Managing Artificial Intelligence-Specific Cybersecurity Risks in the Financial Services Sector

U.S. Department of the Treasury

March 2024

Executive Summary

In response to Executive Order (EO) 14110, *Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*, this report focuses on 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. The report's findings are based on 42 in-depth interviews conducted in late 2023. The interview participants include representatives from the financial services sector, information technology (IT) firms, data providers, and anti-fraud/anti-money laundering (AML) companies. The U.S. Department of the Treasury (Treasury) recognizes both the importance of responsible technological innovation across the financial sector and the increasing opportunities surrounding AI systems in the financial sector. Emerging technologies, however, often come with risks. While the focus of this report is on cybersecurity and fraud, Treasury also recognizes that the use of AI in financial services has important implications beyond these topics and will continue to study these implications.

Financial institutions have used AI systems in connection with their operations, and specifically to support their cybersecurity and anti-fraud operations, for years. Early adopters in the financial services sector may be revisiting AI's potential strategic value and considering new use cases in light of the recent rate of change and rapid developments in AI technology. Thus far, many financial institutions have incorporated AI-related risks into their existing risk management frameworks, especially those related to information-technology, model, compliance, and third-party risk management.

Some of the financial institutions that Treasury met with reported that existing risk management frameworks may not be adequate to cover emerging AI technologies, such as Generative AI, which emulates input data to generate synthetic content. Hence, financial institutions appear to be moving slowly in adopting expansive use of emerging AI technologies. Interview participants generally agreed that the safe adoption of AI technologies requires cross-enterprise collaboration among model, technology, legal, compliance, and other teams, which can present its own challenges. Further, in the case of cybersecurity and anti-fraud AI usage, interviewees largely agreed that effectively managing risks requires collaboration across the financial services sector. Applying appropriate risk management principles to AI development is critical from a cybersecurity perspective, as data poisoning, data leakage, and data integrity attacks can take place at any stage of the AI development and supply chain. AI systems are more vulnerable to these concerns than traditional software systems because of the dependency of an AI system on the data used to train and test it.

Like other critical infrastructure sectors, the financial services sector is increasingly subject to costly cybersecurity threats and cyber-enabled fraud. As access to advanced AI tools becomes more widespread, it is likely that, at least initially, cyberthreat actors utilizing emerging AI tools will have the advantage by outpacing and outnumbering their

targets. At the same time, many industry experts believe that most cyber risks exposed by AI tools or cyber threats related to AI tools can be managed like other IT systems. To counter the threat actors' initial advantage, financial institutions should expand and strengthen their risk management and cybersecurity practices to account for AI systems' advanced and novel capabilities, consider greater integration of AI solutions into their cybersecurity practices, and enhance collaboration, particularly threat information sharing. According to interviewed financial institutions, the adoption of AI technology, including Generative AI, has the potential to significantly improve the quality and cost efficiencies of their cybersecurity and anti-fraud management functions.

To an extent not seen with many other technology developments, technological advancements with AI are dependent on data. In most cases, the quality and quantity of data used for training, testing, and refining an AI model, including those used for cybersecurity and fraud detection, directly impact its eventual precision and efficiency. Interview participants indicated general agreement that more collaboration results in better cyber protection. It is becoming more common for institutions to share anonymized cybersecurity information with vendors to enhance their anomaly detection AI models used for cyber defense functions such as intrusion detection across the vendor's customer base. There are also well-established standards, frameworks, and apparatuses for sharing cyberthreat information, including the Financial Services Information Sharing and Analysis Center (FS-ISAC), launched in 1999.

Collaboration in the fraud-protection space, however, appears to be less coordinated than for cyber protection. Except for certain efforts in banking, there is limited sharing of fraud information among financial firms. A clearinghouse for fraud data that allows rapid sharing of data and can support financial institutions of all sizes is currently not available. The absence of fraud-related data sharing likely affects smaller institutions more significantly than larger institutions. With their broader set of client relationships, large firms have a wider base of historical fraudulent activity data they can use to develop fraud-detection AI models. For example, one large firm noted that it developed AI models trained completely on the firm's own internal historical data, which enabled it to reduce fraud activity by an estimated 50%. Fraud activity blocked by such models would likely shift to more vulnerable corners of the sector like smaller institutions that have neither enough data to replicate the larger firms' base data nor the resources to create the systems needed to digest the necessary data.

The Bank Policy Institute (BPI) and the American Bankers Association (ABA) are both making efforts to close the fraud information-sharing gap across the banking sector. The ABA's initiative is specifically aimed at closing the fraud data gap for smaller financial institutions. Treasury's Financial Crimes Enforcement Network (FinCEN) and core providers might also be well positioned to play a critical role in supporting efforts to ensure that smaller financial institutions are benefitting from the advancements in AI technology development for countering fraud. If smaller financial institutions are not supported in closing this critical

gap, several interviewees indicated that there may be a risk of future consolidation towards larger institutions which are better equipped with both the data and technical talent able to leverage these tools to counter adversary advancements.

Additionally, the importance of data for AI technology and the complexity of AI technology development would very likely increase financial institutions' reliance on third-party providers of data and technology. As a result, it is very likely that often overlooked third-party risk considerations such as data integrity and data provenance will emerge as significant concerns for third-party risk management. Emerging AI solutions may challenge traditional expectations regarding financial institutions' ownership of data, models, and insights. Additionally, the current trend of adopting AI solutions through multiple intermediaries and service providers complicates oversight and transparency. It is becoming increasingly challenging to accurately understand data flows and the use of AI solutions, thus inhibiting understanding and verification of those AI systems' fidelity of insights and decision making.

Table of Contents

1. Introduction

1.1 BACKGROUND AND PURPOSE OF THIS REPORT

This report fulfills the requirement in EO 14110 on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence 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.*

This report provides an overview of the state of AI use in the financial services sector for cybersecurity purposes and discusses its implications for financial institutions based on 42 interviews with industry stakeholders. While this report is focused on the use of AI in cybersecurity and the risks that are attendant to that use, many of the findings and best practices may have applicability to other AI use cases. This report does not address many other issues related to AI and financial services, including those related to consumer and investor protection, disparate impact, financial stability, and financial regulatory questions. Treasury expects to continue to study the impact of AI on financial services in the coming months and years. The observations reflect the participating stakeholders' perception of the state of AI and is not intended as an authoritative assessment of actual AI usage or terminology. Potential follow-on workstreams will address aligning perceptions toward a common understanding of the current state and enhancements to AI cybersecurity and risks.

1.2 REPORT ORGANIZATION

This report is composed of five main components, beginning with *Artificial Intelligence in Financial Services: Cybersecurity and Fraud Protection*. That section provides an overview of how financial institutions of all sizes are using AI, focusing specifically on cybersecurity and anti-fraud use cases. Next, the report outlines *AI Cybersecurity and Fraud Threats for Financial Institutions*. The following section, *Regulatory Landscape for Artificial Intelligence in Financial Services*, offers an overview of the current regulatory environment regarding

the use of AI in cybersecurity and fraud management by financial services firms. The report then outlines *Best Practices for Managing AI-Specific Cybersecurity Risks.* Finally, the report concludes with recommendations for addressing *Next Steps: Challenges & Opportunities* that Treasury identified while developing this report and offers a way forward on the most pressing issues.

1.3 SCOPE AND METHODOLOGY

The focus of this report is on the best practices for financial institutions to manage AI-specific cybersecurity risks as directed by the EO. The report also reviews anti-fraud efforts because the use of AI by financial institutions to identify fraud offers important lessons in managing AI cybersecurity risk. Additionally, cybersecurity and anti-fraud efforts are becoming ever more tightly coupled as fraud becomes increasingly enabled by cybersecurity gaps.

Following the release of Treasury's report *Financial Services Sector's Adoption of Cloud Services*[1] in February 2023, Treasury launched a public-private partnership dedicated to bolstering regulatory and private sector cooperation.[2] This partnership—the Cloud Executive Steering Group (CESG)—is chaired by leaders in the financial sector with expertise in financial sector cybersecurity. The CESG provides a forum for convening financial sector AI stakeholders across the member agencies of the Financial Stability Oversight Council (FSOC),[3] the Financial and Banking Information Infrastructure Committee (FBIIC),[4] and the Financial Services Sector Coordinating Council (FSSCC).[5]

To develop this report, Treasury, with assistance from CESG's private sector members and several FSSCC committees, met with a broad array of organizations and financial firms to learn from practitioners about the use and implications of AI in the financial services sector. This included 42 in-depth discussions with representatives from financial institutions of all sizes, financial sector trade associations, cybersecurity and anti-fraud service providers that include AI features in their products and services, consulting firms assisting financial institutions in the development of AI, regulatory advocacy groups,

1 U.S. Department of the Treasury, *The Financial Service Sector's Adoption of Cloud Services* (Feb. 2023), https://home.treasury.gov/system/files/136/Treasury-Cloud-Report.pdf.

2 U.S. Department of the Treasury, *U.S. Department of the Treasury Kicks Off Public-Private Executive Steering Group to Address Cloud Report Recommendations* (May 25, 2023), https://home.treasury.gov/news/press-releases/jy1503.

3 Established in 2010 under the Dodd-Frank Wall Street Reform and Consumer Protection Act, the FSOC's mission is to identify risks to U.S. financial stability, promote market discipline, and respond to emerging threats to the stability of the U.S. financial system. For more about FSOC see https://www.fsoc.gov.

4 Chartered under the President's Working Group on Financial Markets, the FBIIC is charged with improving coordination and communication among financial regulators, promoting public-private partnerships within the financial sector, and enhancing the resiliency of the financial sector infrastructure overall. For more about the FBIIC see https://www.fbiic.gov/.

5 Established in 2002, the FSSCC is an industry-led, non-profit organization that coordinates critical infrastructure and homeland security activities within the financial services industry. FSSCC members consist of financial trade associations, financial utilities, and the most critical financial firms. For more about the FSSCC see https://fsscc.org/.

payment service providers, financial services technology service providers, and general technology companies focused on AI products and services. This outreach was focused on the state of AI use by financial sector firms for cybersecurity and fraud protection, viewpoints on the risks posed by AI, and best practices for managing risks related to the use of AI.

In preparing this report, Treasury also considered a report published by FSSCC in February 2024 entitled *Artificial Intelligence in the Financial Sector: Cybersecurity and Fraud Use Cases and Risks.* The FSSCC report is included as an annex to this report. The report provided Treasury with the FSSCC's perspective on cybersecurity and anti-fraud AI-use cases, AI-specific risks, and considerations for approaching ongoing AI-induced changes in the financial sector. In addition to publishing the report, FSSCC's Research and Development (R&D) Committee invited Treasury to participate in the discussions the Committee organized with experts from across the FSSCC to gather information about the application of cybersecurity and anti-fraud use cases across the financial sector.

1.4 WHAT DO WE MEAN BY ARTIFICIAL INTELLIGENCE?

This report uses the definition of AI set forth in the EO:

> The term "artificial intelligence" or "AI" has the meaning set forth in 15 U.S.C. 9401(3): a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. Artificial intelligence systems use machine- and human-based inputs to perceive real and virtual environments; abstract such perceptions into models through analysis in an automated manner; and use model inference to formulate options for information or action.

Treasury found that there is no uniform agreement among the participants in the study on the meaning of "artificial intelligence." While the above definition is broad, it may still not cover all the different concepts associated with the term "artificial intelligence." In fact, one participant said they do not use "artificial intelligence" and instead prefer to call these systems "augmented intelligence" to emphasize that users of these systems still need to be the ultimate decision-makers. Participants also expressed frustration with understanding various AI product and service offerings because vendors described their AI systems in ways that do not always align with how financial institutions understand them to work, especially for Generative AI.

Recent commentary around advancements in AI technology often uses "artificial intelligence" interchangeably with "Generative AI." This report considers Generative AI as a subset of AI and will explicitly differentiate between the two, using the definition of Generative AI in the EO[6]:

6 Generative is capitalized throughout the report to help clearly distinguish between different AI technologies.

The term "Generative AI" means the class of AI models that emulate the structure and characteristics of input data in order to generate derived synthetic content. This can include images, videos, audio, text, and other digital content.

