

PENN PROGRAM ON REGULATION
UNIVERSITY of PENNSYLVANIA

Navigating the Neural Network: Artificial Intelligence in Finance and Recalibration of the Regulatory Framework

Sarah Hammer
The Wharton School
University of Pennsylvania

Penn Program on Regulation Research Paper Series
No. 24-01

October 2024

Navigating the Neural Network: Artificial Intelligence in Finance and Recalibration of the Regulatory Framework

Sarah Hammer†

Abstract

This paper examines the transformative impact of artificial intelligence on the financial sector and the complex regulatory challenges it presents. It explores various applications of AI in finance, from client-facing services like advice to institutional uses such as risk management and regulatory compliance. The paper discusses both the potential benefits of AI in optimizing financial models and achieving efficiencies and the risks it poses, including issues of bias, ethical concerns, market disruption, and institutional safety and soundness.

The paper proposes a taxonomy for AI explainability in financial services, categorizing AI applications based on their potential impact and associated risks. This framework aims to guide regulators in tailoring explainability requirements to different AI applications in the financial sector. The paper also assesses existing regulatory structures and emerging AI-specific initiatives, both in the United States and internationally, to assist in identifying areas where current frameworks may be insufficient to address the unique challenges posed by AI.

To address the fragmented nature of AI regulation in finance, the paper recommends the establishment of a senior AI function within financial sector policy and regulation. This proposed function would focus on regulatory harmonization within the US, collaborate with international regulatory entities and standard-setting bodies, evaluate potential innovation opportunities and enhancements to economic growth, and contribute to assessments of potential systemic risks.

The paper concludes that while explainability provides a foundation for AI-related policymaking in finance, it is not a panacea. More research is needed, and there may be tradeoffs between a model's explainability and its performance. A multifaceted approach is necessary to balance innovation with risk management and ensure AI's responsible development and deployment in the financial sector. The appendix to the paper provides an inventory of existing laws and regulations that require or have the propensity to require model explainability, offering a thorough overview of the current regulatory landscape.

---

† Executive Director, The Wharton School of the University of Pennsylvania.

This paper is forthcoming in the *Oxford University Press Handbook on the Foundations and Regulation of Generative AI.*

Navigating the Neural Network: Artificial Intelligence in Finance and Recalibration of the Regulatory Framework

Sarah Hammer

The AI Revolution in Finance

Artificial intelligence has the potential to transform the financial system radically. From client-facing applications such as the delivery of financial advice or use of chatbots to institutional use cases such as managing lending, complying with capital & liquidity regulations, enterprise-wide adoption, and transforming antiquated computer code, AI is already being leveraged by a vast range of financial institutions to optimize models, achieve efficiencies, and assist with compliance obligations.

The number of applications for AI in finance continues to proliferate, potentially outpacing policymakers' ability to evaluate whether and how existing regulatory frameworks address newfound risks in the use of AI in finance. Since the launch of generative AI in 2022, these applications have grown even more rapidly. To be sure, these applications of AI in finance implicate some of the critical challenges of broader use of AI across many industries — bias, ethical issues, model transparency, and data management.

Unfortunately, the use of generative AI also triggers new and significant concerns about sector-wide harms. Bad actors employ generative AI to enable financial fraud, perpetrate cyber attacks, and inject systemic risk into the financial system. For example, fraudsters use generative AI to create synthetic identities and develop cataclysmic tactics. Criminals use generative AI to analyze and improve existing malware to make even more powerful variants that evade detection and mitigation strategies.[1] Undoubtedly, as generative AI and its capabilities increase, so will dangerous scenarios.

The challenge for US policymakers is to evaluate the current financial regulatory framework and determine whether those rules and regulations apply to generative AI practices by financial institutions. In some cases, such as consumer protection, current regulations may directly apply. In other instances, certain generative AI practices may pose new and novel challenges that current rules and regulations do not address. Policymakers are charged with evaluating whether existing laws and regulations should be amended or implemented.

It is also worth noting that not all regulations related to AI in the financial sector come from financial regulatory agencies. Indeed, many of the challenges related to AI may be better addressed through AI-specific regulations and perhaps even AI-specific regulators. For example, the regulation of foundation models doesn't directly fall within financial regulation. Therefore,

---

[1] Ali, Syed and Ford, Frank, "Generative AI and Cybersecurity: Strengthening Both Defenses and Threats," Bain & Company Technology Report, September 18, 2023, available at https://www.bain.com/insights/generative-ai-and-cybersecurity-strengthening-both-defenses-and-threats-tech-report-2023/ (last visited October 15, 2024).

regulators will also need to evaluate whether certain issues are under their jurisdiction or the jurisdiction of other authorities. They will also need to assess any need to collaborate amongst themselves. This is no easy task given the dynamic of ever-evolving AI technology combined with the complexity and challenges of the US regulatory system.

From Theoretical Concepts to Financial Application

It is important to first define artificial intelligence and distinguish it from generative AI. Artificial intelligence is the use of technology, including machine learning (ML) and generative AI, to analyze data, automate tasks, and improve decision-making. AI is not new; it dates back to 1948, when Alan Turing wrote the first AI manifesto, "Intelligent Machinery." Turing theorized that to be considered intelligence, a machine should be able to imitate human behavior such that it would be indistinguishable from actual humans.

Also in the 1940s, neurophysiologist Warren McCulloch and mathematician Walter Pitts published a paper, "A Logical Calculus of Ideas Imminent in Nervous Activity." The paper focused on how neurons in the brain might work and modeled a simple neural network using electrical circuits. In 1950, Turing published "Computing Machinery and Intelligence," proposing the Turing test as a way to measure a machine's ability. Fast forward decades where other significant inventions surfaced: SNAR, the first neural network computer; IBM's Deep Blue chess champion; Facebook's DeepFace, which recognizes faces; and the release of GPT3 by OpenAI.

As AI has evolved from these early conceptual foundations to practical applications, various approaches to machine learning have emerged. At a high level, machine learning algorithms fall in the categories of supervised, unsupervised, semi-supervised, and reinforcement learning. Supervised ML models are trained on a labeled dataset with input and output parameters. Therefore, they can map points between inputs and outputs. Unsupervised learning refers to ML techniques in which the AI algorithm discovers patterns and relationships using unlabeled data. Unsupervised learning doesn't give the algorithm labeled target outputs. Instead, it discovers patterns or similarities in the data. Reinforcement ML models improve their performance through trial, error, and delay. Each time these models are given more data, they learn and improve. Semi-supervised learning is ML that works between supervised and unsupervised models to use both labelled and unlabeled data.

