or (e) it involves a transfer of funds to a new bank account or confidential information to an unfamiliar email address. Training should specifically note that employees will not face any adverse action for following company verification protocols when presented with such a request, and that failing to follow verification protocols, even at the request of the CEO, could result in discipline.

Consider Creating Model Diligence Questions and Contract Terms for AI Vendors. AI vendor diligence is a complicated process. Creating model diligence questions and contract terms will help standardize AI third-party risk management, but companies need to determine:

- Which kinds of AI vendors are covered (e.g., does the program apply to vendors who leverage AI on their own systems to provide goods and services to the company?);
- What counts as AI (e.g., does it apply to complex algorithms that do not involve machine learning but make important decisions or otherwise present significant reputational or regulatory risk?);
- How much of the program can be standardized (e.g., can diligence and contract terms for risks associated with IP and confidentiality apply to all AI vendors, while risks like bias or antitrust would only be addressed for certain vendors?); and
- Which risks are addressed through vendor-risk management and which are addressed separately through the internal use case approval process.

Working through these issues will increase the likelihood that cybersecurity-related risks posed by AI vendors are identified and addressed.

Monitor AI Use Cases in Production for Mission Creep. For AI use cases that have been approved subject to certain limitations (e.g., no use of confidential data, only one approved AI tool can be used, etc.) it is important to periodically check to ensure that the actual use of AI is consistent with those limitations, and that unanticipated cybersecurity (and other risks) have not materialized.

To subscribe to the Data Blog, please click here.

The Debevoise Data Portal is an online suite of tools that help our clients quickly assess their federal, state, and international breach notification and substantive cybersecurity obligations. Please contact us at dataportal@debevoise.com for more information.

The cover art used in this blog post was generated by DALL-E.

* * *

Please do not hesitate to contact us with any questions.

Charu A. Chandrasekhar
Partner, New York
+ 1 212 909 6774
cchandra@debevoise.com

Luke Dembosky
Partner, Washington, D.C.
+ 1 202 383 8020
ldembosky@debevoise.com

Avi Gesser
Partner, New York
+ 1 212 909 6577
agesser@debevoise.com

Erez Liebermann
Partner, New York
+ 1 212 909 6224
eliebermann@debevoise.com

Marshal Bozzo
Counsel, New York
+ 1 212 909 6797
mlbozzo@debevoise.com

Johanna Skrzypczyk
Counsel, New York
+1 212 909 6291
jnskrzypczyk@debevoise.com

Ned Terrace
Associate, New York
+ 1 212 909 7435
jkterrac@debevoise.com

Mengyi Xu
Associate, San Francisco
+ 1 415 738 5725
mxu@debevoise.com