1.5 FINANCIAL SERVICES SECTOR PROFILE

The financial services sector is highly diverse and includes thousands of financial institutions, including depository institutions, providers of investment products, insurance companies, other credit and financing organizations, and the providers of the critical financial utilities and services that support these functions. Financial institutions vary widely in size and presence, ranging from some of the world's largest global companies with hundreds of thousands of employees and trillions of dollars in assets, to community banks and credit unions with a small number of employees serving individual communities.[7]

Financial institutions are generally organized and regulated based on the services that the institutions provide. Collectively, these organizations form the backbone of the nation's financial system and are a vital component of the global economy. These organizations are tied together through a network of electronic systems with innumerable entry points. Financial institutions face an evolving and dynamic set of risks, including operational, liquidity, credit, legal, and reputational risks. Each financial institution has unique security and resilience needs, resources, and plans depending on the functions it performs and its approach to risk management.[8]

1.6 CYBERSECURITY AND FRAUD TRENDS

The number of cybersecurity incidents continues to increase each year as various cyber threats, from ransomware to data theft, remain significant challenges to organizations of all kinds.[9] The costs of these incidents also continue to rise every year. According to IBM, the average cost of a data breach reached an all-time high of $4.45 million in 2023.[10] Interestingly, participants in the same IBM study that used AI and other automated technologies to detect risk reported lower costs associated with data breaches and a shorter timeframe for detecting an incident.[11]

Losses from fraud also continue to rise every year. According to Juniper Research, online payment fraud is expected to cumulatively surpass $362 billion by 2028.[12] Separately,

7 See U.S. Department of the Treasury, U.S. Department of Homeland Security, and FSSCC, *Financial Services Sector-Specific Plan* (2015), https://www.cisa.gov/sites/default/files/publications/nipp-ssp-financial-services-2015-508.pdf.

8 *Ibid*.

9 Verizon, *2023 Data Breach Investigations Report* (Jun. 2023), https://www.verizon.com/business/resources/Te7c/reports/2023-data-breach-investigations-report-dbir.pdf.

10 IBM, *Cost of Data Breach Report 2023* (Jul. 2023), https://www.ibm.com/downloads/cas/E3G5JMBP.

11 *Ibid*.

12 Juniper Research, *Online Payment Fraud: Market Forecasts, Emerging Threats & Segment Analysis 2023-2028* (Jun. 2023), https://www.juniperresearch.com/research/fintech-payments/fraud-identity/online-payment-fraud-research-report/.

the Federal Bureau of Investigation reports that cumulative losses from business email compromises reached over $50 billion at the end of 2022.[13] Verizon reports that personal information, which is often used to commit fraud, remains the most common type of information stolen from financial institutions during a data breach.[14] Synthetic identity fraud, which involves fraudsters leveraging the personally identifiable information of individuals, costs financial institutions more than $6 billion per year.[15] Unfortunately, these are just some of the most well-known types of fraud. Costs of other types of fraud have not been fully estimated.

13 FBI, *Business Email Compromise: The $50 Billion Scam* (Jun. 2023), https://www.ic3.gov/Media/Y2023/PSA230609.

14 Verizon, *2023 Data Breach Investigations Report* (Jun. 2023), https://www.verizon.com/business/resources/Te7c/reports/2023-data-breach-investigations-report-dbir.pdf.

15 The Federal Reserve FedPayments Improvement, *Payments Fraud Insights July 2023: Mitigating Synthetic Identity Fraud in the U.S. Payment System* (Jul. 2023), https://fedpaymentsimprovement.org/wp-content/uploads/frs-synthetic-identity-payments-fraud-white-paper-july-2020.pdf.

2. Artificial Intelligence in Financial Services: Cybersecurity and Fraud Protection

2.1 AI SYSTEMS AND MITIGATION OF CYBERSECURITY AND FRAUD RISKS

Many financial institutions reported that they currently use AI systems in a wide variety of operations, and some are evaluating or piloting Generative AI-based tools to support employee efficiency with research and report-writing tasks. While most financial institutions have reported that they have used AI systems for years, maturity in utilization and deployment of AI systems varies by institution and continues to evolve. In particular, AI tools for fraud detection, including machine learning (ML)-based tools, have been used by a wide range of financial institutions as part of risk management strategies for more than a decade, as reported by interview participants.

Some financial institutions reported that they started using AI for cybersecurity several years ago. Various types of cybersecurity tools that financial institutions typically use to mitigate cybersecurity risks now incorporate AI, making the institutions reportedly more agile than they were in the past. Financial institutions provided examples of incorporating advanced anomaly-detection and behavior-analysis AI methods into existing endpoint protection, intrusion detection/prevention, data-loss prevention, and firewall tools. AI-driven tools are replacing or augmenting the legacy, signature-based threat detection cybersecurity approach of many financial institutions. AI tools can help detect malicious activity that manifests without a specific, known signature. This capability has become critical in the face of more sophisticated, dynamic cyberthreats that may leverage legitimate system administration tools, for example, to avoid triggering signature detection.

According to interview participants representing financial institutions, their increasing adoption of AI, including Generative AI, has the potential to significantly improve the quality and cost efficiencies of their cybersecurity and anti-fraud management functions. Many firms rely heavily on automation for time-consuming and labor-intensive anti-fraud and cybersecurity-related tasks; AI and Generative AI can augment those processes by capturing and processing broader and deeper data sets and utilizing more sophisticated analytics. These technologies can also help institutions employ more proactive cybersecurity and fraud-prevention postures. For example, Generative AI could be used to provide opportunities for educating employees and customers about cybersecurity and fraud detection and prevention measures or for analyzing internal policy documents and communications to identify and prioritize gaps in those measures.

Despite views on its potential, participants in our outreach stated that they are taking a cautious and risk-based approach to integrating Generative AI into their cybersecurity and anti-fraud operations. Most participants representing firms reported that they are proceeding with caution on Generative AI and are trying to address Generative AI

risks by providing guardrails and developing internal policies for the acceptable use of this technology. In general, financial institution representatives stated that they are implementing processes to manage risks posed by emerging and critical technologies as part of their financial, operational, reputational, or legal risk management.

Specifically, many financial institution representatives believe that their existing practices align with aspects of the National Institute of Standards and Technology (NIST) AI Risk Management Framework (RMF), which was released in January 2023. However, while interview participants representing financial institutions stated that they have implemented the core elements of the NIST RMF (govern, map, measure, manage), many also noted that it is challenging to establish practical and enterprise-wide policies and controls for emerging technologies like Generative AI. Discussion participants noted that while their risk management programs should map and measure the distinctive risks presented by technologies such as large language models (LLM), these technologies are new and can be challenging to evaluate, benchmark, and assess in terms of their cybersecurity.

2.2 MIXED USE OF IN-HOUSE AND THIRD-PARTY AI SYSTEMS

Treasury's discussions with financial institutions revealed that their use of in-house or third-party AI systems varies significantly by institutional size. Many institutions that Treasury interviewed—regardless of size—indicated that they engaged a third-party service or product provider for at least some of their cybersecurity or fraud-related AI systems. Many noted that some of the capabilities they use are supplied by a fourth-party provider or product (in other words, that their vendors are using their own vendors that may be providing the AI). Global systemically important banks (G-SIBs) and many midsize financial institutions, particularly in the cybersecurity and anti-fraud context, also reported using a combination of AI models developed internally based on proprietary data or acquired commercially and fine-tuned with proprietary data. Some G-SIBs stated that they employ hundreds of AI system developers and have been able to significantly reduce fraudulent activity. In some cases, financial institutions acquire an AI model from a third-party entity which is developed based on sample or anonymized data provided by the financial institution.

Differences in financial institutions' proprietary data and adoption of cloud services will likely influence the adoption of AI by each institution. Most representatives of financial institutions stated that they are trying to leverage their own data for use cases that are unique to their firm while relying on vendors for more general applications. As computing capacity and analytic tools become more readily available through cloud service providers, more institutions may develop their own models. Institutions that have already adopted cloud services, or those with large amounts of proprietary data, will likely be able to take advantage of these AI tools sooner than those that have not. In particular, institutions that have already migrated some of their systems and data into cloud computing platforms will likely be able to take advantage of these developments sooner

than others because their data will already be available for training or processing by AI in the cloud. While smaller institutions may be able to access these tools through vendors, 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.

2.3 CAUTIOUS ADOPTION OF GENERATIVE AI SYSTEMS

Discussion participants representing financial institutions generally acknowledged that Generative AI models are still developing, currently very costly to implement, and very difficult to validate for high-assurance applications. As a result, many firms are instead prioritizing low-risk, high-return use cases, such as code-generating assistant tools for imminent deployment. Aside from some small institutions that seem to not be using Generative AI at this time, most interviewed financial firms have not used public access or public application programming interface (API) offerings of Generative AI providers and have instead opted for an enterprise solution deployed in their own virtual cloud network, tenant, or multi-tenant. However, for a more tailored approach, several institutions reported using the retrieval augmented generation (RAG) method, where an open- or closed-source foundation model is selected, and the institution's proprietary data is then converted and stored in a vector database and used to enrich the prompt to provide more relevant and accurate responses. This technique may be combined with fine-tuning of the open model with non-sensitive proprietary data.

2.4 CYBERSECURITY AND RISKS RELATED TO AI MODELS[16]

Financial institutions' primary cybersecurity goal generally remains the protection of their important assets: their personnel and customers, and their data, including internal information and information received from counterparties. Regulatory requirements and guidance provide a framework for institutions to implement controls to mitigate AI-related cybersecurity risks. However, more advanced AI technologies, such as Generative AI, may require institutions to extend these controls or adopt new ones.

Data poisoning, data leakage, and data integrity attacks can occur at any stage of the AI development and supply chain. AI systems are more vulnerable to these concerns than traditional software systems because of the dependency of an AI system on the data used

16 Other issues specifically implicated by AI, but not necessarily relevant to cybersecurity and anti-fraud use cases, include biases, unethical uses, and false outputs that are becoming more noteworthy with the increasing use of foundational models such as LLMs and Generative AI. This report generally does not address these challenges. Financial-sector firms may be able to address these kinds of issues by leveraging model risk management (MRM) practices common to the banking industry. MRM may help manage the risks surrounding the use of these models by highlighting the need for transparency and risk management of the models themselves. However, it remains to be seen whether additional guidance or risk management practices are needed to manage these types of risks. Additionally, financial institutions that participated in our study reported keeping humans involved in the development of AI models and eventual applicability of AI models' outputs as they work to resolve model-risk concerns. How any of these issues manifest for institutions, however, largely depends on how AI systems are deployed. Some AI systems, for example, may not be appropriate for high-assurance applications that require explanation of the outputs to ensure repeatable, unbiased decision making.

to train and test it. Data ingested by an AI system in training or even in testing can directly inform the production processing of the AI system. Source data, training datasets, testing datasets, pre-trained AI models, LLMs themselves, prompts, and prompt and vector stores can all be subject to data attacks, making the security of data throughout the development and production cycle as important as protecting production data.

2.5 FRAUD DETECTION AND THE IMPACT OF DATA

As highlighted in Treasury's report *Assessing the Impact of New Entrant Non-bank Firms on Competition in Consumer Finance Markets*,[17] there is an ongoing evolution of the consumer finance value chain due to an increase in consumer demand for digital financial services and the widespread availability of new financial services-related digital technologies, infrastructure services, and consumer data. While these rapid developments can benefit businesses and consumers, they have also brought about an increase in the number and sophistication of fraudulent activities.

Currently, firms do not share fraud data with each other to the extent that would be needed to train anti-fraud AI models. Similar to cybersecurity tools, fraud detection and prevention technologies have evolved from traditional rule-based engines, deny lists, and device fingerprints to more advanced ML-based systems. However, the accuracy of ML-based systems in identifying and modeling fraudulent behavioral patterns correlates directly with the scale, scope (variety of datasets), and quality of data available to firms. Unlike data on cybersecurity, there is reportedly little fraud information sharing across the financial sector, which limits the ability to aggregate fraud data for use by AI systems. Most financial institutions Treasury spoke with expressed the need for better collaboration in this domain, particularly as fraudsters themselves have been using AI and ML technologies. Sharing of fraud data would support the development of sophisticated fraud detection tools and better identification of emerging trends or risks.

However, while collecting, analyzing, and sharing fraud data can improve detection, it also raises privacy concerns that need to be managed through robust data protection and privacy practices. The collection, storage, and processing of sensitive financial information or other personal data poses risks. Transaction histories and personal behaviors are examples of sensitive financial information that may be used as inputs in AI systems. Although beyond the scope of this report, financial sector representatives have stated that the historical data used to train fraud-detection models could contain biases, such as the overrepresentation of certain demographics in anti-fraud cases.[18] Participating firms noted that data-anonymization techniques and algorithmic transparency could help to mitigate some of these issues.

17 U.S. Department of the Treasury, *Assessing the Impact of New Entrant Non-bank Firms on Competition in Consumer Finance Markets* (Nov. 2022), https://home.treasury.gov/system/files/136/Assessing-the-Impact-of-New-Entrant-Nonbank-Firms.pdf.

18 Chen, Changye, et al, *Ethical perspectives on AI hazards to humans: A review* (Dec. 2023), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10695628/.

3. AI Cybersecurity and Fraud Threats for Financial Institutions

Complex and persistent cyber threats continue to grow, and some experts from financial institutions believe the availability of advanced AI tools such as Generative AI will, at least initially, give threat actors an upper hand. However, many also believe that most of the risks related to AI tools do not go beyond traditional software vulnerabilities and the adversary tactics and techniques associated with them. Concerns identified by financial institutions 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. It can also help less-skilled threat actors to develop simple but effective attacks.

