January 31, 2024

Information Technology Laboratory
ATTN: AI E.O. RFI Comments
National Institute of Standards and Technology
100 Bureau Drive
Mail Stop 9800
Gaithersburg, MD 20899-8900

Via electronic mail: ai-inquiries@nist.gov

The American Bankers Association (ABA)[1] appreciates the opportunity to comment on the request for information (RFI) related to National Institute of Standards and Technology's (NIST) assignments under sections 4.1, 4.5 and 11 of the October 30, 2023 Executive Order on Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence as published in the Federal Register on December 21, 2023 (EO).[2]

NIST is uniquely positioned to build upon the success of its AI Risk Management Framework (AI RMF)[3] by developing a complement to address the unique opportunities and challenges presented by generative AI (GenAI) and prompt-based large language models (LLMs). We urge NIST to continue to approach any new content for the AI RMF as industry-agnostic, customizable, and non-binding. At the same time, NIST should recognize the prominent role it enjoys in the ecosystem, which will only increase in the wake of the EO. The AI RMF and any supplemental materials may become an integral part of each sector's best practices, integrated into contractual provisions, and referenced by regulatory agencies. Accordingly, NIST must thread the needle to find the appropriate balance between universality and fitting into existing regulatory regimes.

In summary, the ABA:

- Encourages NIST to consider the banking industry's approach as a possible solution for other industries given that banks are at the forefront of the responsible AI movement due to the mature and flexible risk management framework at its core, which is subject to oversight by industry-focused regulatory agencies;
- Requests NIST to develop voluntary standards in harmony with regulatory requirements given that financial institutions are subject to numerous regulatory requirements with respect to the use of AI, including third-party risk management, model risk management and cybersecurity;
- Requests NIST to clarify what outputs need to be subject to AI governance;

[1] The American Bankers Association is the voice of the nation's $23.4 trillion banking industry, which is composed of small, regional and large banks that together employ approximately 2.1 million people, safeguard $18.6 trillion in deposits and extend $12.3 trillion in loans.
[2] https://www.federalregister.gov/documents/2023/12/21/2023-28232/request-for-information-rfi-related-to-nists-assignments-under-sections-41-45-and-11-of-the
[3] https://www.nist.gov/itl/ai-risk-management-framework

- Encourages NIST to define "red-teaming" given that the term is used in many ways and lack of clarity could raise concerns;
- Applauds efforts to reduce the risk of synthetic content given GenAI can generate synthetic content leading to the creation of deepfakes that can be used to perpetrate fraud and causing mis- and dis-information but also encourages NIST to recognize the positive uses of synthetic data to advance privacy and trade secret protection and ensure it does not inadvertently prevent GenAI's development; and
- Generally supports efforts to advance global technical standards that accommodate the risk management needs (including regulatory compliance requirements), that are technology neutral, risk-based, and tailored to use cases.

Financial Sector Assessment of AI on Cybersecurity and Fraud

The ABA and our member banks have been actively assessing the impact of AI, especially the use of GenAI. The ABA established two working groups in 2023 that focus on AI opportunities and risks. The AI Working Group is an interdisciplinary group of data scientists, technologists, compliance, legal, risk management, security, and human resources experts. For the past six months the working group has been examining opportunities and risks across the spectrum of AI applications generally and to help formulate the ABA's policy positions. Meanwhile, the ABA AI Cyber and Fraud Subgroup has been looking at strategies banks employ to counter AI risks, including the use of "deepfake" voices to defeat voice recognition authentication systems and deepfake images to defeat biometric controls. In the Fall of 2023, the ABA organized a series of meetings with financial sector experts on behalf of the Financial Services Sector Coordinating Council's R&D Committee. Collectively, these efforts reveal the following:

- Financial institutions have a long history of deploying and controlling risk related to novel technologies. This is true with AI and machine learning, where institutions have utilized graph-based analytics to, among other use cases, detect anomalous and suspicious activity and flag for further investigation. The technologies 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. Many vendors now integrate AI/machine learning technology into commercially available solutions that are widely used across the sector.
- 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.
- The effectiveness of AI underscores the necessity for rapid and safe implementation of AI technologies in the financial sector. 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 financial 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.[4]

- 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, jailbreaking, and difficulties in verifying accuracy.
- Some institutions shared challenges pertaining to third-party risk management, with variability in the depth of information provided by vendors regarding their testing 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.
- 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.

