MITRE

# Request for Information and Comment on Financial Institutions' Use of Artificial Intelligence, Including Machine Learning

## Response to Agency/ID:

Board of Governors of the Federal Reserve System:
Docket No. OP-1743

Bureau of Consumer Financial Protection:
Docket No. CFPB-2021-0004

Federal Deposit Insurance Corporation:
RIN 3064-ZA24

National Credit Union Administration:
Docket No. NCUA -2021-0023

Office of the Comptroller of the Currency:
Docket ID OCC-2020-0049

June 24, 2021

# Table of Contents

# Introduction

The MITRE Corporation is pleased to respond to the “Request for Information and Comment on Financial Institutions' Use of Artificial Intelligence, Including Machine Learning” released on behalf of the Board of Governors of the Federal Reserve System (Docket No. OP-1743), Bureau of Consumer Financial Protection (Docket No. CFPB-2021-0004), Federal Deposit Insurance Corporation (RIN 3064-ZA24), National Credit Union Administration (Docket No. NCUA -2021-0023), and Office of the Comptroller of the Currency (Docket ID OCC-2020-0049).

As a not-for-profit organization, MITRE works in the public interest across federal and state, local, tribal, and territorial governments, as well as industry and academia. We operate federally funded research and development centers (FFRDCs) established under Federal Acquisition Regulation (FAR) part 35.017—unique organizations that are chartered to assist the United States government with scientific research and analysis, big data analytics, artificial intelligence (AI) and machine learning (ML), development and acquisition, and systems engineering and integration. Through our FFRDCs and public-private partnerships, we work across government and in partnership with industry to tackle challenges to the safety, stability, and well-being of our nation.

As the operator of the U.S. Department of Treasury’s FFRDC, we work with the Securities and Exchange Commission, National Institute of Standards and Technology, Treasury, and other key stakeholders to mitigate cyber risks and engineer faster, more secure payments to and from every American. We provide trusted data for policy makers, businesses, and citizens, and advance the federal financial and economic regulatory and intelligence ecosystem. Working closely with our specific agency partners, we are helping to transform fiscal policy and financial management by collecting, analyzing, and increasing accessibility to data that impacts American finance, trade, and commerce.

MITRE has a long history of partnering with federal agencies to apply the best elements of AI and ML, while developing and supporting ethical guardrails to protect people and their personal data. Our team is committed to anticipating and solving future needs that are vital to the success and safety of the public and the country.

MITRE values the opportunity to contribute to this important discussion. We are also eager to engage further with this community of regulators and financial institutions. In the following pages, we have focused on answering five specific questions from the RFI:

- Question 6 – Overfitting
- Question 7 – Cybersecurity Risk
- Question 8 – Dynamic Updating
- Question 12 – Fair Lending
- Question 16 – Other Considerations

# Response to Question 6 (Overfitting):

*How do financial institutions manage AI risks relating to overfitting? What barriers or challenges, if any, does overfitting pose for developing, adopting, and managing AI? How do financial institutions develop their AI so that it will adapt to new and potentially different populations (outside of the test and training data)?*

***Opportunities exist for financial institutions to mature their AI processes, and avoid or mitigate many of the risks associated with overfit before, and during, deployment.***

---

Overfitting data sets while training models can cause them to underperform on operational data and increase false positives, false negatives, or both. This could have implications on loan determination processes, leaving an institution's portfolio riskier than intended, or causing loans to be improperly denied in a discriminatory manner. Fraud detection and cybersecurity systems suffering from overfitting could miss suspicious activity or identify too many false positives resulting in unnecessary costs from added analyst investigations, and analysts could miss actual incidents due to *alarm fatigue*. This could result in increased identity theft, data theft, ransomware attacks, and financial losses. In the realm of customer service, overfitting could lead to customer dissatisfaction, loss of consumer confidence, and degraded or discriminatory consumer decision-making.

Opportunities exist for financial institutions to mature their AI processes, and avoid or mitigate many of the risks associated with overfit before, and during, deployment. A proactive approach that is enabled by mature processes is important because regulators, watchdogs, and the public would not have the frame of reference to detect most indicators of overfitting without intrusive and expensive inspection regimes and data disclosures. The effects of overfitting may eventually become observable, but likely after some harm has been sustained, and unnecessary organizational and societal costs are realized.