Generative AI, exemplified by technologies such as large language models (LLMs), differs from traditional machine learning in that it can create novel content and solutions based on training data rather than solely recognizing patterns and making predictions from input data. Generative AI relies on sophisticated models, often using deep learning architectures that identify patterns and representations from vast amounts of data. Generative AI can, therefore, produce new content based on what is learned from the input data.

Diverse Applications of AI in the Financial Sector

At the time of this writing, the number of use cases for AI in finance is already very significant. Most, if not all, types of financial institutions are using AI, including banks, asset managers, hedge funds, private equity firms, venture capitalists, insurance companies, and infrastructure providers. Applications include front-office responsibilities such as portfolio optimization and economic forecasting, institutional management such as risk management and fraud detection, and back-office operations such as trade reconciliation and document processing.

In addition, external contractors working with financial institutions are also increasingly adopting generative AI. The US Securities Exchange Commission (SEC) and Financial Stability Oversight Council (FSOC) noted this. For example, customer service contractors partnering with banks and asset managers are increasingly adopting AI-powered chatbots and virtual assistants to handle customer inquiries and support requests. These AI systems are trained on vast amounts of customer data. They can understand natural language queries, provide accurate responses, and even handle tasks like account balance inquiries or transaction disputes.

Another example is third-party vendors who manage data. These vendors play a vital role in enabling financial institutions to leverage AI effectively and aim to ensure data quality, security, and accessibility. These vendors use AI for data cleansing, anomaly detection, and automated categorization to maintain high-quality, reliable data essential for building accurate and effective AI models. Data management vendors also use generative AI techniques such as natural language processing (NLP) to uncover patterns, trends, and correlations in customer transactions, market trends, and even social media that may not otherwise be apparent. Some risks of using third-party data vendors include privacy and cybersecurity risks.[2]

Concerns About AI in Finance

Many of the concerns about the use of AI in finance relate specifically to the use of generative AI. With this, concerns largely fall into two high-level categories: opaqueness and data challenges. For example, a generative AI system could develop an algorithm for investment advice, lending, or insurance underwriting based on data alone, with no human analysis or oversight and a limited understanding of how the model arrived at a decision. In this case, it is difficult to determine whether the output is appropriate and should be used. Other concerns relate specifically to data, such as limitations based on the scope of the data set, human error in manual labeling of data, and the use of synthetic data in testing, which may impair the ability to detect bias or assess out-of-sample performance.[3]

Developing an Effective Regulatory Approach

Regulating AI in finance requires a broad and well-coordinated framework to address these concerns. To develop an approach, I begin by comparing three use cases for AI in finance.

---

[2] 2023 Annual Report of the Financial Stability Oversight Council, available at https://home.treasury.gov/system/files/261/FSOC2023AnnualReport.pdf (last visited October 15, 2024).

[3] SAS Institute Comment Letter on Federal Reserve Board Request for Information on Financial Institutions' Use of Artificial Intelligence, June 9, 2021, available at https://www.federalreserve.gov/SECRS/2021/August/20210810/OP-1743/OP-1743_070121_138251_324677604799_1.pdf (last visited October 15, 2024).

Example 1: Generative AI in Financial Advisory Services

Generative AI is now employed in the development and provision of financial advice. In the past, financial advice has been generated by a human financial advisor after data gathering, an intake meeting to discuss goals and risk tolerance, and human analysis supplemented with simple financial models. The result was the production of a financial and investing plan for the client and ongoing contact between the human advisor and client in order to execute that plan and track whether the client's goals were being met.

With generative AI, large language models have been found to be able to provide financial advice. Still, this can only be done by adding supplemental domain-specific models incorporating finance-specific information.[4] Generative AI combined with domain-specific models can produce domain-specific knowledge relevant to the consumer's situation. Generative AI can also personalize financial advice in tone, content, and delivery. Also, LLMs can be used to answer ongoing financial questions similar to how a human would respond. These uses may also cultivate and build relationships with clients.

One significant concern about AI-generated financial advice is that it may inherit biases present in the historical data used to train the model. For example, if the training data contains biases related to gender or race, the generative AI may produce advice that discriminates against those same groups. Taking this logic a step further, the bias in the advice could result in those groups receiving less favorable investment recommendations, loan terms, or financial product offerings.[5] When such bias is believed to occur, a deep understanding of the models used is essential to understanding how the bias developed.

Example 2: AI-Driven Trading and Market Dynamics

A second example of the use of AI in finance involves generative AI to power trading models. In the "olden days," trading in equities, fixed income, commodities, and currencies was a human-centric process that relied heavily on the expertise and judgment of individual financial professionals. Traders, together with analysts, would evaluate the value of an asset using quantitative measures such as financial ratios and qualitative factors such as assessment of company leadership.

This process involved reviewing financial statements, reading research reports, and discussing the assets with colleagues. The actual trading was done over the phone or verbally through a "box," where brokers would manually match buy and sell orders. The brokers' deep knowledge of the market and their network of relationships facilitated the price discovery and order-matching process. Both traders and brokers had a role in providing liquidity to the markets.

[4] "Can generative AI provide trusted financial advice?" MIT Ideas to Made to Matter, April 8, 2024, available at https://mitsloan.mit.edu/ideas-made-to-matter/can-generative-ai-provide-trusted-financial-advice (last visited October 15, 2024).
[5] "Can generative AI provide trusted financial advice?" MIT Ideas to Made to Matter, April 8, 2024, available at https://mitsloan.mit.edu/ideas-made-to-matter/can-generative-ai-provide-trusted-financial-advice (last visited October 15, 2024).

Over time, trading and market structure have evolved due to electronic trading, algorithmic trading, and high-frequency trading. Supporters of this progression have argued that it increases efficiency and market liquidity and reduces volatility. Detractors say that the technologization of markets increases volatility, among other negatives.[6]

Of course, ML has been used in trading for decades. Today, the human element of trading may be somewhat reduced because AI is integrated into trading and asset management models. A 2021 survey showed that 48% of hedge funds use AI to develop trading decisions (Statistica, 2021). These AI-powered models are often based on deep learning techniques, analyzing vast amounts of financial data, news sentiment, and market trends to identify complex patterns and generate novel investment strategies. As the use of AI in trading and asset management proliferates, ensuring adequate oversight is important.