The banking industry is at the forefront of the responsible AI movement due to the mature and flexible risk management framework at its core, which is subject to oversight by industry-focused regulatory agencies. This is a model that may inform other industries' usage of AI (including GenAI), and NIST should consider banks' approach as a possible solution for other industries.

Regulatory Considerations

ABA appreciates NIST's role in developing standards for use by multiple industries and governmental agencies, including regulators that rely on NIST standards. We recognize that NIST develops standards that are industry agnostic and customizable. Given that US financial regulators rely on NIST standards in constructing regulatory and supervisory requirements it is

[4] The three lines of defense for risk management are the business line, second line corporate risk management, and a third line of auditing risk controls. With this model, the business line is responsible for managing risk associated with its AI systems. The second line provides risk management structures to support the business lines managing AI-specific risk and enables risk related communications and decisions about AI system use up through the board. The third line audit assures the right monitoring, controls, and reporting are in place to manage the context specific risks posed by artificial intelligence.

important that NIST avoid being too prescriptive but offer examples of various risk profiles in order to help banks apply the standards more effectively.

ABA encourages NIST to take into account that financial institutions are subject to numerous regulatory requirements with respect to the use of AI, including third-party risk management, model risk management and cybersecurity. Federal financial regulators have recently reminded the industry in public remarks that AI technologies like any new technology utilized by a financial institution must be done so in a way that complies with existing law.[5] Senior officials from the Federal Reserve, FDIC, OCC and CFPB stressed that banks are ultimately responsible for how AI technology is deployed, even if they are contracting with third parties to provide AI-powered products and services. They emphasized that financial institutions must understand what model risk management is and how they are expected to conduct it.[6] These agencies have released statements and guidance, which underscore their positions.[7] These issuances build upon a strong foundation of model risk management expectations that have been in place for over a decade.[8]

In general, the ABA believes that existing laws and regulatory guidance that apply to banks are sufficient to cover the risk associated with the potential outcomes of AI technology, particularly if the existing laws and guidance are technology agnostic. However, some requirements and guidance may need to be reassessed to ensure they continue to provide a practical risk-based approach to governance in the context of GenAI. We encourage NIST to create specific risk profiles for the AI RMF that allows financial institutions to 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.

Further, ABA encourages NIST to recognize that different organizations have different levels of maturity. Accordingly, the AI RFM should provide for different stages of sophistication, ranging from a minimally viable product to an optimal endgame. Finally, NIST should continuously update its glossary and validate that the terminology is used consistently in future releases. This will aid in understanding and drive operationalization.[9] The need for a common lexicon extends even to such basic issues as the definition of GenAI and traditional AI.

---

[5] https://www.politico.com/newsletters/morning-money/2024/01/22/law-order-ai-edition-00136879
[6] Id.
[7] See, e.g., https://files.consumerfinance.gov/f/documents/cfpb_joint-statement-enforcement-against-discrimination-bias-automated-systems_2023-04.pdf; https://www.consumerfinance.gov/data-research/research-reports/chatbots-in-consumer-finance/chatbots-in-consumer-finance/; and https://www.consumerfinance.gov/compliance/circulars/circular-2023-03-adverse-action-notification-requirements-and-the-proper-use-of-the-cfpbs-sample-forms-provided-in-regulation-b/.
[8] See https://www.federalreserve.gov/supervisionreg/srletters/sr1107.htm; https://www.occ.gov/news-issuances/bulletins/2011/bulletin-2011-12a.pdf; and https://www.fdic.gov/news/financial-institution-letters/2017/fil17022.html.
[9] See https://airc.nist.gov/AI_RMF_Knowledge_Base/Glossary.