As advanced AI capabilities become more commonly available, customers, employees, processes, and IT systems of financial institutions can be targeted with malicious attacks using AI. Effective attacks against AI systems themselves are also expanding and can target different stages of an AI system's life cycle, including design, implementation, training, testing, evaluation, and deployment. Additionally, AI system dependency on data may amplify existing challenges and introduce new data security and privacy challenges for institutions, including those related to their third-party providers and their software and data supply chains. Most representatives from financial institutions stated that while there is a range of cyber threats related to AI and some cyberthreat actors may have experimented such attacks, most successful real-world attacks so far have been primarily related to social engineering and identity spoofing.

3.1 CYBERTHREAT ACTOR USES OF AI

Ways in which cyberthreat actors can use AI include:

- **Social Engineering**: Financial institutions have cited an increase in the scope and scale of sophisticated social-engineering techniques. Threat actors can utilize LLMs to conduct more 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. These attacks can further be tailored for high-value customers using other data points, such as social media posts or messages to enhance their mimicry.
- **Malware/Code Generation**: Threat actors using Generative AI might be able to develop new malware code or a new variant of an existing malware more quickly. Slightly modified malware will have a higher chance of evading an automated signature-based detection system. Many participants stated that the impact of these developments will, at least initially, be limited to exploitation of less-sophisticated

malware. One example is using Generative AI to create a false copy of a financial institution's website entirely to harvest customers' credentials. Threat actors can also use Generative AI products that generate code to develop, or refine code, to automate their exploitation tools. Such development could reduce the time from vulnerability discovery to exploitation.

- **Vulnerability Discovery:** Threat actors can use advanced AI-based tools that are typically used for cyber defense by developers and testers to discover vulnerabilities and identify weaknesses in an institution's IT network and application security measures. This potential reduction in vulnerability discovery time, combined with potential reduction in exploitation time, could give cyber threat actors an upper hand against cybersecurity practitioners by outpacing patching or intrusion detection.
- **Disinformation:** Threat actors can increase a targeted attack's efficiency by conducting parallel disinformation campaigns. Using Generative AI technology, which is becoming increasingly compelling in conveying human language characteristics and personalities, threat actors could spread disinformation more broadly and with higher rates of consumption. Threat actors may also incorporate different forms of AI-generated content, such as images, audio, and videos, to enhance their malicious campaign's influence. This report further explores the risks of deepfakes and synthetic identity in a later subsection.

The few major providers of Generative AI models today have stated that they have implemented safeguards to prevent malicious use of their products. However, there are documented instances of researchers bypassing these safeguards through carefully crafted prompt engineering techniques.[19]

3.2 CYBERTHREATS TO AI SYSTEMS

A few of the financial institutions Treasury interviewed indicated that they have started examining the types of cybersecurity exploits unique to AI and to model those threats using references such as NIST's Adversarial Machine Learning technical publication.[20] Relatedly, AI systems are not immune to insider threats. Trusted insiders are more likely to have direct access to AI systems and may be able to harm those systems more readily. Cyberthreats to AI systems mainly fall under the following categories:

- **Data Poisoning:** Threat actors corrupt the training data or model weights directly to impair the training process or gain a desired output of a model.

19 Zou, A., et al, *Universal and transferable adversarial attacks on aligned language models* (Dec. 2023), https://arxiv.org/abs/2307.15043.

20 Vassilev, A., Oprea, A., Fordyce, A. and Andersen, H., *Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations* (Jan. 2024), NIST Trustworthy and Responsible AI, National Institute of Standards and Technology, https://www.nist.gov/publications/adversarial-machine-learning-taxonomy-and-terminology-attacks-and-mitigations.

- **Data Leakage During Inference:** Threat actors gain access to confidential data through model inversion and programmatic query of the model during the inference phase.
- **Evasion:** Threat actors corrupt a model to gain a desired output by providing a strategic query or input to the model.
- **Model Extraction:** Threat actors steal an AI model itself by constructing a functionally equivalent model via iteratively querying the model.

3.3 IDENTITY IMPERSONATION AND SYNTHETIC IDENTITY

AI allows bad actors to impersonate individuals, such as employees and customers of financial institutions, in ways that were previously much more difficult. Although deepfakes are not new, AI that mimics human features has advanced in recent years to become much more believable.[21] Fraudsters can use AI to mimic voice, video, and other behavioral identity factors that financial institutions use to verify a customer's identity. Financial institution representatives that participated in our discussions consistently identified this kind of sophisticated identity impersonation as a chief concern about the malicious use of AI.

Fraudsters have previously used AI to impersonate people. In one of the earliest documented cases, fraudsters reportedly used audio generated by AI to impersonate a company's chief executive officer (CEO).[22] The fraudsters allegedly used the CEO's fake voice to ask a subsidiary company in the United Kingdom (UK) to transfer money to a supplier for a loss of nearly $250,000. The fraudsters subsequently attempted to trick the UK subsidiary into transferring more money but could not convince the victims to do so again. Fraudsters reportedly used a similar scheme in 2024 to trick an employee in Hong Kong into transferring $25 million by posing as the company's chief financial officer.[23]

Voice, video, or other identity-verification mechanisms like measuring keystroke dynamics are all susceptible to AI-enabled impersonation. Depending on the circumstances, fraudsters may be readily able to impersonate any of these identity factors with AI technology available today. It appears that even live video interactions with a known client may be no longer sufficient for identity verification because of advances in AI-driven video-generation technology.

Separately, AI may enhance fraudsters' abilities to create synthetic identities. Synthetic identities represent fake people, usually created using a composite of personal information

21 Brooks, Tina, et al. *Increasing Threat of Deepfake Identities* (2021), https://www.dhs.gov/sites/default/files/publications/increasing_threats_of_deepfake_identities_0.pdf.

22 Trend Micro, *Unusual CEO Fraud via Deepfake Audis Steals US$243,000 From UK Company* (Sep. 2019), https://www.trendmicro.com/vinfo/mx/security/news/cyber-attacks/unusual-ceo-fraud-via-deepfake-audio-steals-us-243-000-from-u-k-company.

23 Chen, Heather and Magramo, Kathleen, *Finance worker pays out $25 million after video call with deepfake 'chief financial officer'* (Feb. 2024), CNN, https://www.cnn.com/2024/02/04/asia/deepfake-cfo-scam-hong-kong-intl-hnk/index.html.

from various real people, that fraudsters can use to conduct a range of financial activities, like opening accounts and lines of credit at financial institutions. These synthetic identities typically exhibit normal financial behaviors over time to appear as if they are real, even going as far as to develop fake credit profiles so that fraudsters can use the identities to secure large loans and then disappear. This kind of fraud is difficult to measure, but "[o]ne widely reported analysis by Auriemma Group suggested that synthetic identity fraud cost U.S. lenders $6 billion and accounted for 20% of credit losses in 2016."[24]

A recent report by Wakefield Research commissioned by Deduce describes an acceleration in the growth of synthetic identity fraud.[25] Wakefield's findings indicate that 87% of the companies they surveyed had extended credit to synthetic customers and that synthetic identity fraud increased by 17% over the two years prior to October 2023.[26] The companies participating in the Wakefield study offer credit in such forms as personal loans and credit cards but not mortgages. Notably, 61% of companies in that study extending credit to those using synthetic identities proactively offered credit to the fraudsters, highlighting the challenge of discerning real people.[27] While the researchers were not able to explicitly tie the increase to fraudsters' use of AI, Generative AI may allow fraudsters to create more sophisticated synthetic identities, exacerbating an already challenging problem, contributing to an increase in synthetic identity fraud.

FinCEN's recent Financial Trend Analysis (FTA) on identity-related suspicious activity found that approximately 42% of the 3.8 million Bank Secrecy Act reports filed in 2021, equivalent to $212 billion in reported activity, was related to identity.[28] The volume of these types of exploitations or cyber-enabled attacks is likely to rise as technological developments like Generative AI reduce the cost, complexity, and time required to leverage gaps in our digital infrastructure.

3.4 THIRD-PARTY RISKS AND MAGNIFIED DATA SECURITY AND PRIVACY

Whether a financial institution decides to develop an AI solution in-house or acquire one through a vendor, 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.

24 The Federal Reserve FedPayments Improvement, *Payments Fraud Insights July 2023: Mitigating Synthetic Identity Fraud in the U.S. Payment System* (Jul. 2023), https://fedpaymentsimprovement.org/wp-content/uploads/frs-synthetic-identity-payments-fraud-white-paper-july-2020.pdf.

25 Wakefield, *Financial Jeopardy: Companies Losing Fight Against Synthetic Fraud* (Oct. 2023), https://www.deduce.com/resource/wakefield-research-report/.

26 *Ibid*.

27 *Ibid*.

28 FinCEN Financial Trend Analysis, *Identity-Related Suspicious Activity: 2021 Threats and Trends* (Jan. 2022), https://www.fincen.gov/sites/default/files/shared/FTA_Identity_Final508.pdf.

Multiple factors—including whether there are proprietary data or licenses associated with the training data, how the data is handled, how the data was gathered, prepared, and tagged, and the quality of training data—can expose institutions to financial, legal, and security risks. Beyond legal compliance, interview participants point to a growing demand for ethical considerations in AI development and deployment. This includes ensuring AI systems are fair, transparent, and accountable, and do not perpetuate biases or discrimination.

4. Regulatory Landscape for Artificial Intelligence in Financial Services

Throughout the interview process, financial institutions emphasized the collaborative and cooperative nature of their engagements with regulators. Firms mostly noted clear expectations, non-prescriptive solutions, and an openness to collaborative dialogue. Financial institutions stated their interest to remain robustly engaged in the regulatory conversation. Some participants reported that regulators are focused on institutions' enterprise risk frameworks for handling AI and are exploring how AI fits into existing risk management practices for cybersecurity and anti-fraud efforts. Several stakeholders highlighted that smaller institutions may benefit from more granular guidance, but larger entities typically preferred a more flexible approach at this stage. As discussed later in this report, participants emphasized the importance of establishing a common lexicon to improve their communication with regulators.

4.1 U.S. FINANCIAL SECTOR REGULATORY LANDSCAPE

Financial regulatory agencies generally do not issue regulations or guidance on specific technologies, but instead address the importance of effective risk management, governance, and controls regarding the use of technology, including AI, and the business activities that those technologies support. Regulators have emphasized that it is important that financial institutions and critical infrastructure organizations manage the use of AI in a safe, sound, and fair manner, in accordance with applicable laws and regulations, including those related to consumer and investor protection. Controls and oversight over the use of AI should be commensurate with the risk of the business processes supported by AI. Regulators have noted that it is important for financial institutions to identify, measure, monitor, and manage risks arising from the use of AI, as they would for the use of any other technology. Advances in technology do not render existing risk management and compliance requirements or expectations inapplicable.

Various existing laws, regulations, and supervisory guidance are applicable to financial institutions' use of AI. 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. The focus of this report is on cybersecurity and fraud protection, as they relate to AI and financial services, and thus, this report does not comprehensively list the regulatory frameworks and issues that could potentially apply to the use of AI in financial services. This report also does not weigh in on whether there are regulatory gaps or issues that may necessitate additional legislation, regulation, or guidance.

Key examples of risk management and control principles common across financial sector laws, regulations, and supervisory guidance that are applicable to the use of AI regarding cybersecurity and fraud issues include:

- **Risk Management:** Many financial regulatory agencies have rules and guidance addressing standards for risk management and information security,[29] including heightened standards for certain institutions[30] and financial market utilities.[31] Effective risk management processes for the introduction of new technologies and business activities, such as AI-supported activities, help financial institutions to effectively identify, measure, monitor, and control the risks associated with these new activities.[32] Effective risk management over emerging technologies and business activities includes performing appropriate risk assessments and due diligence prior to implementing these technologies, determining whether emerging technologies are appropriate for the intended business purpose, and evaluating whether the institution or utility has the necessary staffing, expertise, and other resources to manage the risks associated with those technologies.
- **Model Risk Management:** Appropriate governance and controls over the use of AI and other tools is an important aspect of managing risks. Supervisory guidance on model risk management has principles applicable to managing risks from AI, including assessing conceptual soundness, confirming underlying data, considering model complexity and transparency, assessing performance, and evaluating implementation.[33] Regardless of whether an AI tool or service is formally considered a model within the context of model risk management, appropriate risk management, including validation and testing, helps ensure AI tools and services operate as intended. Ongoing performance monitoring helps assess model implementation and whether the model is performing as intended. Similar considerations apply to

29 See, e.g., *Interagency Guidelines Establishing Standards for Safety and Soundness* (12 CFR Part 30, App. A (OCC); 12 CFR Part 208, App. D–1 (FRB), 12 CFR. Part 364 App. A (FDIC)); *Interagency Guidelines Establishing Information Security Standards*, 12 CFR 30, Appendix B (OCC), 12 CFR 208, Appendix D-2 (FRB, for state member banks), 12 CFR §§ 211.24 (FRB, for uninsured state-licensed branches or agencies of foreign banks), 12 CFR 225, Appendix F (FRB, for bank holding companies), 12 CFR 362, Appendix B (FDIC), *Computer-Security Incident Notification Requirements for Banking Organizations and Their Bank Service Providers*, (codified at 12 CFR part 53 (OCC); 12 CFR 225, subpart N (FRB); 12 CFR 304, subpart C (FDIC); *Interagency Guidance on Third-Party Relationships: Risk Management* (Board SR 23-4, FDIC FIL 29-2023, and OCC Bulletin 2023-17); 17 C.F.R. 240.15c3-5 (risk management controls for brokers or dealers with market access); 17 C.F.R. 242.1000-1007 (Regulation Systems Compliance and Integrity); 15 U.S.C 78o(g) (prevention of misuse of material nonpublic information by brokers or dealers); 17 C.F.R. 248.1-30 (Regulation S-P: Privacy of Consumer Financial Information and Safeguarding Personal Information).