Researchers and regulators are exploring how AI technology impacts market power, information rents, price informativeness, market liquidity, and misplacing. They have expressed concerns that the widespread adoption of AI-generated trading models could lead to collusive supra-competitive profits.[7] As the financial sector continues to adopt AI, there may be more consensus on the selection of parameters, such as the type of learning model and the loss function. In the meantime, opaque ("black box") trading models may make it difficult for policymakers to assess the likelihood of these potential scenarios.

Example 3: AI Integration in Banking Operations and Risk Management

Going back decades, banks relied heavily on human personnel to manage all aspects of their operations and ensure regulatory compliance. Bank tellers handled customer transactions, and back-office staff processed paperwork and maintained records. The industry was driven by human intelligence and manual labor. After the Great Financial Crisis of 2008-2009, the Dodd-Frank Act and the Basel III international standards for capital and liquidity management presented banks with a complex new set of rules and regulations designed to increase financial stability. The responsibility for ensuring compliance fell on the shoulders of humans — compliance officers, risk managers, and lawyers. This human-centric approach was labor-intensive, time-consuming, and prone to errors.

Today, some banks use artificial intelligence to manage their capital and liquidity positions in light of increasingly complex bank regulations. A 2020 survey by the Basel Committee on Banking Supervision found that 60% of banks were at that time using AI and machine learning for risk management purposes, including liquidity risk management.[8] AI models can process

---

[6] "High Frequency Trading: Background, Concerns, and Regulatory Developments," the Congressional Research Service, June 19, 2014, available at https://crsreports.congress.gov/product/pdf/R/R43608 (last visited October 15, 2024).
[7] Dou, Winston Wei; Goldstein, Itay; and Ji, Yan, "AI-Powered Trading, Algorithmic Collusion, and Price Efficiency," Jacobs Levy Equity Management Center for Quantitative Financial Research Paper, The Wharton School Research Paper, May 30, 2024, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4452704 (last visited October 15, 2024).
[8] "What is the outlook for the European banking sector?" Conversation between Kerstin af Jochnick, Member of the Supervisory Board of the ECB, and Chris Hallam, Managing Director at Goldman Sachs, at the Twenty-Eighth

vast amounts of financial data, including transaction records, market trends, and client behavior, to predict liquidity needs and optimize capital allocation. By leveraging AI, banks can simulate stress test scenarios, identify potential liquidity gaps, and make data-driven decisions to mitigate risk.

Using AI in bank models to manage the institution poses risks that require careful consideration. Opaqueness in these models can complicate banking supervision and hinder the regulators' ability to assess the soundness of the bank's risk management practices. If the model isn't understood, it will also be a challenge to identify flaws, biases, or unintended consequences that lead to the misallocation of resources such as capital and liquidity. If model flaws are widespread across the banking industry, this could likewise pose widespread risks. While stress testing of models is a requirement for banks, humans are required to oversee and, if necessary, intervene.

Cross-Sectoral Challenges in Finance

While each of the above examples illustrates several risks of using AI at a high level, all three examples demonstrate the challenges of using AI models that are not transparent. Suppose an AI model is producing gender-biased or racially-biased advice. In that case, it is essential to understand how the model used the data and domain-specific inputs to create that advice. Likewise, if generative AI could result in market disruptions, one must critique the trading models and associated processes for implementing those models. Finally, for the supervision of bank safety and soundness, a clear understanding of internal models used for managing capital and liquidity is important for assessing the safety and soundness of a single bank and for assessing whether systemic risk might be present across the industry.

Navigating the Complex Regulatory Framework

Financial institutions already operate in a very complex regulatory landscape, facing oversight from multiple layers of financial and AI-specific regulations. At the federal level in the US, they must comply with regulations set forth by agencies such as the Federal Reserve and the Securities Exchange Commission. Where they exist, state-level financial regulations add another layer of requirements.

Internationally, standard-setting bodies like the Financial Stability Board and the International Organization of Securities Commissions and foreign regulators like the European Central Bank shape global financial practices. Industry associations like FINRA also play a role. As AI becomes increasingly prevalent in finance, institutions must also navigate additional AI-specific regulations and standards, including the US AI Bill of Rights and Executive Order on AI, NIST Standards from the Department of Commerce, international laws such as the EU AI Act, and international standards such as those outlined by the OECD. Jurisdiction-specific rules in countries like China further add to the requirements for multinational institutions.

Annual European Financials Conference, hosted by Goldman Sachs in Madrid, June 5, 2024, available at https://www.bankingsupervision.europa.eu/press/interviews/date/2024/html/ssm.in240605_1~5851fe586f.en.html (last visited October 15, 2024).

Applicable Laws, Regulations, Standards, and Recommendations

- US Federal Financial Regulation
- International Standard-Setting Bodies for Financial Institutions (FSB, IOSCO)
- The Bretton Woods Institutions (IMF, World Bank)
- Foreign Financial Regulators (ECA, UK Financial Conduct Authority, ESMA)
- Finance Industry Associations & Regulators (FINRA)
- AI Specific US Executive Orders (AI Bill of Rights, Executive Order)
- AI Standards in the US (NIST)
- Int'l AI Regulators (UK Safety Institute, EU AI Act)
- Int'l AI Standard Setters (OECD)
- Foreign Jurisdictions Financial and AI Regulations (Singapore, China)
- US State Financial and AI Regulations
- United Nations High-Level Advisory Body on AI Recommendations

Existing Financial Regulatory Structures

In the United States, the financial regulatory framework operates under high complexity and, in many cases, severe fragmentation. Five federal regulatory agencies oversee financial regulation — three banking agencies and two capital markets agencies.[9] A sixth regulatory agency, the Consumer Financial Protection Bureau, is charged with regulating consumer-related financial products and services, protecting against predatory and other harmful practices, and providing education. The US Treasury, the National Economic Council, and the Council of Economic Advisors play lead roles in determining financial sector policy.

At the same time, the United States financial sector operates within the global financial system. It interfaces with international standard-setting bodies such as the Financial Stability Board and international organizations such as the International Monetary Fund. The US regulatory agencies interface with foreign regulators, including the Bank of England, the UK Financial Conduct Authority, the European Central Bank, the European Securities and Markets Authority (ESMA), and the Monetary Authority of Singapore. It is, therefore, essential for the US to engage with these regulators and standard-setting bodies to ensure strong protections against harm and to prevent circumvention.