There are numerous approaches for measuring an institution's AI maturity including the Government Services Agency's AI Capability Maturity Model[1] and MITRE's Organizational AI Maturity Model, which is discussed in our response to question 16. Institutions can assess and increase their AI maturity through organic capabilities, contracted capabilities, or through enhanced support from AI-enabled application vendors. For example, in the MITRE AI Maturity Model, a highly AI-mature entity will institutionalize the following:

[1] https://coe.gsa.gov/2020/10/28/ai-update-2.html

- **AI strategy** and will collaborate with other organizations to understand emerging problems and opportunities to collectively improve AI initiatives
- **Structures, roles, and processes** that are codified across the organization, using metrics to improve outcomes
- **Workforce development** and manage gaps to ensure capability and competency to address current and future needs
- **Test and evaluation standards and processes** for development, testing, and deployment that continuously improve AI performance and alignment with enterprise objectives
- **Sufficient AI infrastructure**, incorporating monitoring and analytics for continuous evaluation and improvement
- **Enterprise-wide data governance framework** to continuously monitor and improve data interoperability, quality, and compliance
- **Due diligence process** to identify, analyze, and learn from incidents.

A regulator may be able to expand industry-wide AI maturity by providing financial institutions with incentives (e.g., regulators could extend a certain degree of cooperation to supervised institutions facing examination for unintended outcomes instigated by AI-enabled applications) to advance their respective maturity levels. AI-mature financial institutions should be able to harness their AI capabilities to lower operating costs, improve customer satisfaction, lower risk, and improve profitability. Institutions that choose to deploy AI-enabled applications but do not invest in achieving a high level of AI maturity would remain fully accountable for any potential negative outcomes and should not expect regulatory cooperation.

## Response to Question 7 (Cybersecurity Risk):

*Have financial institutions identified particular cybersecurity risks or experienced such incidents with respect to AI? If so, what practices are financial institutions using to manage cybersecurity risks related to AI? Please describe any barriers or challenges to the use of AI associated with cybersecurity risks. Are there specific information security or cybersecurity controls that can be applied to AI?*

***Regulations should reward organizations that seek to collaborate effectively to build dynamic and responsive AI security frameworks that facilitate the sharing of threat intelligence and vulnerabilities.***

---

AI and ML systems represent a unique and rapidly expanding attack surface with associated risks not addressed by traditional cybersecurity controls and frameworks. Unlike traditional cybersecurity vulnerabilities that are caused by “bugs” or human error in code, vulnerabilities specific to AI are enabled by inherent limitations in the underlying AI algorithms. They are not tied to specific software development or deployment stacks,

so they cannot be patched in a traditional cybersecurity sense. As a result, data can now be weaponized in new ways, which will drive changes in how data is collected, stored, and applied.

AI and ML are advancing quickly, and new applications are being developed at such fast speeds, that lines are often blurred between research and production-ready systems. In the same way AI specific vulnerabilities continue to evolve rapidly in both the steps necessary to weaponize the vulnerability and the resulting effects the attack can achieve. These attacks can be highly effective in exfiltrating sensitive consumer data, stealing intellectual property, or in subverting the AI system for malicious purposes, even when the AI models are effectively secured using traditional cybersecurity measures.

These AI specific vulnerabilities can be weaponized to achieve a broad range of effects from targeted attacks that attempt to subvert AI-enabled facial recognition systems to gain malicious access to user accounts and financial data, intentionally poisoning data being used to train algorithms for trading on the stock market, and defeating fraud detection algorithms to allow illicit financial transactions to occur. As financial markets become more heavily integrated and automated, the number of transactions governed solely by algorithms will grow. Although this level of automation leads to significant benefits in the form of decreased costs, increased customer convenience, and more efficient capital allocation, it also dramatically increases the potential damage malicious actors can cause to domestic and global markets.

Increased automation creates a unique challenge for regulators as these AI-specific vulnerabilities are highly dynamic, evolving rapidly, and they are often inseparable from the significant benefits of these technologies. This causes a dynamic where overly restrictive regulations focused on AI security could quickly stifle beneficial innovation without enhancing security. It will be challenging for regulators to avoid situations where underlying technology and associated attacks rapidly outpace regulator efforts, leaving institutions implementing policies to satisfy audits of compliance instead of focusing on addressing legitimate security concerns.