30 See 12 CFR 30, appendix D, OCC Guidelines Establishing Heightened Standards for Certain Large Insured National Banks, Insured Federal Savings Associations, and Insured Federal Branches; and 12 CFR 52, Enhanced Prudential Standards issued by the Federal Reserve Board; and Federal Reserve SR Letter 20-24: *Sound Practices to Strengthen Operational Resilience* (Nov. 2, 2020).

31 See, e.g., 17 CFR 39 subpart C, establishing heightened risk management and cybersecurity requirements for derivatives clearing organizations that have been designated as systemically important; and 12 CFR part 234, establishing, among other things, operational risk management standards for certain systemically important financial market utilities supervised by the FRB.

32 OCC Bulletin 2017-43, New, Modified, or Expanded Bank Products and Services: Risk Management Principles.

33 OCC Bulletin 2011-12, *Sound Practices for Model Risk Management: Supervisory Guidance on Model Risk Managemen*t; FDIC Financial Institution Letter (FIL)-22-2017, Adoption of Supervisory Guidance on Model; Federal Reserve SR Letter 11-7, Supervisory Guidance on Model Risk Management.

financial market utilities that rely on models as risk management tools in setting financial resources, including for purposes of margining.[34]

- **Technology Risk Management:** Regulatory agencies set standards and expectations for sound risk management and evaluate controls for the use of technology, including AI technologies.[35] This should include appropriate development and acquisition, testing, and change management when introducing new and updated technologies. Sound technology risk management should include maintenance of an inventory of AI technologies being implemented, assessment of the level of risk associated with each AI use case, expectations for testing and ongoing validation, and issue and incident tracking. Effective information security, cybersecurity, resilience, privacy, and operational and fraud-related controls are also important for use of AI.
- **Data Management:** AI relies on large volumes of data; therefore, understanding data sources, quality, and limitations is a key aspect of sound risk management for AI. Effective data management and governance help to monitor the soundness and appropriateness of data leveraged by AI.[36] The Gramm-Leach-Bliley Act requires financial institutions—companies that offer consumers financial products or services like loans, financial or investment advice, or insurance—to explain their information-sharing practices to their customers and to safeguard sensitive data.
- **Third-Party Risk Management:** Financial institutions and financial market utilities should ensure they have prudent risk management over all activities, including AI, whether conducted in-house or through a relationship with a third party.[37] When implementing third-party AI, especially when limited model design documentation may be available, conducting and maintaining evidence of testing and validation performed, including any assumptions and model limitations, promotes effective risk management. Frequent and in-depth performance monitoring by the financial institution, including ongoing internal validations of AI results, can also be a compensating control. Financial institutions retain responsibility for the integrity of operations performed by third parties and contingency planning for potential disruptions.[38] Third-party guidance from regulators may also address the indirect use of AI through fourth, fifth, and other indirect parties.

34 See, e.g., 17 CFR 39.13 and 17 CFR 39.36, setting forth risk management considerations for systemically important derivatives clearing organizations.

35 See, e.g., 17 CFR Part 242 Subpart ECFRe106e84e67e2bc9, Regulation SCI—Systems Compliance and Integrity.

36 See FFIEC IT Examination Handbook InfoBase - III.A Data Governance and Data Management for relevant expectations available here: https://ithandbook.ffiec.gov/it-booklets/architecture-infrastructure-and-operations/iii-common-aio-risk-management-topics/iiia-data-governance-and-data-management/.

37 FRB, FDIC, and OCC, Interagency Guidance on Third-Party Relationships: Risk Management (Jun. 2023), https://www.federalregister.gov/documents/2023/06/09/2023-12340/interagency-guidance-on-third-party-relationships-risk-management.

38 See, e.g., 17 CFR. 39.18(d) concerning retention of responsibility in the event of outsourcing.

- **Insurance:** In accordance with the National Association of Insurance Commissioners (NAIC)'s 2020 Principles on Artificial Intelligence,[39] the NAIC adopted the Model Bulletin on the Use of Artificial Intelligence Systems by Insurers[40] in December 2023. The model bulletin was developed as principles-based guidance on applying a state's existing statutory framework to AI, rather than creation of a new standard. The model bulletin applies to all phases of an AI system's lifecycle, including design, development, validation, implementation, use, ongoing monitoring, updating, and retirement. The bulletin reminds insurers that AI-supported decisions impacting consumers must comply with all applicable insurance laws and regulations. The bulletin sets forth expectations for how insurers should govern the development, acquisition, and use of AI, to include the implementation of a written program in line with the insurer's risk assessment.

FBIIC partner agencies also advised of two additional topics not directly related to cyber, anti-fraud, or operational resilience.

- **Compliance and Consumer/Investor Protection:** Regardless of how AI is used in the activities of a financial institution, the institution is responsible for adherence to applicable laws and regulations. Complex interactions among data points within an AI model that are not readily observable or explainable by humans can produce unintended or problematic outcomes, and the source of any unfair outcomes may be masked by the model's complexity. Institutions are required to closely monitor outcomes of activity generated by AI and other models or tools to ensure that they adhere to fair lending laws, unfair and deceptive acts and practices prohibitions, and other investor and consumer protection requirements.[41] The intersection of AI, regulatory compliance, and consumer and investor protection is complex and is generally outside the scope of this report.
- **Securities Market Access Risk Management:** The Securities and Exchange Commission proposed new rules and amendments to address certain conflicts of interest associated with the use of predictive data analytics by broker-dealers and investment advisers in investor interactions. The proposal would require (1) such a firm to eliminate or neutralize the effect of conflicts of interest associated with the firm's use of covered technologies in investor interactions that place the firm's or its associated person's interest ahead of investors' interests; (2) such a firm that has any investor interaction using covered technology to have written policies and procedures reasonably designed to prevent violations of (in the case of investment

39 NAIC, *Principles on Artificial Intelligence* (Aug. 2020), https://content.naic.org/sites/default/files/inline-files/NAIC%20Principles%20on%20AI.pdf.

40 NAIC, *Model Bulletin on the Use of Artificial Intelligence Systems by Insurers* (Dec. 2023), https://content.naic.org/sites/default/files/inline-files/2023-12-4%20Model%20Bulletin_Adopted_0.pdf.

41 See, e.g., U.S. Securities and Exchange Commission IM-Staff Guidance Update No. 2017-2, Robo-Advisers (Feb. 2017) (discussing the unique considerations robo-advisers should keep in mind as they seek to meet their legal obligations under the Investment Advisers Act of 1940); CFTC Staff Advisory 14-21, https://www.cftc.gov/PressRoom/PressReleases/6866-14 (providing best practices for protecting the privacy of customer information).

advisers) or achieve compliance with (in the case of broker-dealers) the proposed rules; and (3) recordkeeping related to the proposed conflicts rules.[42]

Financial regulatory agencies continue to update regulations, rules, and supervisory guidance as technologies evolve and change the nature of markets and financial services.[43] The FSOC identified the use of AI in financial services as a potential vulnerability in the financial system.[44] The FSOC highlighted the importance of financial institutions, market participants, and regulatory and supervisory authorities deepening expertise and capacity to monitor AI innovation and usage and identify emerging risks. Many financial regulatory agencies have established dedicated units focused on innovation and emerging financial technologies in the financial sector. Several regulatory agencies are taking steps to better understand AI adoption in the financial sector[45] and are beginning to address AI in their communications.[46] Financial regulatory agencies have indicated that they will continue to assess whether updates to regulatory references and resources are needed as the use of AI grows and evolves in the financial sector.

The potential for further benefits as AI gains more widespread adoption could be significant. However, as with the implementation of any new technology or businesses process, unintended risks and consequences may occur if effective governance and controls are not implemented. Risks and consequences may be magnified as adoption of AI technologies becomes more widespread. Treasury, U.S. prudential regulators, and other U.S. government agencies also participate in a range of international fora concerning the regulation of AI in financial services that are more specific to client-facing products, including at the Financial Stability Board (FSB),[47] the Basel Committee on Bank Supervision (BCBS),[48] and the Organisation for Economic Co-operation and Development (OECD).[49]

---

42 See SEC's Fact Sheet, *Conflicts of Interest and Predictive Analytics* (Jul. 2023), https://www.sec.gov/files/34-97990-fact-sheet.pdf; and SEC Press Release, SEC Proposes New Requirements to Address Risks to Investors from Conflicts of Interest Associated with the Use of Predictive Data Analytics by Broker-Dealers and Investment Advisers, https://www.sec.gov/news/press-release/2023-140.

43 See SEC, *Conflicts of Interest Associated with the Use of Predictive Data Analytics by Broker-Dealers and Investment Advisers*, https://www.sec.gov/files/rules/proposed/2023/34-97990.pdf (Proposing Release, July 26, 2023). The proposal would also require a firm to have written policies and procedures reasonably designed to achieve compliance with the proposed rules and to make and maintain books and records related to these requirements. See also SEC, *Cybersecurity Risk Management for Investment Advisers, Registered Investment Companies, and Business Development Companies*, https://www.sec.gov/files/rules/proposed/2022/33-11028.pdf (Proposing Release, Feb. 9, 2022) and *Outsourcing by Investment Advisers*, https://www.sec.gov/files/rules/proposed/2022/ia-6176.pdf (Proposing Release, Oct. 26, 2022).

44 FSOC, *Annual Report 2023*, https://home.treasury.gov/system/files/261/FSOC2023AnnualReport.pdf.

45 FRB, OCC, FDIC, CFPB, and NCUA, *Request for Information and Comment on Financial Institutions' Use of Artificial Intelligence, Including Machine Learning* (Mar. 2021), https://www.federalregister.gov/documents/2021/03/31/2021-06607/request-for-information-and-comment-on-financial-institutions-use-of-artificial-intelligence.

46 FHFA, Advisory Bulletin 2022-02, *Artificial Intelligence/Machine Learning Risk Management* (Feb. 2022), https://www.fhfa.gov/SupervisionRegulation/AdvisoryBulletins/Pages/Artifical-Intelligence-Machine-Learning-Risk-Management.aspx.

47 FSB, *Artificial Intelligence and Machine Learning in Financial Services* (Nov. 2017), https://www.fsb.org/2017/11/artificial-intelligence-and-machine-learning-in-financial-service.

48 BIS, *Newsletter on artificial intelligence and machine learning* (Mar. 2022), https://www.bis.org/publ/bcbs_nl27.htm.

49 OECD, *Artificial Intelligence, Machine Learning and Big Data in Finance* (Oct. 2021), https://www.oecd.org/finance/financial-markets/Artificial-intelligence-machine-learning-big-data-in-finance.pdf.

5. Best Practices for Managing AI-Specific Cybersecurity Risks

Interviewed financial institution representatives stated that they rely on existing risk management methodologies to mitigate threats related to emerging technologies, including cybercrime and fraud. As highlighted in the regulatory discussion above, while existing supervisory risk management and operational resiliency expectations may not expressly address AI, risk management principles provide a framework for financial institutions implementing AI to operate in a safe, sound, and fair manner. Regulators note that financial institutions need to identify, monitor, and control risks arising from AI use as they would for the use of any other technology. This section describes some of the best practices shared by participating institutions to mitigate AI-related cyber and fraud risks.

5.1 SITUATING AI RISK MANAGEMENT WITHIN EXISTING ENTERPRISE RISK MANAGEMENT PROGRAMS

Interviewed financial institutions stated that they are embedding AI-specific risk management within their enterprise risk management programs, which vertically integrates AI-specific risk management within broader risk management practices. In 2011, the BCBS articulated the three layers of defense approach to risk management in its *Principles for the Sound Management of Operational Risk*.[50] As that report highlights, "operational risk is inherent in all banking products, activities, processes, and systems,"[51] and AI systems within financial institutions are no exception.

Generally, the three lines of defense for risk management are the business line, corporate risk management, and auditing risk controls. In such a structure, the business line is first responsible for managing risk associated with its AI systems and its business offerings. The second line provides compliance management systems and risk management structures to support the business lines managing AI-specific risk and enables risk-related communications and decisions about AI system use to be elevated to management. The third line audit assures the right monitoring, controls, and reporting are in place to manage the context-specific risks posed by AI.