Arguably, the rapid pace of technological change in AI makes it challenging to mandate definitive regulatory obligations at this early stage. As AI systems evolve and advance at an unprecedented rate, regulators must carefully consider the potential implications of hastily implementing new rules and regulations. Some experts consistently argue that responding too quickly to emerging technologies can lead to unintended consequences and may hinder positive innovation.[10] In some

---

[9] The federal financial regulatory agencies include the Federal Reserve Board of Governors, the Office of the Comptroller of the Currency, the Federal Deposit Insurance Corporation, the Securities Exchange Commission, the Commodity Futures Trading Commission, and the Consumer Financial Protection Bureau.

[10] Professor Jeremias Adams-Prassl, presentation before the International Expert Consortium on the Regulation, Economics, and Computer Science of AI (RECSAI) at the AI for Good Global Summit in Geneva, May 30, 2024.

instances, existing legal frameworks may already provide adequate safeguards to address the risks associated with AI in finance. However, there may also be cases where new principles, standards, or rules are necessary to ensure AI's responsible and ethical deployment in the financial sector.

Emerging AI-Specific Regulatory Initiatives

It is also crucial to acknowledge that other jurisdictions have already taken steps to regulate AI, which has significant implications for US-based companies operating globally. A prime example is the EU AI Act, which aims to establish a comprehensive regulatory framework for AI systems. As US policymakers and regulators navigate the complex landscape of AI governance, they must also consider the extent to which US laws and policies align with or diverge from the rules and regulations of foreign jurisdictions. Inconsistencies or conflicts between legal regimes can create additional challenges for financial institutions seeking to leverage AI across borders. Therefore, international collaboration is essential to ensure a cohesive and effective approach to regulating AI in the financial sector while striking the right balance between fostering innovation and protecting consumers and the financial system.

Towards a Comprehensive Policy Response

Given the still-early stage of regulating AI in finance, it is prudent to establish guardrails to mitigate the risks associated with non-transparent models, such as the type mentioned above. The potential risks of bias, market disruption, and financial institutions' risk may be particularly relevant in areas such as investment advice, algorithmic trading, and bank risk management. Establishing clear guidelines for model transparency and explainability can help financial institutions and regulators navigate the ever-evolving AI landscape and ensure compliance with existing regulations.

There are several options to begin to establish these guardrails. One concept that has been mentioned is AI benchmarking. Benchmarking is a concept that originated from financial services, where data and performance metrics are used to create standards against which financial products or services can be compared. For example, the Morgan Stanley Capital International (MSCI) indexes are benchmarks for various investment portfolios.[11] An equity index fund using the MSCI as its benchmark aims to match the return of the MSCI. However, with regard to AI, the idea of using a benchmark to assess the desired outcome of a model is still in its very early stages. To do this, there would have to be a clear idea of the goal of the benchmark — in other words, policymakers would have to agree on what they are aiming for. Moreover, the rapid development of AI models would make it difficult to identify a firm benchmark. This is not to say that benchmarking AI in finance is not feasible — it would, however, require significant and thoughtful development, which takes time.

A more foundational policy option is model explainability. Explainability in AI refers to understanding and interpreting how a model arrives at its decisions or predictions. By requiring financial institutions to provide clear explanations of their AI models, regulators can better assess the potential risks and ensure that these models are not perpetuating biases, making erroneous decisions, or posing systemic risks. Financial institutions can build trust with regulators, customers, and the public by prioritizing explainability, demonstrating that their AI models are

[11] MSCI, available at https://www.msci.com/ (last visited October 15, 2024).

transparent and accountable. Furthermore, explainable AI can help financial institutions identify and mitigate risks associated with AI models, such as unintended bias or errors, *before* they harm customers or markets. Financial institutions utilizing explainability requirements are taking a step forward in addressing some of the risks associated with AI's opacity.

There are three overall levels of AI explainability. One level is "local model" explainability, which asks specific questions about the model's decisions. Another level is "cohort model" explainability, which uses subsets of data to test accuracy with new and unseen data to identify potential bias in the model. The final level is "global model explainability," which focuses on the features that have the highest impact on model decisions and outcomes.[12]

Regarding specific tactics, common approaches include future importance analysis, which identifies the input variables that have the most significant impact on the model's output; sensitivity analysis, which examines how changes in inputs affect the predictions; and rule extraction, which attempts to distill the model's decision-making process into a set of interpretable rules.[13] These are just a few examples of specific tactics — there are many more.

A Taxonomy for Explainability and Transparency

Not all AI models pose the same level and type of risk, nor are they all used for the same purpose. To address the varying levels of risk and uses associated with different AI applications in finance, it is critical to develop a structured approach to categorizing these applications and their corresponding regulatory responses. For instance, models used for customer segmentation or marketing purposes may not require the same type of explainability as models used for credit scoring, which could have more significant consequences for individuals. This is because bias in credit scoring could result in a disadvantaged population not receiving loans.

On the other hand, one might argue that bias in marketing could have the same negative impact on individuals because it could result in less marketing of helpful products to disadvantaged populations. Another example is investment management. Considerations of whether to require an investment advisor to disclose information about its models should be assessed in light of potential risks and in conjunction with fundamental principles of capital markets operations and the legal protection of proprietary information.

Therefore, whether a model needs a more or less severe type of explainability is debatable. The critical point is that regulators carefully consider the factors and impacts involved before imposing explainability requirements. By developing a clear taxonomy of AI models and their risks, regulators can tailor explainability requirements accordingly. This would ensure that the most critical and impactful models are subject to the highest standards of transparency and accountability without stifling innovation in the industry.

---

[12] Zhang, Ivan, "Explain Yourself. Current Trends in Model Explainability," Harvard Technology Review, March 26, 2023, available at https://harvardtechnologyreview.com/2023/03/26/explain-yourself-current-trends-in-model-explainability/ (last visited October 15, 2024).

[13] Grozdanovski, Ljupcho, "The Explanation One Needs for the Explanation One Gives. The Necessity of Explainable AI (XAI) for Causal Explanations of AI-related harm - Deconstructing the 'Refuge of Ignorance' in the EU's AI liability Regulation," March 27, 2024, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id= 4740419 (last visited October 15, 2024).

In 2021, then-Federal Reserve Board Vice Chair Lael Brainerd pointed out that "there need not be a single principle or one-size-fits-all approach for explaining machine learning models." The level and type of explainability required must take into account who requires the explanation and the purpose of the model. For example, a model developer may need a different type of explanation than a bank compliance officer.[14] Moreover, requirements for explainability should take into account the proprietary nature of AI models and their use by the private sector.