Rather than relying on prescriptive regulations, regulatory bodies could encourage AI security with positive incentives. Regulations should reward organizations who seek to effectively collaborate to build dynamic and responsive AI security frameworks that facilitate sharing threat intelligence, information about vulnerabilities, and community-developed best practices and mitigations.

Both government and industry are challenged to provide an effective framework to incentivize AI-Security incident avoidance while managing the tension between the use of AI to maintain innovative as well as competitive advantage and the responsibility for secure AI-implementation. Given the rapidly expanding number of AI specific attacks and their difficulty to detect, organizations, who have demonstrated sufficient preventative measures to avoid cyber incidents and reasonable responsive actions, should not be overly penalized when an attack inevitably occurs.  Regulators should

consider implementing policies and practices that incentivize financial institutions to report these attacks by eliminating potential fines or other punitive measures that would have otherwise been incurred conditional to rapid and open reporting. This would give the financial community freedom to innovate and to share AI security best practices. The collective security of AI-driven financial technologies would also increase.

This community-based collaboration and best practice development should be implemented in the form of public and private partnerships structured to support the rapid identification of vulnerabilities, development of mitigations, and their public release. A working model for this type of public-private partnership is the MITRE Adversarial Threat Landscape for Artificial-Intelligence Systems (ATLAS). MITRE ATLAS was launched in partnership with Microsoft and a broad range of participating organizations to collect real-world, voluntarily submitted cases, and to study them using a robust and common taxonomy to empower security analysts across industry and within the government to detect, respond, and remediate threats against ML systems.

Released in Fall 2020, ATLAS has already impacted AI security across multiple industry verticals by encouraging security teams from diverse companies such as Bosch, Ant Financial Group, and Airbus to apply the framework to evaluate their systems' AI security. These companies are contributing their experiences and results as case studies to further improve ATLAS. This collaboration also resulted in Microsoft releasing a powerful open-source tool set, *CounterFit,*[2] based on ATLAS. This is a significant step toward mutual security as it gives companies that cannot afford dedicated AI security practitioners with a robust ability to evaluate their own AI-enabled systems.

MITRE ATLAS highlights strong community-based collaboration already taking place to rapidly understand, detect, and mitigate these threats; however, it is not the only way forward. It should serve as a prime example of how a rapid response to evolving AI threats can be achieved without excessive regulation. Regulators should consider policies and practices that reinforce and support similar types of collaboration, which lead to more effective protection of consumer data and intellectual property, reduce the cost of securing critical financial systems, and allow the U.S. financial services industry to focus on delivering high quality, fair, and secure technologies that benefit U.S. consumers and reinforce the U.S. as a global financial leader.

## Response to Question 8 (Dynamic Updating):

*How do financial institutions manage AI risks relating to dynamic updating? Describe any barriers or challenges that may impede the use of AI that involve dynamic updating. How do financial institutions gain an understanding of whether*

[2] AI security risk assessment using Counterfit - Microsoft Security

## *AI approaches producing different outputs over time based on the same inputs are operating as intended?*

***Similar to our response to Question 6 regarding overfit - Opportunities exist for financial institutions to mature their AI processes, and avoid or mitigate many of the risks associated with dynamic updating before, and during, deployment.***

---

MITRE's view on how the financial services industry can manage the risks associated with dynamic updating of models aligns with the views described in MITRE's response about overfitting in question 6. There are significant similarities in terms of the inaccessible frame of reference required for regulators to detect changes in model behavior, the negative consequences that can occur if those changes are not managed properly, and the limited mechanisms available for regulators to influence financial institutions to detect and correct problems with their models.

The performance of models can decay or drift over time as new data diverges from the data that was used to train them. For instance, a model to forecast a consumer's credit use will likely perform worse during an economic downturn. There are several types of model decay. An organization with a high level of AI maturity will understand how to detect and mitigate each type of model decay. Typically, this involves retraining models to account for new data. Some types of AI can automatically update models based on the most recent data, such as reinforcement learning attempts to maximize a reward function by reducing the error between predictions and actual values. Although this dynamic updating can reduce drift by learning from the most recent data, it can also create models that have some altered characteristics. An organization with a high level of AI maturity will be continuously monitoring, evaluating, and adjusting models to ensure they fulfill their intended purpose. There are numerous approaches for measuring an institution's AI maturity including the Government Services Agency's AI Capability Maturity Model and MITRE's Organizational AI Maturity Model, which is discussed in our response to question 16.