GenAI Governance

NIST must recognize the different roles of participants in the AI ecosystem: namely, developers, deployers, and end users. Developers are often "Big Tech" companies but may include the largest banks. Deployers utilize models created by third party developers, and most banks will fit into this category. End users are the ultimate audience for the models; in the case of so-called traditional AI, this will most commonly be financial services employees due to the technical barriers to entry, but for GenAI end users may include customers since GenAI applications are prompt-based and thus more accessible to laypeople.

In developing standards and guidelines, NIST should recommend that AI developers adhere to responsible and ethical design principles and frameworks in the design phase while maintaining appropriate oversight and governance during the design of AI technology to minimize bias, protect privacy, and promote inclusive representation. NIST should consider guidelines on testing that can be used while building an AI model to reduce bias. To the extent developers of AI technologies make their models commercially available and not available through open source, they should be required to provide sufficient information to allow the deployers to conduct appropriate due diligence. The AI developer should be clear about how their AI system is trained, fine-tuned and maintained (i.e., the type and source of the data that was used to train and fine-tune the model, as well as any restrictions on the data to prevent allegations of collusion through AI systems).

While developers of AI systems will need to provide information regarding their systems to deployers, deployers themselves should be held to certain standards relating to transparency, third party risk management, and implementing safeguards to maintain fairness and safety and to manage against discrimination and bias in any AI implementation. Doing so will ensure that deployers are held accountable alongside developers.

Deployers and developers alike will need to ensure that the AI being used is explainable, as explainability is a necessary aspect of transparency. It requires that developers and deployers, to the extent possible, meaningfully explain how AI systems affect decisions and outcomes that impact individuals, while bearing in mind trade-offs between explainability and accuracy. Any explainability requirements should consider a meaningful and accessible approach for explaining how a decision using AI (or any other technology) was made so that an average consumer would be able to comprehend. Careful consideration of which contexts require an explanation or redress is imperative so as not to limit innovation or over-saturate consumers.

Where individuals believe they have been harmed by AI, deployers and developers should create clear options for inquiries, complaints, and redress, where appropriate and necessary. There are existing explainability and redress laws whereby certain decisions, such as decisions not to extend credit, need to be explained to the individual. Those legal requirements are technology agnostic and should not be expanded simply as a result of the technology used to arrive at the outcome.

Any standard should strike a balance between transparency and trade secret laws and deployers' right to protect their intellectual property. Model cards may be a potential solution, as they allow

information to be summarized at a high level in such a way that does not divulge confidential commercial information. However, NIST would need to create a template for model cards to standardize the format to make the information easily digestible and navigable.

Additionally, the ABA encourages NIST to clarify what outputs need to be subject to AI governance. While not all outputs need to be subject to the same governance, it is important to create policies and procedures tailored to the use case and associated risk appetite. For example, customer-facing products should be subject to far stronger controls than models to predict whether a server will remain operational. The framework should not be overly prescriptive and should remain customizable, as context is key. Banks should document any decisions thoroughly and be prepared to explain their rationale to regulators.

Cross-functional Collaboration & Departmental Roles

GenAI requires a cross-functional governance approach to identify and mitigate risk. In the financial services space, participants should include first line, second line, and third lines. It is essential that the first line be involved and have ownership over the resulting risk. Moreover, the process should constantly iterate and mature based on continuous monitoring and periodic holistic reviews. This interdisciplinary focus should apply to both the inputs as well as the outputs of models or neural networks.

While GenAI does not necessarily require a separate system from traditional iterations of AI, it does introduce some novel risk domains that should be identified, assessed, and addressed (for example, privacy & consents, hallucination, and intellectual property). Moreover, the cross-functional team must be diverse to account for different perspectives, and it is important to include both technologists and non-technologists.

Validation Process

The ABA encourages NIST to create guidance and benchmarks for evaluating and auditing AI applications, with a focus on capabilities and limitations through which AI could be used to cause harm. The ABA believes any AI standard that contemplates human oversight requirements should be pragmatic, considering: (1) the fact that the technology continues to evolve; (2) there are different practicalities associated with using AI models; and (3) there are different contexts in which AI is being used.