The 2022 White House Blueprint for an AI Bill of Rights also called for a taxonomy for model explainability. In the Section on Notice and Explanation, the order calls for a type of taxonomy, stating that "Automated systems should provide explanations that are technically valid, meaningful and useful to you and to any operators or others who need to understand the system and *calibrated to the level of risk based on the context*."

To address the need for a comprehensive approach to the regulation of artificial intelligence in finance, I propose the following taxonomy as a potential framework for classifying AI use cases based on their potential impact and associated risks. This proposed taxonomy aims to guide regulators in tailoring explainability requirements to different AI applications in the financial sector. The proposed taxonomy categorizes AI applications into three risk levels.

The taxonomy categorizes AI as high-risk where it may result in bias or discrimination, significantly impact financial markets, or is responsible for mission-critical operations at a financial institution. Medium-risk AI would include privacy, data security, and data integrity risks. Low-risk AI includes non-material operations such as process automation. Again, the factors used and the assessment of risk should be carefully debated before setting forth a taxonomy.

Level of Risk	**Characteristics**	**Applications**
High Risk	Propensity for bias and discrimination, significant impact on financial markets, responsibility for fraud and AML detection, possibility of large client losses	AI-generated financial advice, insurance underwriting for individuals, home loan decision-making, HR uses (e.g., hiring)
Medium Risk	Handle sensitive customer information, susceptible to data privacy breaches, potential manipulation of customer behavior	Customer service chatbots, document analysis for internal operations purposes
Low Risk	Relate to process automation and optimization, perform analysis of non-sensitive documents	Marketing materials, administrative tasks (e.g., calendar management, data entry), note taking for meetings

[14] Brainard, Lael, "Supporting Responsible Use of AI and Equitable Outcomes in Financial Services," at the AI Academic Symposium hosted by the Board of Governors of the Federal Reserve System, Washington, D.C. (Virtual Event), January 12, 2021, available at https://www.federalreserve.gov/newsevents/speech/brainard20210112a.htm (last visited October 15, 2024).

Notably, there are alternative ways to establish a taxonomy for explainability. For example, approaches may be developed based on the type of deep learning model and understanding of its modules.[15] This proposed taxonomy is intended as a starting point for discussion and refinement. As the field of AI in finance evolves, so too should our frameworks for understanding and regulating its applications.

Areas of the Law that Require or Provide Propensity to Require Explainability

In the financial sector, several regulatory frameworks and laws already require some level of model explainability to ensure transparency, fairness, and accountability in financial decision-making. Other frameworks offer the propensity for explainability requirements based on core requirements of the particular law, e.g., US consumer protection laws. The appendix details some key areas where model explainability is emphasized in the US and other jurisdictions. The list is by no means exhaustive.

The appendix covers the Basel member organizations, which consist of 45 members from 28 jurisdictions, including central banks and authorities with formal responsibility for bank supervision, as well as observers such as supervisory groups and international organizations. The appendix also includes the EU, the UK, and US jurisdictions. It encompasses 14 different regulatory frameworks that relate to banking, markets and trading, asset management, investment advisors, robo-advisors, broker-dealers, and lenders, as well as data and AI-specific rules. Based on this list, one can begin to imagine the difficulty of developing a comprehensive approach.

Potential Drawbacks to Mandating Explainable AI

While explainability is a valuable tool, it is not without limitations. In some cases, such as low-impact use cases, the need for explainability may be minimal or even unnecessary. This must be weighed against the resource-intensive nature of implementing explainability requirements. As a result, costs could be increased for financial institutions and, by extension, their customers. There's also a risk that over-emphasis on explainability could stifle innovation, as firms might move away from more complex but potentially more effective AI models in favor of simpler, more easily explainable ones. Without proper controls, providing detailed explanations of AI models could also compromise intellectual property.

Increased transparency can also potentially lead to system manipulation. For instance, if an investment client becomes aware of specific characteristics required by an onboarding model for a trading platform, they might alter their profile merely to gain acceptance, potentially compromising the model's integrity. Finally, and very importantly, it's crucial to recognize that enhancing a model's explainability may come at the cost of its performance, creating a delicate balance between transparency and efficacy. There may be scenarios where relying on a complex and tremendously potent model produces the optimal outcome. Thus, while explainability offers significant benefits, its implementation should be carefully considered within the context of each specific application.

---

[15] Saranya A. and Subhashini R., "A systematic review of Explainable Artificial Intelligence models and applications: Recent developments and future trends," Decision Analytics Journal 7 (2023).

Developing a Cohesive Approach Across the Financial Sector

Given the vast and fragmented regulatory landscape outlined here, it is clear that a more coordinated approach to AI governance in the financial sector is needed. To address this challenge and ensure effective oversight of AI in finance, I propose the following recommendation: the establishment of a senior AI function within financial sector policy and regulation to focus on regulatory harmonization within the US. This senior function should also collaborate with international regulatory entities and standard-setting bodies. It should be well-resourced and have a nonvoting seat on the Financial Stability Oversight Council, which would contribute to the FSOC's assessments of potential systemic risk.

As prescribed by statute, the FSOC is exclusively focused on risks to financial stability.[16] Therefore, in addition to domestic regulatory harmonization, this senior function should also include evaluating potential innovation opportunities and enhancements to economic growth. *Partnering harmonization of the regulatory framework with cutting-edge knowledge of AI innovations will enable foreword-looking policy approaches and a better understanding of where regulatory gaps exist*. The senior function could potentially be exercised by the US Treasury's Chief AI Officer, the Director of the White House National Economic Council, or elsewhere, so long as it meets the above criteria and collaborates with the US Treasury and the NEC.

Envisioning the AI Financial Ecosystem

As we navigate the complexity of AI in finance, it is clear that a multifaceted approach is necessary to address the challenges and opportunities presented by the technology. Explainability provides a foundation for AI-related policymaking in the financial sector. Understanding and interpreting how a model arrives at its decisions or predictions will enable regulators to better assess the potential risks and forge appropriate responses. Moreover, financial regulation and standard-setting already offer a vast range of existing laws and regulations requiring or having the propensity to require explainability. Due to the complexity of this framework, regulatory harmonization is imperative.