In the absence of such a risk management structure, the NIST RMF suggests having a principles-based approach in which senior leadership determines the overall goals, values, policies, and risk tolerance within the organization and aligns the technical aspects of AI risk management to these goals. Regardless of the approach, NIST recommends that AI risk management governance should be a continual and intrinsic requirement over an AI system's lifespan and organizational hierarchy. Furthermore, transparency is required to improve human review processes and establish accountability in AI systems' teams.

50 Basel Committee on Banking Supervision, *Principles for the Sound Management of Operational Risk* (Jun. 2011), https://www.bis.org/publ/bcbs195.pdf.

51 *Ibid*.

Proper inventory of systems, appropriate documentation, and effective communication can bolster transparency throughout the organization.

5.2 DEVELOPING AND IMPLEMENTING AN AI RISK MANAGEMENT FRAMEWORK

Participating financial institution representatives stated that they are developing AI-specific risk management frameworks to guide their use of AI systems. While many AI frameworks and guidelines exist, notably NIST's RMF,[52] OECD AI Principles,[53] and Open Worldwide Application Security Project (OWASP) AI Security and Privacy Guide,[54] several interviewed financial institutions stated that they are developing their own tailored AI frameworks that leverage the aforementioned guidelines.

Having an AI framework in place allows financial institutions to identify AI risks associated with their use of AI systems in the context of their institution and desired AI use cases. Interviewed financial institutions report that they use their AI framework to map identified AI risks to existing controls across the enterprise. This mapping provides an overview of the existing coverage of controls related to AI risk and highlights gaps specific to AI that institutions can build into risk mitigation plans.

5.3 INTEGRATING RISK MANAGEMENT FUNCTIONS FOR AI

Financial institution participants stated that they are horizontally integrating risk management functions to cover the range of risks posed by AI systems. Some organizations have put this aspect of AI risk governance under the control of a single AI lead or assigned the responsibility to an existing official, like the Chief Technology Officer (CTO) or Chief Information Security Officer (CISO), while others have created AI-specific centers of excellence to address risks and opportunities posed by AI. The board of directors has assumed this role at some institutions.

Regardless of the location within the governance structure, financial institutions are advised to integrate AI plans into their enterprise risk management functions and connect them with other parts of the organization to address the multifaceted risks posed by AI. The most common integration within enterprise risk management occurs across model risk, technology risk, cybersecurity risk, and third-party risk management, which are all core functions associated with the implementation and use of AI systems. AI-specific integration at the enterprise level should include representation from legal, compliance, data science, marketing, and other business functions like operations, product management, and design, depending on the organization.

52 NIST, *Artificial Intelligence Risk Management Framework* (Jan. 2023), https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf.

53 OECD, *OECD AI Principles overview* (May 2019), https://oecd.ai/en/ai-principles.

54 OWASP, *OWASP AI Security and Privacy Guide*, https://owasp.org/www-project-ai-security-and-privacy-guide/.

5.4 EVOLUTION OF THE CHIEF DATA OFFICER ROLE AND MAPPING THE DATA SUPPLY CHAIN

According to interviewed financial institutions, the data requirements to power AI models have been a propelling factor for them to go beyond compliance and take a more proactive approach to data acquisition, curation, privacy, security, and monetization. Structured and unstructured data undergo different phases of transformation and integration from its raw state before reaching the business intelligence teams or data scientists who may be responsible for building the AI models. Additionally, raw data may be originated from different sources and systems internal or external to the organization spanning different teams, offices, and jurisdictions. All these aspects constitute a complex supply chain that is becoming more significant as more key functions rely on AI systems. A number of financial institutions reported building a comprehensive inventory and mapping of their data supply chain under a corporate data lead.

In this model, typically, the corporate data lead integrates data across the organization through participation in a cross-functional AI risk management team to streamline data requirements across the firm and to capitalize on data opportunities for AI systems. This includes both internal use of AI systems and vendor solutions that have embedded AI.

Depending on the corporate structure of a financial institution, the corporate data lead may not be always called the Chief Data Officer (CDO) but may be empowered with similar responsibilities. This role will likely continue to evolve as financial institutions adopt more AI systems and AI systems become embedded in more products. With careful implementation, communication, and leadership support, the CDO can drive innovation and manage data as a business asset.[55]

In Treasury's own AI adoption strategy, Treasury has broadly leveraged the Chief Data Officer role to empower both technology transformation and AI development. For example, Treasury established a CDO position for the Office of Terrorism and Financial Intelligence (TFI), which includes FinCEN, the Office of Foreign Assets Control (OFAC), the Office of Intelligence and Analysis (OIA), the Office of Terrorist Financing and Financial Crimes (TFFC), and Treasury Executive Office for Asset Forfeiture (TEOAF). The TFI CDO was empowered to create an executive-level data governance board to drive decision making on cloud adoption strategies, federated data lake development, and AI use cases with a focus on scalability, accessibility, access control, and protection of data required to enable the future development of AI models.

In 2023, this multi-year effort led by TFI CDO resulted in the launch of the CACHE platform, a consolidated data platform for open source, subscription, and proprietary Treasury data sources. Treating data as the foundation for AI development coupled with the

55 Eastwood, Brian. *Chief data officers don't stay in their roles long. Here's why.* (Sept. 2022), MIT Sloan School of Management, https://mitsloan.mit.edu/ideas-made-to-matter/chief-data-officers-dont-stay-their-roles-long-heres-why.

availability of large quantities of curated data enabled the rapid development of AI models by Treasury data scientists. As a result, the technical, legal, and security aspects of this work were integrated into the technical and data transformation infrastructure, enabling employees to develop solutions in a secure environment.

5.5 ASKING THE RIGHT QUESTIONS OF VENDORS

When considering vendors that offer AI systems, or products and services relying on AI systems, financial institutions should consider expanding their typical third-party due diligence and monitoring to account for AI-specific factors. In addition to common third-party risk-related questions, financial institutions should consider inquiring about AI technology integration, data privacy, data retention policies, AI model validation, and AI model maintenance. Financial institutions should also consider asking their vendors if they rely on other vendors for data or models and if so, how they manage and account for these factors.

Some requests that financial institutions should consider asking their vendors include:

- Notify the financial institution if the third party makes changes or updates to products or services that use AI systems.
- Disclose the scope of AI system use in their products or services and notify them of material changes.
- Describe the model and data lifecycles, when an AI system is significant to a product or service.
- Explain the impact the AI systems could have on the financial institution's customers and how the financial institution can explain this impact to their customers.
- Describe the implemented security practices, including patch management and vulnerability assessment process of the infrastructure hosting the AI system.
- Describe any incorporated underlying third-party AI models.

FS-ISAC recently published a *Generative AI Vendor Evaluations & Qualitative Risk Assessment Guide* as well as a *Generative AI Vendor Evaluation & Qualitative Risk Assessment Tool* that may be helpful for financial institutions as they plan for and engage with Generative AI vendors.[56]

5.6 SURVEYING NIST'S CYBERSECURITY FRAMEWORK TO IDENTIFY OPPORTUNITIES FOR AI USE

Many of the major financial institutions Treasury interviewed have reported that they are looking at how they can use AI systems to improve cybersecurity by optimizing workflows,

56 FS-ISAC, *Financial Services and AI: Leveraging the Advantages, Managing the Risks* (Feb. 2024), https://www.fsisac.com/knowledge/ai-risk.

assisting human practitioners, and identifying trends and patterns in both threat and defense data.

In the category of using AI for cybersecurity, interviewed financial institutions report they are using the NIST Cybersecurity Framework (CSF) to identify opportunities for AI solutions to enhance cybersecurity.[57] Several financial institutions expressed that they are analyzing their current practices against the NIST CSF's key functions and breaking down the processes, procedures, and systems that they have in place to manage cybersecurity risk. With this reevaluation of their posture, some participating financial institutions are considering where it makes sense to use AI systems to augment existing processes and procedures or fill gaps in coverage across the cybersecurity lifecycle.

5.7 IMPLEMENTING RISK-BASED TIERED MULTIFACTOR AUTHENTICATION MECHANISMS

Interviewed financial institutions reported that they are aware that they should be implementing and extending multifactor authentication solutions to enhance both cybersecurity and fraud protections against AI-powered threats. Criminal use of AI will challenge the effectiveness of current identity solutions for customers of financial institutions, including the use of biometrics like voice or video recognition and soft biometrics like keystrokes or other behavioral patterns, which are currently considered to be the best authentication methods. In their place, financial institutions may continue to develop and adopt authentication methods, such as offering out-of-band identity tokens (digital identity credentials) to their customers to authenticate their identity for authorizing access to financial accounts or other information, especially to conduct account-based activity like transferring funds.

NIST recommended deprecating short message service (SMS) for multifactor authentication several years ago.[58] In its place, many options exist that offer better assurance but come with increasing costs:

- Hardware-based (e.g., Fast Identity Online, or FIDO-compliant) authentication devices
- App-based passkeys
- Other password-less solutions

Financial institutions should be wary of disabling any of these factors, like geolocation or device fingerprinting.

57 NIST, *Cybersecurity Framework*, https://www.nist.gov/cyberframework.

58 NIST, *Questions…and buzz surrounding draft NIST Special Publication 800-63-3*, https://www.nist.gov/blogs/cybersecurity-insights/questionsand-buzz-surrounding-draft-nist-special-publication-800-63-3.

5.8 PICKING THE RIGHT TOOL FOR THE JOB AND RISK TOLERANCE

Financial institutions should consider their enterprise risk tolerance when implementing AI systems, particularly Generative AI. Regardless of whether a financial institution is considering onboarding a completely new AI system or augmenting an existing system with new AI functionality, it should take into account the level of risk of the function that the new system supports. While enterprise IT system developers and practitioners may be pressured to fully capitalize on recent advancements of Generative AI, financial institutions should evaluate any solution or vendor product based on its capabilities, applicability, and limitations.

The use case for AI systems should account for the risk tolerance associated with current Generative AI shortcomings. If a higher level of explainability[59] 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. Financial institutions should take these considerations into account when determining how to use Generative AI and how to fit these concerns within the risk tolerances of both the particular use case and the overall risk appetite of their firms.

5.9 CYBERSECURITY BEST PRACTICES CLOSELY APPLY TO AI SYSTEMS

Financial institutions should use current cybersecurity best practices to secure AI systems. Financial institutions should map security controls to AI applications and ensure that AI systems are subject to at least the same levels of cybersecurity as that of any other IT system used by a financial institution. For example, data loss prevention standards should extend to Generative AI as they do to other tools used by financial institutions if a firm's data loss prevention policy prohibits entering customer information, firm's documents, or files into a search tool or an external system, it should also cover entering them into a Generative AI tool.

Secure-by-design principles are not only applicable to AI systems but have become more essential as AI systems' risk mitigation solutions are much more demanding. AI systems are increasingly more complex and minor changes to the system may require changes in the models, which may need retraining or retuning. Additionally, AI systems' integration into an organization's IT systems requires investment in training multiple teams across that organization. The process of applying frequent patches for security vulnerabilities after integration is more complex. Therefore, it is paramount that cybersecurity and other risks are considered in the design and development phases of AI systems. Financial institutions are encouraged to communicate their security thresholds and requirements to

59 See NIST, *The Four Principles of Explainable AI*, NISTIR 8312, https://nvlpubs.nist.gov/nistpubs/ir/2021/NIST.IR.8312.pdf.

their vendors before the delivery of the products or, in the case of in-house development, to incorporate secure design considerations using recommendations such as the Guidelines for Secure AI System Development.[60]

Furthermore, in developing AI systems, especially ML models and LLMs, financial institutions should pay close attention to cybersecurity practices around data. They should consider extending best practices for data security to training data and test data, which might have been overlooked as low risk in the past but can influence model behavior and attestation because of the correlation between the models and the data they ingest.

60 See UK National Cyber Security Center, et al, *Guidelines for Secure AI System Development* (Nov. 2023), https://www.ncsc.gov.uk/files/Guidelines-for-secure-AI-system-development.pdf.

6. Next Steps: Challenges & Opportunities

There are many direct and indirect AI-related challenges and opportunities for the financial services sector, including those regarding consumer and investor protection, disparate impact, financial stability, and financial regulatory concerns. Treasury will continue to assess different aspects of AI impact for the financial services sector with respect to gaps and emerging concerns. This section lays out some 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.

6.1 NEED FOR A COMMON AI LEXICON

Financial institutions, regulators, and consumers would benefit from a common lexicon specific to AI. In interviews, many participants stated that "artificial intelligence" itself is an imprecise term and could mean many different things. There was little agreement among participants about what the label meant. What everyone did agree on, however, was the need for a common lexicon to promote common understanding. A common lexicon would not only facilitate appropriate discussion with third parties and regulators but could help improve understanding of the capabilities AI systems may have to improve risk management or to amplify new risks.