Like the challenges of monitoring overfitting, regulators must establish a cohesive strategy and regulatory framework to properly balance and leverage the technologically derived benefits and the burdensome mitigation of technologically induced risks. One possible solution is that regulators could incentivize and empower financial institutions to detect and solve these problems themselves with appropriate reporting and examination responsibilities in place. Specific to AI, regulators may wish to consider the AI maturity of financial institutions and provide incentives such as reduced inspection frequency and greater cooperation for institutions that exhibit high levels of AI maturity. Although the industry is not currently positioned to effectively use an AI maturity model in this manner, it will be important for regulators to encourage a common model and for financial institutions to begin reporting their data to create a baseline for future comparison.

# Response to Question 12 (Fair Lending):

*What are the risks that AI can be biased and/or result in discrimination on prohibited bases? Are there effective ways to reduce risk of discrimination, whether during development, validation, revision, and/or use? What are some of the barriers to or limitations of those methods?*

***Developers of consequential ML systems should emphasize fairness risks across the lifecycle.***

---

There are clear demonstrations that AI can result in discrimination against protected classes.[3] Concern about unfair algorithmic outcomes in the financial sector is also the focus of international attention;[4] worry about algorithmic unfairness in sophisticated predictive models (without AI) long used in the financial sector was already a regulatory focus (including for example Model Risk Management [MRM] guidance such as SR 11-7). How does adopting AI and ML change the nature and scope of the risks and the methods for mitigation and oversight, and change the expectations to be established for financial institutions? The rest of this section focuses on ML because it is widely used in rapidly advancing AI capabilities. ML techniques involve building a system that "learns" (i.e., develops specific algorithms) from exposure to data or to feedback, so it can address problems that are too complex to be explicitly coded by developers. This can also make ML very brittle if it experiences data or circumstances outside of its training (see the discussion of overfitting), and it often leaves the specific logic that has been implemented unclear. This places added constraints and demands on oversight and assessment across the lifecycle, shifts more of the burden of establishing justified confidence in the system to runtime scrutiny and analysis, and undermines some of the assumptions of traditional test and evaluation. These challenges are an issue for assuring fairness in ML systems.

There is a need for developers of consequential ML systems to emphasize fairness risks across the lifecycle, what MITRE calls "left of algorithm, right of result."

- "Left of Algorithm" refers to the steps needed well before model training begins:

[3] https://science.sciencemag.org/content/366/6464/447 and http://faculty.haas.berkeley.edu/morse/research/papers/discrim.pdf
[4] https://www.mas.gov.sg/schemes-and-initiatives/veritas and http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/science-and-technology-committee/algorithms-in-%20decisionmaking/oral/77536.html

  - Establish stakeholder consensus on the definition of fairness for this specific application,[5] then identify derived requirements and metrics for the development.
  - Use hazard analysis techniques (e.g., System Theoretic Process Analysis) to identify risks of unfairness in the sociotechnical system in which the model will be used.
  - Identify a priori key aspects of model performance and the sociotechnical impact of the model prior to development so that proper instrumentation of the system and its effects can be part of the design.

- "Right of Result" refers to vigilance in ensuring the model in operation is not contributing to inequitable outcomes:
  - Testing for fairness before use is necessary, but not sufficient; system instrumentation, monitoring, and analysis (IMA) is required.

    *"Confidence in … AI systems requires insight into the development process and the operational performance of the system. Insight into the development process is supported by capturing decisions and development artifacts for review; insight into operational performance is supported by runtime instrumentation and monitoring to capture details of execution."*[6]

  - The need for runtime assurance and IMA to complement pre-release testing for fairness in ML applications is a logical extension of MRM revalidation and challenge requirements.