However, explainability alone is not a panacea. Much more research is needed to build an explainability approach, and there are downsides, such as a tradeoff between the explainability of the model and its performance. Explainability is one key component of overall AI-related regulations. Other AI-specific regulations may relate to hallucinations or cybersecurity. Additionally, as seen above, six relevant federal financial agencies and several overseas regulatory bodies and standard-setters have jurisdiction over AI in the financial sector. Due to the fragmented nature of the applicable regulatory frameworks, there may be areas of regulation where gaps or contradictions exist.[17]

[16] Dodd-Frank Wall Street Reform and Consumer Protection Act, Pub. L. No. 111 - 203, § 929-Z, 124 Stat. 1376, 1871 (2010) (codified at 15 U.S.C. § 78o).
[17] Hammer, Sarah, "The Regulation of Artificial Intelligence in Finance," (unpublished working paper), December 4, 2023, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4621225 (last visited October 15, 2024).

To be sure, there may be areas where regulation should be enhanced to protect the public and the financial system. For example, AI and data governance are key areas for attention. Ensuring the security, architecture, and integrity of data used to train and operate AI models is essential for building robust and trustworthy AI systems. Likewise, developing a clear system for institutional AI governance is critical. A detailed discussion of data management and governance is crucial but beyond the scope of this paper.

In the final analysis, implementing a taxonomy for AI explainability in finance is a strategic imperative. This framework could enhance compliance, foster trust, mitigate risks, and drive innovation by effectively matching explanatory techniques to specific AI applications and stakeholder needs. A senior AI function within US financial sector policy and regulation could enable this process, assess innovation opportunities, promote forward-looking policy approaches, and ensure coherent implementation. As AI continues to reshape finance, such coordinated efforts are crucial in balancing advanced algorithms with the principles that underpin our financial system.

APPENDIX

EXISTING FRAMEWORKS REQUIRING OR HAVING PROPENSITY TO REQUIRE EXPLAINABILITY IN THE FINANCIAL SECTOR

Note: Where foreign frameworks or regulations are listed, they include provisions establishing jurisdiction over US companies domiciled in that jurisdiction or operating within that jurisdiction or doing business with individuals or companies within that jurisdiction.

Global —

Basel III - International:

- **Pillar 3 Disclosure Requirements - Updated Framework:** Requires banks to disclose their risk assessment processes and the *methodologies used in their internal models*. Moreover, banks are expected to *explain the main drivers of differences between internally modeled amounts disclosed that are used to calculate their capital ratios and amounts disclosed should the banks apply the standardized approach.*[18]

European Union —

General Data Protection Regulation (GDPR):

- **Article 22**: GDPR includes provisions for *the right not to be subject to a decision based solely on automated decision-making*, including profiling, which significantly affects individuals and produces legal effects concerning him or her.[19]
- **Recital 71**: States that decision-making based on automated processing of personal data evaluating personal aspects relating to a natural person should be allowed where expressly authorized by Union or Member State law, such as for fraud and tax-evasion monitoring and prevention. *Further states that such processing should be subject to suitable safeguards, including* specific information to the data subject, the right to obtain human intervention, the right to express his or her point of view, and *the right to obtain an explanation of the decision reached after such assessment and to challenge the decision.*[20]

[18] Basel Committee on Banking Supervision, Standards, "Pillar 3 disclosure requirements – updated framework," December 2018, available at https://www.bis.org/bcbs/publ/d455.pdf (last visited October 15, 2024).
[19] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance), OJ L 119, 4.5.2016, p. 1–88, available at https://gdpr.eu/article-22-automated-individual-decision-making/ (last visited October 15, 2024).
[20] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance), available at https://gdpr-info.eu/recitals/no-71/ (last visited October 15, 2024).

Markets in Financial Instruments Directive II (MiFID II):

- **Section 17(1):** Requires financial firms to have in place effective systems and risk controls suitable to the business it operates to ensure that its trading systems are resilient and have sufficient capacity, are subject to appropriate trading thresholds and limits and prevent the sending of erroneous orders that may create or contribute to a disorderly market.
- **Section 17(2):** The competent authority of the home Member State of the investment firm *may require the investment firm to provide, on a regular or ad-hoc basis, a description of the nature of its algorithmic trading strategies, details of the trading parameters or limits to which the system is subject, [and] the key compliance and risk controls it has in place.*[21]

EU AI Act:

- **Annex XI with regard to Article 11 Technical Documentation:** The EU AI Act has entered into force. Annex 11 with regard to Article 11 requires explainability and transparency in AI applications. *The Act requires technical documentation of an AI system including, but not limited to, general and detailed descriptions of the AI system, detailed information about monitoring, functioning, and control of the AI system, a detailed description of the risk management system, and a description of any change made to the system over its lifecycle*.[22]
- **European Commission - Jurisdiction:** The EU AI Act *applies to both public and private sector actors inside and outside the EU*, as long as the AI system is placed on the European Market or has an impact on people located in the EU.[23]

United Kingdom —

Financial Conduct Authority (FCA):

- **AI Update Section 3.6:** The UK Government's AI principles have identified five principles as key when it comes to the regulation of AI in the UK, including "*(2) appropriate transparency and explainability*."[24]

---

[21] MiFID II Directive 2014/65/EU of the European Parliament and of the Council of 15 May 2014 on markets in financial instruments and amending Directive 2002/92/EC and Directive 2011/61/EU, available at https://www.esma.europa.eu/publications-and-data/interactive-single-rulebook/mifid-ii/article-17-algorithmic-trading (last visited October 15, 2024).

[22] ANNEXES to the Proposal for a Regulation of the European Parliament and of the Council LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS {SEC(2021) 167 final} - {SWD(2021) 84 final} - {SWD(2021) 85 final}, April 21, 2021, available at https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52021PC0206 (last visited October 15, 2024).

[23] European Commission, Artificial Intelligence Questions and Answers, August 1, 2024, available at https://ec.europa.eu/commission/presscorner/api/files/document/print/en/qanda_21_1683/QANDA_21_1683_EN.pdf (last visited October 15, 2024). See, primarily, Art 2(1).

[24] "AI Update," Financial Conduct Authority, available at https://www.fca.org.uk/publication/corporate/ai-update.pdf (last visited October 15, 2024).