Careful consideration of terminology may help address the current lack of clarity around measuring and identifying risks, especially with the rapid adoption of Generative AI. As noted in the introduction, terminology can have implications for the common understanding of AI technology and its associated risks as well. For instance, one firm uses "augmented intelligence" to shift the responsibility to the user by emphasizing that the system is augmenting the user's intelligence, rather than having its own intelligence. Similarly, the use of "hallucination" to describe false outputs by Generative AI suggests these Generative AI systems intend meaning in their outputs when what they are supplying is probabilistic semantics.[61] This anthropomorphism may misleadingly imply intention and create a false sense of trust in a system. For this reason, one firm said they use "prompt" and "response" rather than "question" and "answer," indicating an attempt to neutralize the language associated with these systems to retain the primacy of human agency.

As a first effort, Treasury has included a glossary section in this report, based on NIST's AI RMF and Adversarial Machine Learning documents. Treasury intends to collaborate with FBIIC partner agencies and FSSCC members to develop a common lexicon of AI terminologies most relevant to financial institutions.

61 Stening, Tanner, *What are AI chatbots actually doing when they 'hallucinate'? Here's why experts don't like the term* (Feb. 2024), Northeastern Global News, https://news.northeastern.edu/2023/11/10/ai-chatbot-hallucinations/.

6.2 ADDRESSING THE GROWING CAPABILITY GAP

There appears to be a widening capability gap between the largest and smallest financial institutions when it comes to building AI systems in-house. The largest, most sophisticated firms report that they are experimenting and developing AI systems in ways smaller institutions do not have the resources to do. One firm has stated that it has approximately 400 employees working on fraud-prevention AI systems, and AI service providers noted being approached with thousands of use cases by larger firms. Smaller firms report that they do not have the IT resources or expertise to develop their own AI models; therefore, these firms solely rely on third-party or core service providers for such capabilities.

With the tight intersection between AI and cloud services, financial institutions that have already moved data and services to the cloud may have an advantage when it comes to leveraging AI in a safe and sound manner. This gap is not insurmountable if firms can transition to the cloud, although firms already in the cloud will have had more time to experiment and refine their AI systems. These early adopters may be able to shorten the AI ramp-up time for later cloud adopters. Some smaller financial institutions noted issues with getting their own data from their core providers.

Treasury will seek to facilitate conversations between FSSCC members and critical core providers, along with their direct and indirect oversight agencies and consumer advocate organizations, to better understand how core providers are working toward developing AI-enhanced capabilities for financial institutions in the cybersecurity and anti-fraud space. In addition, Treasury will explore opportunities to collaborate with trade associations and other stakeholders to extend smaller institutions' access to AI capabilities.

6.3 NARROWING THE FRAUD DATA DIVIDE

As Generative AI increases in usage, there appears to be a significant gap in data available to financial institutions for training their models to prevent fraud. Financial institutions represented that they are working with the data they currently have to develop their models. Ramifications of this data divide are especially apparent for anti-fraud use cases where larger institutions generally have much more internal data. Interviewed smaller institutions noted they have insufficient in-house data to build their own anti-fraud models even if they could overcome other resource limitations in building AI systems.

Cybersecurity information sharing in the financial sector, especially through the FS-ISAC, has matured over the years, but little progress has been made to enhance data sharing related to fraud. The ABA is working to design, develop, and pilot a new information-sharing exchange focused on fraud and other illicit finance activities.[62] The U.S. Government, with its collection of historical fraud reports, may be able to assist with this effort to contribute to a data lake of fraud data that would be available to train AI, with appropriate and necessary safeguards.

62 American Banker, *ABA to launch information-sharing exchange to help banks fight fraud* (Nov. 2023), https://www.americanbanker.com/news/aba-to-launch-information-sharing-exchange-to-help-banks-fight-fraud.

Treasury can be a leader in this space and will work with the financial sector, including the ABA and FS-ISAC, to improve fraud data sharing from Treasury. As reported in a February 28, 2024, Treasury press release, Treasury's Bureau of the Fiscal Service (BFS) recovered over $375 million as a result of its implementation of an enhanced fraud detection process that utilizes AI.[63] Building on the work already implemented by BFS, the Treasury Chief AI Officer will convene an AI anti-fraud surge team from across the Department to ensure that Treasury appropriately leverages all available information streams to inform both the sector and Treasury's own internal operations.

6.4 REGULATION OF AI IN FINANCIAL SERVICES REMAINS AN OPEN QUESTION

Every participant in this study noted that the current financial sector domestic regulatory environment allows for prudent innovation. Several participants noted they were working cooperatively with their regulators on AI issues to make sure they were developing and using AI consistent with the applicable regulatory framework. Some financial institutions, however, expressed concern about the possibility of regulatory fragmentation as different financial sector regulators at both the state and federal level consider regulations around AI. This concern also extends to firms operating under different international jurisdictions.

Treasury will work with FBIIC and FSSCC to map major existing and anticipated regulatory regimes relevant to financial sector firms and their vendors in the cybersecurity and fraud space. This effort will explore potentially enhancing coordination across regulators with the goal of fostering responsible AI advancements, while addressing risk, and understanding applicable regulatory regimes. The coordination actions could include the recommendation to establish AI-specific coordinating groups, as allowable, to assess enhancing shared standards and regulatory coordination options.

6.5 EXPANDING THE NIST AI RISK MANAGEMENT FRAMEWORK

The NIST AI RMF could be enhanced to include more substantive information related to AI governance, particularly as it pertains to the financial sector. Many financial institutions that participated in this study noted drawing on the NIST AI RMF when developing their own AI frameworks. Some of the early movers in this space mentioned having to develop their own AI frameworks without the benefit of the NIST RMF because it had not yet been published.

The financial sector's maturity with both AI and risk management, including enterprise risk management, could help inform AI governance models applied to other industries.

Treasury will assist NIST's U.S. AI Safety Institute (USAISI) to establish a financial sector-specific working group under the new AI consortium construct with the goal of extending the AI RMF toward a financial sector specific profile.

63 U.S. Department of the Treasury, *Treasury Announces Enhanced Fraud Detection Process Using AI Recovers $375M in Fiscal Year 2023* (Feb. 2024), https://home.treasury.gov/news/press-releases/jy2134.

6.6 BEST PRACTICES FOR DATA SUPPLY CHAIN MAPPING AND "NUTRITION LABELS"

Participating financial institutions stated that they have significant concerns about the data used to train commercial Generative AI because of privacy and liability concerns. The recent rapid advancements in Generative AI have exposed the importance of carefully monitoring data supply chains to make sure that Generative AI models are using accurate and reliable data that provides for appropriate privacy and safety considerations. Additionally, financial institutions noted that they are paying closer attention to the state of their internal data and whether it is tagged with proper permission and restrictions to prevent it from inadvertently becoming available to third-party AI systems for training purposes. Additionally, financial institutions should know where their data is going and how it is being used.

Financial institutions have stated that they would benefit from the development of best practices concerning the mapping of data supply chains and data standards. This mapping would enable firms to understand restrictions and user rights throughout the training data supply chain and AI model-output data chain while at the same time enabling privacy and other data protection concerns. Relatedly, financial institutions stated that they would benefit from a standardized description, similar to a nutrition label, for vendor-provided Generative AI systems to clearly identify what data was used to train a model, where it came from, and how any data submitted to the model will be incorporated.

Treasury will work with the financial sector, NIST, the National Telecommunications and Information Administration (NTIA), and the Cybersecurity and Infrastructure Security Agency (CISA) to identify whether such recommendations should be explored further. To the extent feasible, this line of effort can potentially build upon the collaborative methods used for the Software Bill of Materials (SBOM) workstreams led by NTIA and CISA covering traditional software components. The best practices could recommend repeatable methods to label incoming and outgoing data, including sharing a firm's data with external partners or ingesting data to train a model.

6.7 DECIPHERING EXPLAINABILITY FOR BLACK BOX AI SOLUTIONS

Most of the interviewed firms described explainability of advanced ML models, and specifically Generative AI, as a challenge for financial institutions. Explainability refers to the system's ability to provide the rationale for the output of the system on a consistent basis.

Financial institutions report that they are currently limiting Generative AI systems to use cases where lower explainability levels may be deemed by the financial institution as sufficient because they are less likely to have to provide any insight into how the outputs were generated. However, going forward, more sensitive use cases that raise concerns like safety, privacy, and consumer protection may increase the need for explainability.

Interviewed industry representatives have stated that they would benefit from research and development surrounding explainability solutions for black box systems like those that Generative AI offers. These could consider the data used to train the models and the outputs, along with before and after states of the models themselves. Additionally, these solutions could include robust testing and auditing of the models. Absent workable solutions for providing black box explainability, financial institutions should establish best practices for using Generative AI without explainability. These 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. Currently there seems to be a lack of an appropriate comprehensive framework for the testing and audit of black box AI solutions that would guide firms through the critical steps to assess inputs, outputs, training the models, and the underlying models themselves. Such a framework should be repeatable and scalable to firms of varied sizes and complexity.

Treasury will work with the financial sector, NIST, and relevant R&D programs to further study the implication of black box AI explainability for cyber and fraud-related issues for the financial services sector.

6.8 GAPS IN HUMAN CAPITAL

The rapid change of pace for AI development has exposed a large talent gap in the workforce. While IT workforce challenges are not new, the rate of change may make some of the AI-related workforce gaps difficult to close, leaving most institutions entirely reliant on third parties for the development or implementation of AI systems. Investing in an AI workforce alongside the already significant IT and data requirements for AI systems may be hard for many financial institutions to sustain. Therefore, financial institutions would benefit from a set of tailored best practices for less skilled practitioners around how to use AI systems.

Further, the talent gap is not limited to those building and deploying AI systems. The technical competency gap also exists across the core teams managing AI risk, such as those in the legal and compliance fields. Financial institutions would benefit from role-specific AI training offerings to help educate those critical enabling roles outside of IT. Closing this gap will be the key to the safe and effective use of AI by financial institutions regardless of where the AI systems originate.

AI tools may be applied to use cases across a financial institution, including business lines, enterprise IT, customer service, compliance, and human capital. An important aspect of managing any technology risk is ensuring that all employees and vendors are properly trained in technologies they may encounter. AI is no exception to this, especially Generative AI, which may have even greater potential to create or magnify errors. Financial institutions should consider providing training for all employees on how to understand and properly use AI tools, as well as recognize malicious activities where AI may be involved.

6.9 UNTANGLING DIGITAL IDENTITY SOLUTIONS

The rise in remotely delivered financial services, together with the lack of secure, privacy-preserving, and accurate digital identity solutions have presented challenges to identity-related efforts to combat fraudulent finance activities. At present, remote customer identity verification at account opening is based on digitizing a patchwork of largely paper-based federal, state, and local government identifiers and credentials, while identity authentication to authorize account access and conduct transactions may rely on weak passwords or partial social security numbers that may have been previously exposed in data breaches. Cybercriminals exploit gaps at all stages of financial institutions' customer identity processes.

An emerging set of international, industry, and national digital identity technical standards, infrastructure, and governance frameworks, as well as pilots and government and private sector digital identity solution rollouts is underway. These efforts support the development and implementation of digital identity credentials—digital objects or data structures that authoritatively bind or link an identity (i.e., an established record of the individual's identity attributes or identifiers such as name, address, date of birth, or tax identification number) to the individual via at least one associated authenticator possessed or controlled by the identified individual that can be used to authenticate identity for both enterprise (internal-facing) and public (external-facing) use cases. [64]

Robust enterprise digital identity solutions may help financial institutions combat fraud and insider threats and strengthen cybersecurity. However, digital identity systems, solutions, and credentials differ in their technologies, governance, and security and, therefore, offer different levels of assurance regarding their accuracy, privacy, and overall effectiveness. In addition, although outside the scope of this report, digital identities may implicate concerns about fairness and inclusion, and may create challenges for certain individuals or populations.

To further inform our efforts to encourage responsible innovation, Treasury will continue to monitor the development of digital identity solutions and how they are being used by government and the financial services industry. Over the past several years, Treasury has leveraged its interagency and public-private partnerships to share information and explore best practices for mitigating the threats financial institutions face from gaps and vulnerabilities in identity processes—including the potential use of digital identity solutions and infrastructure to strengthen anti-money laundering and countering the financing of terrorism (AML/CFT) compliance and facilitate financial inclusion and equity. Treasury is also tracking and, as appropriate, supporting the efforts by federal and state agencies to develop standards and to mitigate risks in these emerging systems and solutions. These include the following:

64 See NIST, *Digital Identity Guidelines* (Dec. 2022), 800-63-4 IDP, https://csrc.nist.gov/pubs/sp/800/63/4/ipd.

- NIST is considering state mobile driver's licenses (mDLs) and verifiable credentials in its revision of its Digital Identity Guidelines and mapping its standards to the European Union's standards for digital identity wallets.[65]
- The Transportation Security Administration is currently setting minimum privacy and security standards for federal use of state mDLs and conducting large-scale acceptance pilots of the International Organization for Standardization (ISO)-compliant state mDLs and private sector-derived digital identities at airport check-ins.
- The NIST National Cybersecurity Center of Excellence is working with the U.S. Department of Homeland Security Science and Technology Directorate (DHS S&T) and the private sector, including financial institutions, on customer identity verification use cases to convert mDL standard into code and creating a toolkit for relying parties. DHS S&T is also testing onboarding and authenticating solutions' ability to detect fake documents, selfie matching, and liveness testing.