    *"It is essential to evaluate whether changes in products, exposures, activities, clients, or market conditions necessitate adjustment, redevelopment, or replacement of the model and to verify that any extension of the model beyond its original scope is valid."*[7]

In some cases, the legal and policy approaches to mitigating protected class discrimination designed to counter human bias are poorly suited for ML models. For example, in some instances, removing any reference to a protected class in training and operational data is counterproductive and can add to disparity.[8] Addressing this may require changes to policy and requires closer analysis as AI adoption in regulated institutions increases. The issue is made clear in an analysis of the tension between technical methods and legal constraints:

> **Tennessee Law Review Article: *Reconciling Legal and Technical Approaches to Algorithmic Bias***

[5] There are many definitions of algorithmic fairness, some of which are mutually exclusive and require agreement on the operational definition to inform algorithm development (e.g., 1, 2, 3).
[6] See page 380 of https://www.nscai.gov/wp-content/uploads/2021/03/Full-Report-Digital-1.pdf
[7] SR 11-7
[8] https://lawreviewblog.uchicago.edu/2020/10/30/aa-ho-xiang/

*The key issue this Article tackles is the extent to which well-meaning algorithm developers can use protected class variables to address algorithmic bias in light of existing anti-discrimination law jurisprudence. The key risk that this Article seeks to mitigate is the possibility that technical and legal approaches to mitigating bias will diverge so much that laws prohibiting algorithmic bias will fail in practice to weed out biased algorithms and technical methods designed to address algorithmic biases will be deemed illegally discriminatory.*[9]

# Response to Question 16 (Additional Considerations):

*To the extent not already discussed, please identify any additional uses of AI by financial institutions and any risk management challenges or other factors that may impede adoption and use of AI.*

***Regulators may be able to promote improved, industry-wide AI maturity by providing supervised financial institutions with incentives such as increased regulatory cooperation on unintended outcomes involving AI-enabled applications.***

---

Financial Institutions with a high level of AI maturity are more likely to apply AI effectively and without avoidable bias. Mature institutions promote a stable and resilient financial system and can more easily demonstrate their compliance with applicable standards and regulations.

There are numerous approaches for measuring an institution's AI maturity, including the Government Services Agency's AI Capability Maturity Model and MITRE's Organizational AI Maturity Model (AI MM).

The MITRE Organizational AI MM and its associated AI Assessment Tool (AI AT) serve to assess and guide effective readiness, adoption, and use of AI/ML across an organization. The AI MM defines the dimensions and levels of AI maturity and provides the foundation for an assessment using the AI AT.

The AI MM identifies five broad types of obstacles, known in the model as "pillars," which can impede the adoption and effective use of AI:

- **Strategy and Budget**: Does the organization have an implementation plan, partnerships with other agencies, and governance processes?

[9] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3650635

- **Organization**: Does the organization have a risk-tolerant culture that supports innovation, defined roles for AI development, and a plan for recruiting, training, and retaining AI talent?
- **Technology Enablers**: Does the organization have an approach for identifying and using new AI innovations, a method for verifying and validating proposed solutions, and the compute power needed to develop, deploy, and maintain these solutions?
- **Data**: Does the organization have data governance processes and audit capabilities in place to monitor compliance with AI/ML standards, for sharing and for the appropriate storage, retention, and access control?
- **Performance and Application**: Does the organization have an approach for integrating AI into business workflows, have monitoring processes in place to measure how well they support strategic outcomes, and have a due diligence process that promotes calibrated user trust and protects against unintended consequences in AI solutions?

Operationalizing AI and ML is not easy. Although many enterprises initiate AI and ML projects, the results often fall short of expectations, highlighting the need for organizations to prepare for AI and ML initiatives.

> *"…nearly eight of 10 organizations engaged in AI and machine learning said that projects have stalled, according to a Dimensional Research report. The majority (96%) of these organizations said they have run into problems with data quality, data labeling necessary to train AI, and building model confidence."*[10]

The combined MITRE Organizational AI Maturity Model and Assessment Tool can help organizations see more than one aspect of AI adoption, and provide a systematic path to success. It can provide the framework for understanding and a roadmap for success.

[10] 96% of organizations run into problems with AI and machine learning projects. Macy Bayern. May 24, 2019.

# MITRE Author and Contact Information

*Please contact the author below to share your feedback or to learn more about MITRE's Regulation and Artificial Intelligence/Machine Learning work.*

---

Corresponding Author:

Matthew Dinger, Finance, Regulation, and Economics Department Manager
mdinger@mitre.org