- **AI Update Section 3.35:** Although the UK's regulatory framework does not specifically address the transparency or explainability of AI systems, there are a number of high-level requirements and principles in its approach to consumer protection.[25]
- **AI Update Section 3.36:** Principle 7 of the AI Principles requires firms where Consumer Duty does not apply to "*pay due regard to the information needs of clients* and communicate with them in a way that is clear, fair and not misleading."**[26]**
- **AI Update Section 3.37:** With regard to processing personal data, *"data controllers must provide data subjects with certain information about their processing activities, including the existence of automated decision-making and profiling."***[27]**

United States —

Dodd-Frank Wall Street Reform and Consumer Protection Act:

- **Section 1033**: Requires a consumer financial services provider to *make available to a consumer information* in its control or possession concerning the financial product or service.[28]
- **Consumer Financial Protection Bureau Circular 2022-03:** The CFPB confirmed that federal anti-discrimination law requires companies to e*xplain to applicants the specific reasons for denying an application for credit or taking other adverse actions, even if the creditor is relying on credit models using complex algorithms.*[29]

Fair Credit Reporting Act (FCRA):

- **Section 615 (Adverse Actions):** Requires that if a consumer is denied credit, employment, insurance, or any other service based on information in a credit report, *the entity must provide the consumer with an adverse action notice*.[30]
- **Section 609 (Disclosures to Consumers):** Under this section, consumers have the right to request and obtain all information in their credit file, including the sources of information and the entities that have accessed their credit report. *While not explicitly stipulated in the*

[25] "AI Update," Financial Conduct Authority, available at https://www.fca.org.uk/publication/corporate/ai-update.pdf (last visited October 15, 2024).
[26] "AI Update," Financial Conduct Authority, available at https://www.fca.org.uk/publication/corporate/ai-update.pdf (last visited October 15, 2024).
[27] "AI Update," Financial Conduct Authority, available at https://www.fca.org.uk/publication/corporate/ai-update.pdf (last visited October 15, 2024).
[28] 12 CFR Chapter X [Docket No. CFPB–2020–0034] RIN 3170–AA78 Consumer Access to Financial Records, available at https://www.federalregister.gov/documents/2020/11/06/2020-23723/consumer-access-to-financial-records (last visited October 15, 2024).
[29] Consumer Financial Protection Circular 2022-03: "Adverse action notification requirements in connection with credit decisions based on complex algorithms," Consumer Financial Protection Circular, available at https://www.consumerfinance.gov/compliance/circulars/circular-2022-03-adverse-action-notification-requirements-in-connection-with-credit-decisions-based-on-complex-algorithms/ (last visited October 15, 2024).
[30] 15 U.S.C. 1681m, available at https://www.ffiec.gov/exam/InfoBase/documents/02-con-fair_credit_reporting_act-000799.pdf (last visited October 15, 2024).

section, "all information" and "sources of information" arguably includes an explanation of a decision-making model used, if there was one.[31]

Equal Credit Opportunity Act (ECOA):

- **Title VII of the Consumer Credit Protection Act (15 USD 1691-1691f):** *Requires creditors to "provide applicants, upon request, with the reasons underlying decisions to deny credit."*[32]

Federal Reserve's Guidance on Model Risk Management (SR 11-7):

- **SR 11-7:** Guidance on Model Risk Management is a document issued by the Federal Reserve Board and the Comptroller of the Currency in 2011 that provides guidance and regulatory expectations related to model risk management. The guidance defines a model as a set of software tools and techniques used to generate outputs based on certain inputs. This covers a wide range of models including those created in Microsoft Excel as well as AI/ML models.[33]
    - The guidance defines model risk as the risk of incorrect or inappropriate model usage, incorrect model outputs, or model implementation errors.
    - It outlines how financial institutions should design, implement, and maintain a comprehensive model risk management framework. The framework should include a set of model risk management policies and procedures that cover all aspects of model development and usage. It should also include an independent model risk oversight process, model validation and review process, model risk reporting and monitoring processes. The guidance also provides expectations on model governance, data quality, model development processes, model validation, model risk management policies and procedures, model risk oversight, and model risk reporting.
    - The guidance emphasizes that the model risk management framework should be consistent with an institution's risk management culture and be tailored to the institution's business strategy, products, services, and risk profile. It also specifically notes that *"where models and model output have a material impact on business decisions, including decisions related to risk management and capital and liquidity planning, and where model failure would have a particularly harmful impact on a bank's financial condition, a bank's model risk management framework should be more extensive and rigorous."*
- **2021 Interagency Statement on Model Risk Management for Bank Systems Supporting Bank Secrecy Act/Anti-Money Laundering Compliance (BSA/AML):** In this 2021 statement, the three federal banking regulators (the Federal Reserve Board, the Office of

[31] 15 U.S.C. 1681, available at https://www.ffiec.gov/exam/InfoBase/documents/02-con-fair_credit_reporting_act-000799.pdf (last visited October 15, 2024).
[32] 15 U.S.C. §§ 1691-1691f, available at https://www.ftc.gov/legal-library/browse/statutes/equal-credit-opportunity-act (last visited October 15, 2024).
[33] SR 11-7: Guidance on Model Risk Management, April 4, 2011, available at https://www.federalreserve.gov/supervisionreg/srletters/sr1107.htm (last visited October 15, 2024).

the Comptroller of the Currency, and the Federal Deposit Insurance Corporation) note that the 2011 supervisory guidance does not have the force of law. They note that the Model Risk Management Guidance is principles-based and provides flexibility for banks in developing, implementing, and updating models. The Statement also discusses the use of third-party models and stipulates that, ultimately, the bank is responsible for BSA/AML compliance.

Securities Exchange Commission (SEC):

- **Regulation Systems Compliance and Integrity (Reg SCI):** Reg SCI is not explicitly about AI. However, it requires certain market participants to have policies and procedures to ensure the capacity, integrity, resiliency, and security of their technological systems. It does not specifically set forth explainability requirements.[34]
- **Investment Advisers Act of 1940, Section 206:** This is the anti-fraud provision of the Act. It requires investment advisers to act as fiduciaries to their clients. The SEC has interpreted this to mean that advisers using AI or algorithmic methods must be able to *explain their investment strategies and methodologies to clients*.
- **SEC Guidance on the Use of Robo-Advisers, No. 2017-02:** The SEC's 2017 guidance for the use of robo-advisors includes a recommendation, "Explanation of Business Model." That provision recommends that a robo-adviser *should disclose a description of the algorithmic functions used to manage client accounts, a description of the assumptions and limitations of the algorithm used to manage client accounts, and a description of the particular risks inherent in the use of an algorithm to manage client accounts*.[35]
- **SEC Rule on Risk Management Control for Brokers or Dealers with Market Access, Rule 15c3-5:** This rule requires broker-dealers with market access to "*establish, document, and maintain*" a system for regularly reviewing the effectiveness of the risk management controls and supervisory procedures and for promptly addressing any issues; and no less frequently than annually, conduct a review of its business activity in connection with market access to assure the overall effectiveness of such risk management controls and supervisory procedures and document that review.[36]
- **SEC Proposed Rule on Conflicts of Interest Associated with the Use of Predictive Data Analytics by Broker-Dealers and Investment Advisers, 17 CFR Parts 240 and 275:**[37] In 2023, the SEC *proposed* new requirements to address risks to investors from conflicts