6.10 INTERNATIONAL COORDINATION

Similar to the domestic regulatory landscape, regulation of AI in financial services remains an open question internationally. Treasury will continue engaging with foreign counterparts on the risks and benefits of AI in financial services, including through ongoing bilateral financial and regulatory dialogues with the United Kingdom and the European Union. And through the FSB, Treasury will continue working with international partners to assess the risks of AI to global financial stability. The FSB will deliver an updated assessment of financial stability risks of AI to the Group of 20 (G20) by the end of 2024. Treasury is also participating in work at the OECD on the risks and benefits of AI to financial services, which is hosting an international workshop on AI jointly with the FSB in May 2024. Additionally, Treasury co-chairs the Group of 7 (G7) Cyber Experts Group (CEG) with the Bank of England. The CEG will continue to monitor the AI landscape and will develop additional lines of effort where it identifies cybersecurity gaps.

65 See Draft NIST, EU-*US TTC WG-1 Digital Identity Mapping Exercise Report* (Dec. 2023), https://www.nist.gov/system/files/documents/2023/12/22/EU-US%20TTC%20WG1_Digital_Identity_Mapping_Report_Final%20Draft%20for%20Comment_22122023.pdf.

7. Conclusion and Other Treasury AI Work

While the scope of this report is limited to cyber and fraud issues, Treasury remains focused on a range of AI-related matters, ranging from developing Treasury's own AI use cases to the potential risks and benefits associated with financial institutions' use of AI, including issues broadly impacting third-party oversight and consumers.

AI presents opportunities for the financial sector, but also significant potential risks, particularly for consumers and historically marginalized groups. Applications of AI in consumer financial services, insurance underwriting, fraud detection, and other areas of financial services have the potential to perpetuate or amplify existing inequities. Use of AI in these contexts also raises questions regarding consumer privacy and data security.

Treasury and financial regulators are also considering other potential risks and benefits associated with the use of AI by financial institutions. As discussed in the report in the context of cyber risk management, longstanding principles for sound risk management, including model risk management and third-party risk management, are critical to addressing the risks of AI more broadly. Treasury continues to monitor the development and use of these tools and consider risks to financial institutions and the financial system. To better understand the use of AI by financial institutions and its impact on consumers and investors, as well as the sufficiency of existing regulatory frameworks, Treasury is exploring opportunities for deeper engagement with the public.

Additionally, as highlighted in the 2023 FSOC Annual Report,[66] Treasury continues to monitor the deployment of AI in financial services overall to identify risks (besides cybersecurity and fraud) that could undermine the financial sector's integrity and stability.

66 See U.S. Department of the Treasury, *2023 FSOC Annual Report*, https://home.treasury.gov/system/files/261/FSOC2023AnnualReport.pdf.

Annex A: FSSCC R&D Committee Paper:
Artificial Intelligence in the Financial Sector: Cybersecurity and Fraud Use Cases and Risks

February 2024

EXECUTIVE SUMMARY

In an increasingly digitized financial ecosystem, the role of artificial intelligence (AI) is both pivotal and multifaceted.[67] Many financial institutions already use AI-powered solutions to manage cybersecurity and fraud risks to their core product and service offerings. Most institutions are now assessing novel AI technologies to enhance core business, customer, and risk management activities. The integration of AI offers the sector increased efficiency, precision, and adaptability, as well as the potential to bolster the resiliency of institutions' systems, data, and services. This opportunity is set against the reality of a complex and persistent threat landscape that is also adopting AI for malicious purposes.

This white paper aims to provide policymakers with an understanding of opportunities and risks, particularly cybersecurity-risks, of AI deployment within the financial sector. To achieve this, the paper first examines the current and anticipated use cases of cybersecurity and fraud AI solutions within the sector. The paper then focuses on the novel risks introduced alongside the evolution of AI technologies, including how adversaries are leveraging these advancements to create new threats. Finally, the white paper offers ideas on how policymakers can encourage AI innovation while effectively managing risk across the sector.

INTRODUCTION

The Financial Services Sector Coordinating Council's Research and Development Committee convened a series of discussions in the Fall of 2023 to discuss how advances in AI (including the release of AI tools such as ChatGPT) could impact cybersecurity, fraud prevention, third party risk management, and governance within the financial services sector.[68] These discussions were organized to inform the U.S. Treasury Department, which was tasked to write a report by the Biden Administration's Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence (EO).[69]

67 The term "artificial intelligence" or "AI" has the meaning set forth in 15 U.S.C. 9401(3): a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. Artificial intelligence systems use machine- and human-based inputs to perceive real and virtual environments; abstract such perceptions into models through analysis in an automated manner; and use model inference to formulate options for information or action.

68 The FSSCC is composed of 50 financial trade associations, financial market utilities, and the most critical financial firms. The FSSCC coordinates across sector participants to enhance the resiliency of the financial services sector, one of the nation's critical infrastructure sectors. The FSSCC proactively promotes an all-hazards approach to drive preparedness through its collaboration with the U.S. Government for the benefit of consumers, the Financial Services Sector, and the national economy. Additional details are available on the FSSCC website: https://www.fsscc.org.

69 https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/.

Since the advent of artificial intelligence and machine learning, the financial sector has looked to deploy these technologies for business, security, and risk management use cases. More recently, the emergence of Generative AI (GenAI) and Large Language Models (LLMs) has spurred widescale efforts within the industry to understand how the technology might be put to work and what governance processes are necessary to control risks associated with novel technology. This most recent integration of AI into the financial sector has witnessed increased scrutiny due to the emergence of widely available AI tools and broader understanding of its potential to revolutionize operations and decision-making. The rising prominence of AI has increased concern for both potential misuse, manifesting as cyber threats and AI-driven fraud, as well as inadvertent errors, where outputs are mistakenly assumed to be correct. This dynamic has intensified overall concerns about AI's application and trustworthiness for critical functions within the sector. The focus of most early adopters of AI technologies in the financial services sector is on in-house implementation of the technology. It is worth noting that the security and intellectual property protections of the AI services supply chain are still maturing.

USE CASES: USING AI TO REDUCE CYBER AND FRAUD RISKS

CYBERSECURITY

Financial institutions have a long history of deploying and controlling risk related to novel technologies. This is true with artificial intelligence and machine learning (AI/ML), where institutions have utilized graph-based analytics to, among other use cases, detect anomalous and suspicious activity and flag for further investigation. Many vendors now integrate AI/ML technology into commercially available solutions that are widely used across the sector.

The recent introduction of widely available GenAI technologies has created new opportunities and risks. At this time, only a few larger financial institutions have deployed limited GenAI solutions in support of enterprise cybersecurity. GenAI is expected to significantly transform the cybersecurity ecosystem, enabling cybersecurity professionals to process data and gain deeper insights in shorter cycle times. The use of GenAI will facilitate the automation of analyzing threat actor behaviors and streamlining alerts, investigations, and responses. Importantly, it should also serve as a countermeasure against AI-driven attacks, allowing for a more robust defense mechanism in the ever-evolving cybersecurity landscape.

Financial institutions described GenAI use cases that include both assisting cybersecurity professionals in comprehending malicious code, as well as aiding internal developers in identifying and mitigating vulnerabilities in their own code. One institution developed an AI tool to assist security analysts in the initial classification of suspicious emails. While this tool is designed to scale up the analysts' work by reducing the time spent on false positives, the institution highlighted the essential role of human verification.

This necessity stems from the fact that the accuracy of such models, though high, is not infallible. By involving human oversight, the institution aims to mitigate potential errors, ensuring the balance between automated efficiency and operational safety is maintained. At the same time, financial institutions have been vigilant about the risk of AI-induced complacency on cyber workforces.

FRAUD DETECTION AND PREVENTION

Financial Institutions that have adopted AI and machine learning (ML) models for fraud detection have seen transformative results. Some smaller institutions have noted being in the initial phases to implement AI to address fraud but recognize the transformative impact the technology has brought to other institutions. These models have a proven track record in detecting a spectrum of fraudulent activities, ranging from fraudulent credit card applications to dubious transactions and check fraud. The models' ability to analyze more extensive and complex data has sharpened their proficiency in identifying patterns predictive of fraudulent behavior.

The effectiveness of AI underscores the necessity for rapid and safe implementation of AI technologies in the financial sector. By doing so, institutions not only leverage AI's effectiveness in reducing the frequency and impact of fraud but also stay one step ahead of fraudsters and cybercriminals who are increasingly using similar advanced technologies. The clear success of AI in enhancing operational processes and safeguarding financial operations affirms its vital role in the ongoing battle against financial fraud and cyber threats. Furthermore, some institutions perceive integrating GenAI technology could significantly amplify current uses of AI for fraud detection and prevention. GenAI's advanced algorithms and learning capabilities can adapt to evolve in response to the ever-changing tactics of fraudsters, providing more dynamic and proactive approach to identifying and mitigating fraudulent activities. The potential synergy of GenAI with existing AI systems promises a more robust and resilient defense against significant threats.

ADVERSARIAL APPROACHES: HOW AI IS BEING UTILIZED BY CYBER

One of the sector's foremost concerns is the acceleration of threat actors' capabilities due to AI, particularly GenAI. While AI technologies, including less sophisticated ones, enable threat actors to deploy advanced attack tactics, the real challenges lie in the reduced cycle time for these actors. Skilled adversaries, aided by AI, can swiftly identify and implement novel breach techniques, potentially outpacing traditional detection strategies. This necessitates a continuous and rapid update in detection methodologies to address cyber threats that financial institutions might face.

In discussing the adversarial use of AI in cybercrime, it is important to differentiate between two key concepts: *adversarial machine learning*, which involves manipulating models to cause misjudgment, and *adversarial usage*, where AI, specifically large language models, is actively employed to execute cyberattacks. The use of AI for generating

credible-looking text, images, code, and malware, while concerning, is just one facet of the challenge at hand. Equally alarming is AI's ability to rapidly evolve these techniques, creating a dynamic threat landscape that traditional security measures may begin to struggle to keep pace with.

The evolution of LLM-generated content and deepfake creation services are of concern to the financial sector, especially smaller and less well-resourced institutions. These technologies not only lower the barrier to entry but also complicate authenticity verification measures. Concerns also extend to prompt injections into various forms of LLMs, with the speed of patching varying depending on deployment methods like stateless LLMs.

AI's capacity to conceal threats within multimodal content, such as images, combined with the prevalence of hallucinations or AI-generated misinformation, presents emergent challenges. While AI providers are implementing guardrails to deter this activity, security researchers have demonstrated that it is not difficult to circumvent the guardrails, necessitating robust, layered defenses. Current capabilities may not be fully equipped to address these novel threats, necessitating enhancements in both technical capabilities and control processes.

While the overall malicious activity levels might appear static, the reality could be more nuanced. The intensification of the challenge lies in detecting and understanding the exploitation of models by sophisticated actors. This is particularly true for GenAI, which is increasingly being used to create and disseminate misinformation at an alarming pace, often bypassing traditional detection methods. The potential for nation state actors to exploit GenAI for misinformation campaigns poses a significant and evolving threat.

Furthermore, the advent of open-source AI introduces additional risks like software supply chain security vulnerabilities and data/model poisoning. Cybercriminal groups are using modified versions of open-source models, such as "WormGPT," to circumvent model restrictions, leading to more sophisticated phishing campaigns with fewer typos and improved formatting. Beyond phishing via email, the rise of social engineering attacks through text and voice technologies poses new challenges.

Recognizing that existing controls for phishing campaigns are vital, financial institutions are also aware of the need to evolve their defenses. This includes keeping pace with cybercriminals' advancements by leveraging novel AI technologies. By reinforcing technological defenses and the human element of cybersecurity, financial institutions aim to adapt to these emerging trends and better protect against the dynamic nature of AI-facilitated cyber threats.

RISK MANAGEMENT APPROACHES FOR AI-DRIVEN CYBERSECURITY

In the financial sector, risk management methodologies are pivotal in maintaining integrity and stability amidst threats, including those from cybercrime and fraud. The Three Lines of defense model serves as a foundation framework within the sector, promoting rigorous oversight and clear delineation of responsibilities among operational management, risk compliance, and internal audit functions. Adhering to best practices allows financial institutions to harness the full potential of AI use for cybersecurity and fraud mitigation while accounting for emerging threats and evolving methodologies.