Human oversight requirements should not be overly prescriptive because different AI use cases will benefit from differing controls based on the particular risks posed. Hence, the deployer of such use case would be in the best position to determine the most appropriate controls. For example, with more traditional supervised AI data models where the model will perform in the manner in which it is designed, having a human examine the output routinely will not yield better results. In that case, back-testing or prospective readiness testing and governance may be a more appropriate risk management tool.

Another example is in the use of AI in facial recognition and other biometric models. These models use neural networks (e.g., computer vision) to authenticate and identify individuals and,

in some cases, to promote compliance with Know Your Customer laws and regulations. Given issues with accurately identifying and authenticating some individuals, it is important that financial institutions rigorously test these facial recognition AI models before they are deployed for identification and/or authentication purposes.

Moreover, if there is a concern that the model will not perform in the manner in which it was designed (e.g., the AI technology hallucinates), more periodic or routine human oversight over the output may be advisable. For organizations that adhere to the principles of software development lifecycle (SDLC) standard, AI/machine learning models have controls in place regarding model degradation (e.g., due to data drift). These include continuous feedback models with model retraining and hyperparameter tuning. In addition, any standard should encapsulate that AI technology itself can be useful to validate the reliability of other AI applications, thereby acknowledging that such AI-assisted reviews may present certain advantages to human oversight.

Red-teaming

The EO directs NIST to establish guidelines to enable developers of AI to conduct AI "red-teaming" to enable deployment of safe, secure and trustworthy systems. The ABA encourages NIST to define "red-teaming" given that the term is used in many ways and lack of clarity could raise concerns and lead to confusion. NIST must provide clarity around traditional meaning of red-teaming, which is often used to assess threats as opposed to causal impact or harm, and avoid repurposing terms if possible and thereby define "red-teaming" in the context in which it is offered.

AI red-teaming can be a very important tool in the toolbox for GenAI, due to the "black-box" nature of GenAI algorithms. Developers or providers of GenAI are reluctant to provide evidentiary artifacts that ensure that training data was not poisoned, no trojans were embedded, or that data sets are free of biases and toxicity. Further, AI developers and providers currently only supply a responsible AI disclosure policy, which falls short of the necessary cybersecurity controls. Like using penetration testing as a security control, any red team activity is a point in time exercise designed to identify vulnerabilities and test the effectiveness of the controls in place.

In the context of GenAI, red-teaming could be a valuable control depending on the AI system being tested and the design and execution of the red team activities. If used, however, certain best practices and guidance for conducting red team activities can be valuable, similar to penetration testing design in the security context, but a rigid definition of what will be tested and how such testing shall occur will not be as valuable as an organization identifying the potential risks and then designing testing/red-teaming to test the controls and study the operation of the GenAI system. To the extent that we can align on best practices and guidance for conducting these red team activities, organizations implementing the GenAI systems can then automate the red team activities, using them to automatically attack or test vulnerabilities in target GenAI systems.

Crucially, the ABA encourages NIST to avoid "overselling" what can be done via red-teaming. For example, red-teaming might only be for outputs of GenAI systems, rather than the inputs. This might be framed as: if full scale validation is not feasible, then red-teaming is an acceptable alternative.

The scope of red-teaming activities should be clarified given the range of possible activities, such as jailbreaking. While red-teaming may be a technique that is beneficial in certain instances, it should be considered one of a multitude of controls that organizations can utilize to help identify risks associated with GenAI. For traditional AI models, where the training data and the outputs are largely conducted under a controlled environment with outcomes that are certain, red-teaming may not be as useful in detecting risks. Importantly, the red team does not necessarily need to be housed within the second line. All that is required is independence. For example, it could be conducted through a peer that sits in a different part of the organization.