[34] 17 CFR Parts 240, 242, and 249 [Release No. 34-73639; File No. S7-01-13] RIN 3235-AL43 Regulation Systems Compliance and Integrity, February 3, 2015, available at https://www.sec.gov/files/rules/final/2014/34-73639.pdf (last visited October 15, 2024).
[35] Investment Management Guidance Update, No. 2017-02, U.S. Securities Exchange Commission, February 2017, available at https://www.sec.gov/investment/im-guidance-2017-02.pdf (last visited October 15, 2024).
[36] 17 CFR PART 240 [Release No. 34-63241; File No. S7-03-10] RIN 3235-AK53 Risk Management Controls for Brokers or Dealers with Market Access, November 3, 2010, available at https://www.sec.gov/files/rules/final/2010/34-63241.pdf (last visited October 15, 2024).
[37] 17 CFR Parts 240 and 275 [Release Nos. 34-97990; IA-6353; File No. S7-12-23] RIN 3235-AN00; 3235-AN14 Conflicts of Interest Associated with the Use of Predictive Data Analytics by BrokerDealers and Investment Advisers, October 10, 2023, available at https://www.sec.gov/files/rules/proposed/2023/34-97990.pdf (last visited October 15, 2024).

of interest associated with the use of predictive data analytics. The proposed rule would operate on the premise that "advisers or brokers are optimizing to place their interests ahead of their investors' interests" by using predictive data analytics models. The proposed rule generally would require a firm to evaluate and determine whether its use of certain technologies in investor interactions involves a conflict of interest in that the firm's interest is placed ahead of the investors' interests. The proposed rule would *require firms to have written policies and procedures reasonably designed to achieve compliance with the proposed rule*. Importantly, the rule is not yet in force.

Financial Industry Regulatory Authority (FINRA):

- **FINRA Rule 3110 (Supervision):** While it does not explicitly require model explainability, FINRA has reminded its members that the rule applies to regulatory obligations of firms using generative AI and large language models.[38] The rule states that "[A] member firm must have a reasonably designed supervisory system tailored to its business."[39]
- **FINRA Rule 2210 (Communications with the Public):** This rule pertains to member firms' communications with the public and is not specific to artificial intelligence. The rule states that "(E) Members must consider the nature of the audience to which the communication will be directed and *must provide details and explanations appropriate to the audience*."[40]

Commodity Futures Trading Commission (CFTC):

- **Request for Comment on the Use of Artificial Intelligence in CFTC-Regulated Markets, January 2024:** In this Request for Comment (RFC), the CFTC sought comment on the definition of AI and its applications including its use in trading, risk management, compliance, cybersecurity, record keeping, data processing and analytics, and customer interactions. The request also seeks comment on the risks of AI, including risks related to market manipulation and fraud, governance, explainability, data quality, concentration, bias, privacy and confidentiality, and customer protection.
- **Technology Advisory Committee Report, May 2024:** In May 2024, the Technology Advisory Committee of the CFTC released a report and recommendations on responsible AI in the financial markets.[41]

---

[38] FINRA Regulatory Notice 24-09: FINRA Reminds Members of Regulatory Obligations When Using Generative Artificial Intelligence and Large Language Models, June 27, 2024, available at https://www.finra.org/rules-guidance/notices/24-09 (last visited October 15, 2024).
[39] FINRA, Rule 3110 (2024), available at https://www.finra.org/rules-guidance/rulebooks/finra-rules/3110 (last visited October 15, 2024).
[40] FINRA, Rule 2210 (August 2019), available at https://www.finra.org/rules-guidance/rulebooks/finra-rules/2210 (last visited October 15, 2024).
[41] "CFTC Technology Advisory Committee Advances Report and Recommendations to the CFTC on Responsible Artificial Intelligence in Financial Markets," Commodity Futures Trading Commission Release Number 8905-24, May 2, 2024, available at https://www.cftc.gov/PressRoom/PressReleases/8905-24 (last visited October 15, 2024).

Executive Branch Actions Specifically Applicable to AI:

- **Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, Section 8, October 30, 2023. Protecting Consumers, Patients, Passengers, and Students.** Independent regulatory agencies are encouraged, as they deem appropriate, to consider using their full range of authorities to protect American consumers from fraud, discrimination, and threats to privacy and to address other risks that may arise from the use of AI, including risks to financial stability, and to consider rulemaking, as well as emphasizing or clarifying where existing regulations and guidance apply to AI, including clarifying the responsibility of regulated entities to conduct due diligence on and monitor any third-party AI services they use, and *emphasizing or clarifying requirements and expectations related to the transparency of AI models and regulated entities 'ability to explain the use of AI models.*[42]
- **Blueprint for an AI Bill of Rights, White House Office of Science and Technology Policy, October 4, 2022:** The AI Bill of Rights was released by the White House with the goal of guiding the design, development, and deployment of artificial intelligence and other automated systems so that they protect the rights of the American people.[43] The "Notice and Explanation" section of the Blueprint states that "Designers, developers, and deployers of automated systems *should provide generally accessible plain language documentation including clear descriptions…and explanations of outcomes that are clear, timely, and accessible*." It further calls for "Automated systems…[to] *provide explanations that are technically valid, meaningful and useful to you and any operators or others who need to understand the system*…". The Blueprint then goes on to state that "Reporting that includes summary information about these automated systems in plain language and assessments of the clarity and quality of the notice and *explanations should be made public wherever possible*."

[42] "Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence," The White House, October 30, 2023, available at 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/ (last visited October 15, 2024).
[43] "Blueprint for an AI Bill of Rights: A Vision for Protecting Our Civil Rights in the Algorithmic Age," The White House Office of Science and Technology Policy, October 4, 2022, available at https://www.whitehouse.gov/ostp/news-updates/2022/10/04/blueprint-for-an-ai-bill-of-rightsa-vision-for-protecting-our-civil-rights-in-the-algorithmic-age/ (last visited October 15, 2024).