Following the release of the NIST AI Risk Management Framework (RMF) in January 2023, many financial institutions have commenced self-assessments of their practices against the framework.[70] The integration of practices noted in AI RMF, along with the adoption of MITRE ATT&CK for Learning Systems (ATLAS), OWASP LLM Top 10 and Machine Learning Security Top 10, is becoming common practice, albeit with variations in how comprehensively each institution embeds these guidelines. Most institutions identified utilizing the Office of the Comptroller of the Currency's Model Risk Management guidance to drive their underlying controls related to model risk. Some members perceive that their existing practices may already align with many aspects of these frameworks, though implemented in different guises through existing risk management policies and frameworks.

Some institutions are strengthening technical controls and initiating risk management programs, specifically tailored to address the distinctive risks presented with GenAI, particularly focusing on LLMs like "co-pilot" LLMs. These models are notably opaque, presenting challenges in terms of auditability and security observability. Their inherent complexity introduces risks beyond traditional biases, including hallucinations, toxicity, data poisoning, and difficulties in verifying accuracy. Members shared varying approaches including developing AI-specific policies and conducting thorough assessments of model risk, model security risk, and validation practices within risk and cybersecurity areas to focus on adversarial tactics related to LLMs.

Many institutions noted their reliance on a "human in the loop" approach, but this raises concerns about the varying experience levels of reviewers and the potential for them to develop a false sense of correctness in the models' outputs. Some institutions will continue to place accountability for accuracy on the users and reviewers through related training to have a balanced approach between technology and human oversight. As AI technology advances, evolving risk management practices to effectively detect and mitigate these complex and emerging risks is crucial.

Some institutions shared challenges pertaining to third party risk management, with variability in the depth of information provided by vendors regarding their testing

70 https://www.nist.gov/itl/ai-risk-management-framework.

procedures and how they address potential biases in AI models and products. To combat this, institutions are employing tools for bias detection and are investing in red team testing for LLMs, going beyond the guidance in relevant frameworks today. The sector is also evaluating resiliency concerns related to third party dependencies.

There is a growing consensus among institutions that there is a need to map various AI-related frameworks together, potentially by mapping informative references to the NIST AI RMF or creating a sector-specific profile of it. This effort highlights a significant opportunity for standardization in managing AI risks within the financial sector. By doing so, financial institutions can adopt a cohesive approach, integrating different methodologies and best practices into a clearer and more unified strategy. Such a strategy would be instrumental in mitigating the evolving threats in the financial sector, ensuring that institutions are not only compliant with established guidelines but also at the forefront of risk management in the AI landscape.

Finally, institutions emphasized that regulatory risks are a critical concern in the adoption of new AI technologies. Institutions are mindful of the balance between deploying effective new technologies while appropriately managing associated risks. However, regulatory uncertainty can slow the deployment of new technologies even when effective controls are in place.

To address this concern, a more dynamic regulatory approach is essential. Regulators should shift their focus toward overseeing comprehensive risk management strategies employed by firms. This approach would ensure adherence to stringent standards and promote swift deployment of AI tools that satisfy rigorous risk criteria. Such a balanced regulatory environment is crucial for empowering institutions to harness AI's full potential in combating sophisticated threats, without being hindered by overly restrictive oversight.

CONCLUSION/CONSIDERATIONS GOING FORWARD

In light of the advancement and challenges presented by AI in the financial sector, particularly in cybersecurity and fraud detection, this paper concludes with several considerations for navigating the evolving landscape.

- **Enhanced Cybersecurity Measures:** Financial institutions should continually refine their cybersecurity strategies to address AI-driven threats. Implementing cutting-edge AI tools for detecting and responding to threats is imperative. However, it is equally vital to maintain skilled human oversight to interpret AI data accurately and mitigate potential AI inaccuracies or biases.
- **Advanced Fraud Detection Mechanisms:** The sector must continue to prioritize the adoption of AI models for fraud prevention. Staying ahead of technologically adept fraudsters requires proactive, AI-enhanced fraud detection methods, particularly leveraging GenAI for early identification and mitigation of fraud.

- **Counteracting Adversarial AI:** Institutions need to prepare for sophisticated AI-enabled attacks, such as complex phishing and social engineering tactics. Investing in comprehensive defense systems and updating threat detection tools are essential to address the unique challenges of adversarial AI, including deepfakes and misinformation.
- **Robust Risk Management Strategies:** Aligning with frameworks like NIST AI RMF is critical. Financial institutions must strengthen their risk management protocols, focusing on emerging risks from the increased availability of AI, especially GenAI models, which includes data positioning and model biases.
- **Sector-wide Collaboration and Standardization:** The Financial sector should collaborate to develop standardized strategies for managing AI-related risk. Creating sector-specific guidelines based on AI frameworks can lead to more effective mitigation of emerging threats and ensure alignment with regulatory requirements and supervisory expectations.
- **Investment in Human Expertise:** Recognizing the irreplaceable role of human judgement in AI applications, it is crucial for institutions to invest in their workforce. Training and development programs should be implemented to equip staff with the right skills necessary for working effectively with AI technologies.
- **Risk-based Regulation:** Regulators should identify clear regulatory outcomes and objectives, while enabling regulated entities the ability to deploy effective risk management techniques based on common standards and best practices.

The evolving digital landscape presents a spectrum of unparalleled opportunities and challenges for financial institutions. It is imperative for all stakeholders across the financial sector to adeptly navigate this terrain, armed with a comprehensive understanding of AI's capabilities and inherent risks, to safeguard institutions, their systems, and their clients and customers effectively.

Annex B: External Participants

Accenture

Ally Financial

Amazon Web Services (AWS)

American Bankers Association (ABA)

Bank of America Corporation

Bank Policy Institute (BPI)

Bitsight

Capital One Financial Corporation

Chesapeake Bank

Citigroup Inc.

Darktrace Federal

Depository Trust & Clearing Corporation (DTCC)

DXC Technology Company

Financial Crimes Enforcement Network (FinCEN)

FinRegLab

Fidelity National Information Services, Inc. (FIS)

Fiserv, Inc.

Financial Services Information Sharing and Analysis Center (FS-ISAC)

Financial Services Sector Coordinating Council (FSSCC) Research and Development (R&D) Committee

Goldman Sachs Group, Inc.

Google Cloud

IBM

International Institute of Finance (IIF)

JPMorgan Chase & Co.

Lewis & Clark Bank

Mastercard Inc.

Microsoft Corporation

Moody's

National Institute of Standards and Technology (NIST)

Northwest Preferred Federal Credit Union

NVIDIA Corporation

Oliver Wyman

Optus Bank

Passumpsic Bank

PNC Financial Services Group, Inc.

Quantexa

Securities Industry and Financial Markets Association (SIFMA)

Teachers Insurance and Annuity Association of America-College Retirement Equities Fund (TIAA)

Trail of Bits

Truist Financial Corporation

U.S. Bank

Wells Fargo & Company

Glossary

Anomaly Detection – The identification of observations, events or data points that deviate from what is usual, standard, or expected, making them inconsistent with the rest of data.[71]

API – Application program interface. A system access point or library function that has a well-defined syntax and is accessible from application programs or user code to provide well-defined functionality.[72]

Business Email Compromise – A scam targeting businesses with foreign suppliers and/or businesses regularly performing wire transfer payments. These sophisticated scams are carried out by fraudsters compromising email accounts through social engineering or computer intrusion techniques to conduct unauthorized transfer of funds.[73]

Community Bank – Community banks are those that provide traditional banking services in their local communities.[74] The Federal Deposit Insurance Corporation (FDIC) lays out a test to determine if a bank qualifies as a community bank, which involves, among other things, a limited geographic profile, a limited asset size, and no more than 50% of assets in a certain specialty, such as industrial loan companies.

Core Provider – Companies that provide financial technology and services across the financial services sector in support of core banking services. Core banking services include taking deposits, making loans, and facilitating payments.

Data Integrity – The property that data has not been altered in an unauthorized manner. Data integrity covers data in storage, during processing, and while in transit.[75]

Data Integrity Attacks – An integrity attack targets the integrity of an ML model's output, resulting in incorrect predictions performed by an ML model.[76]

Data Leakage – May occur during inference attacks, where a threat actor gains access to confidential data through model inversion and programmatic querying of the model.

Data Loss Prevention – A system's ability to identify, monitor, and protect data in use (e.g., endpoint actions), data in motion (e.g., network actions), and data at rest (e.g., data storage) through deep packet content inspection, and contextual security analysis

71 IBM, *What is anomaly detection* (Dec. 2023), https://www.ibm.com/topics/anomaly-detection.

72 NIST, *Securing Web Transactions: TLS Server Certificate Management* (Jun. 2020), https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1800-16.pdf.

73 Federal Bureau of Investigations, *2021 Internet Crime Repot* (2021), https://www.ic3.gov/Media/PDF/AnnualReport/2021_IC3Report.pdf.

74 FDIC, *2020 FDIC Community Banking Study* (Dec. 2020), p. A-1, https://www.fdic.gov/resources/community-banking/report/2020/2020-cbi-study-full.pdf.

75 NIST, *Guideline for Using Cryptographic Standards in the Federal Government* (Mar. 2020), NIST SP 800-175B Rev. 1, https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-175Br1.pdf.

76 NIST, *Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations* (Jan. 2024), NIST AI 100-2e2023, https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-2e2023.pdf.

of transaction (e.g., attributes of originator, data object, medium, timing, recipient/destination, etc.) within a centralized management framework.[77]

Data Provenance – An equivalent term to "chain of custody." It involves the method of generation, transmission and storage of information that may be used to trace the origin of a piece of information processed by community resources.[78]

Data Poisoning – Poisoning attacks occur in the training phase by introducing corrupted data. An example would be slipping numerous instances of inappropriate language into conversation records, so that a chatbot interprets these instances as common enough parlance to use in its own customer interactions.[79]

Endpoint Protection – Safeguards implemented through software to protect end-user machines such as workstations and laptops against attack (e.g., antivirus, antispyware, anti-adware, personal firewalls, host-based intrusion detection and prevention systems).[80]

Explainability – An explainable AI system delivers accompanying evidence or reasons for outcomes and processes; provides explanations that are understandable to individual users; provides explanations that correctly reflect the system's process for generating the output; and only operates under conditions for which it was designed and when it reaches sufficient confidence in its output.[81]

Firewall – An inter-network connection device that restricts data communication traffic between two connected networks. A firewall may be either an application installed on a general-purpose computer or a dedicated platform (appliance), which forwards or rejects/drops packets on a network. Typically, firewalls are used to define zone borders. Firewalls generally have rules restricting which ports are open.[82]

Fourth Party – Companies that contract with an organization's third-party vendors. A security incident affecting a fourth party could result in significant supply chain issues or other business disruptions.

Generative AI – The class of AI models that emulate the structure and characteristics of input data in order to generate derived synthetic content. This content can include images, videos, audio, text, and other digital content.[83]

---

77 Committee on National Security Systems (CNSS), *National Instruction on Classified Information Spillage* (Jun. 2021), CNSSI 1001, https://www.cnss.gov/CNSS/issuances/Instructions.cfm.

78 *Ibid*.

79 NIST, Adversarial Machine Learning (footnote 72).

80 NIST, *Guide for Security-Focused Configuration Management of Information Systems* (Aug. 2011), NIST SP 800-128, https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-128.pdf.

81 NIST, Four Principles (footnote 59).

82 NIST, *Guide to Industrial Control Systems* (ICS) Security (May 2015), NIST SP 800-82 Rev. 2, https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-82r2.pdf.

83 See E.O. 14110 of Oct 30, 2023, *Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*, 88 FR 75191, https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence. Federal Register:: Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence.

G-SIB – Global systemically important bank.[84] Global systemically important banks are assessed based on size, connectedness, substitutability, complexity, and cross-jurisdictional activity. The Financial Stability Board (FSB) publishes a list of G-SIBs annually.

High Assurance Applications – Grounds for justified confidence that a claim has been or will be achieved.[85] In the context of software applications, a high assurance AI application or system reflects a high degree of confidence in the output or results produced by that system.

Intrusion Detection System – A security service that monitors and analyzes network or system events for the purpose of finding and providing real-time or near real-time warning of attempts to access system resources in an unauthorized manner.

Intrusion Prevention System – Software that has all the capabilities of an intrusion detection system and can also attempt to stop possible incidents.

Nutrition Label – Similar to a Software Bill of Materials (SBOM), a nutrition label for an AI system would clearly document what data was used to train a model, where it came from, and how any data submitted to the model will be incorporated. Attributes of a nutrition label could include data quality score, personally identifiable information score, and toxicity score.

Prompt – Natural language text describing the task that an AI should perform.

Signature-based Detection – Detection method which involves scanning software programs running on a computer system and looking for certain patterns (i.e., unique signatures, digital footprint) of computer malware or viruses by comparing to a large database of known signatures.

84 BIS, *Global systemically important banks: assessment methodology and the additional loss absorbency requirement* (Nov. 23), https://www.bis.org/bcbs/gsib.

85 NIST, *Engineering Trustworthy Secure Systems* (Nov. 22), NIST SP 800-160v1r1, https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-160v1r1.pdf.

U.S. Department of the Treasury

TREASURY.GOV