Synthetic Content

The RFI outlines actions NIST is considering taking to reduce the risk of synthetic content. The ABA applauds this effort given that GenAI can generate synthetic content leading to the creation of deepfakes that can be used to perpetrate fraud and cause mis- and dis-information. NIST should consider standards for governance around inputted data (e.g., origin of data, ensuring data meets regulatory requirements such as General Data Protection Regulation, cross-border data clearance, and Anti-Money Laundering).

However, not all synthetic data is problematic. Synthetic data can be extremely valuable from a privacy and trade secrets protection standpoint. In the context of synthetic data, organizations using such data to train models should consider the impact on model performance on AI models. For example, in traditional AI, sensitive data can be tokenized in a format preserving way, as an alternative to using synthetic data and thus meaningfully de-risk the use of the data. Other privacy preserving and enhancing techniques (PETs) (such as pseudonymization and differential privacy) and privacy by design elements can be deployed to reduce risks inherent in using non-synthetic data. In other words, synthetic data is just one of many PETs that can achieve similar results.

NIST should be careful about recommendations on disclosures of the use of synthetic data to ensure they are carefully drawn and narrowly tailored so it does not encourage excessive consumer notifications, which can result in burnout (e.g., cookie disclosures).

We agree that there should be a focus on protecting society from harms that can result from synthetic data and deepfakes and that this would benefit from being a global effort. We encourage the US government to help by investing resources in technical capabilities to detect and alert users to synthetic content, particularly on social media and media sites, and to ensure that these platform owners are taking reasonable measures to prevent synthetic content from causing harm. The ABA agrees that this is a key risk that needs to be resolved on a multi-jurisdictional basis. We encourage the government to help by investing in technical capabilities to detect and alert users to synthetic content where notice can mitigate potential negative impact to consumers.

Global Standards

Finally, the RFI asks whether NIST should advance responsible global technical standards for AI. The ABA generally supports efforts to advance global technical standards that accommodate the risk management needs (including regulatory compliance requirements), that are technology neutral, risk-based, and tailored to use case. It is important that global technical standards allow multinational banks to do business and third-party providers to operate in multiple jurisdictions. However, the standards must not be overly prescriptive or rigid, which would simply stifle innovation without commensurate consumer or ecosystem benefit. We are advocating for the development of industry standard and voluntary risk-management frameworks to ensure fairness and safety in the development and use of AI that is designed to minimize bias and promote inclusive representation.

In addition, we support the creation of internationally recognized and interoperable standards, through the OECD, G7, TTC, or other international organizations, that can serve as guideposts for regulators and promote global harmonization. We support research and development efforts by governmental entities for the study of AI risk management, and adherence to responsible and ethical design.

Conclusion

The ABA and our members appreciate NIST's efforts to implement the mandates outlined in the President's EO on AI. We encourage NIST to consider the banking industry's approach as a possible solution for other industries given that banks are at the forefront of the responsible AI movement due to the mature and flexible risk management framework at its core. We request NIST that develop voluntary standards in harmony with regulatory requirements given that financial institutions are subject to numerous regulatory requirements with respect to the use of AI, including third party risk management, model risk management and cybersecurity. We request that NIST clarify what outputs need to be subject to AI governance and define "red-teaming", given that the term is used in many ways and lack of clarity could raise concerns. We applaud efforts to reduce the risk of synthetic content given that GenAI can generate synthetic content leading to the creation of deepfakes that can be used to perpetrate fraud and cause mis- and dis-information. We also encourage NIST to recognize the positive uses of synthetic data to advance privacy and trade secret protection. Finally, we support efforts to advance global technical standards that accommodate the risk management needs (including regulatory compliance requirements), that are technology neutral, risk-based, and tailored to use cases.

We look forward to continued engagement with NIST, others in the Administration, Congress, and other stakeholders on the seminal issue of GenAI usage and governance. If you have any questions about this comment letter, please contact John Carlson, Senior Vice President, Cybersecurity Regulation and Resilience (jcarlson@aba.com) or Ryan Miller, Vice President & Senior Counsel, Innovation Policy (rmiller@aba.com).