Business Roundtable

# Defining and Addressing Artificial Intelligence Risks

September 2024

# Overview

As artificial intelligence (AI) becomes more widely used, policymakers will need to take a pragmatic approach to understanding the risks and, as necessary, designing effective regulatory guardrails. Those guardrails should focus on evidence-based, real-world threats rather than conjectural harms. Existing regulations and industry risk management practices can be adapted to safeguard against many of the concerns regarding the societal and economic impacts of AI, though some gaps remain.

Policymakers should collaborate with industry stakeholders to develop AI policies that address remaining risks without stifling innovation. When designed properly, policies that promote effective risk management practices can mitigate harms, thereby increasing public trust in AI and advancing adoption of the technology.

In designing effective guardrails to address AI risk, policymakers should:

- **Define high-risk uses of AI by considering the specific contexts or use cases and the potential outcomes;**
- **Address AI risk through existing laws and regulations, where applicable, and apply existing risk management standards to address potential policy gaps; and**
- **Identify clear and measurable strategies for evaluating AI risk throughout design, development and deployment.**

# Policy Considerations and Recommendations

As industry continues to offer new AI models, systems and tools, policymakers should prioritize the following:

## Defining and characterizing AI risk

**Business Roundtable Recommendation:**
Policymakers should focus on high risk and potential outcomes associated with deploying AI models and systems in specific contexts, while avoiding broad classifications of risk for entire sectors, categories of AI or uses of AI. Additionally, policymakers should define "high-risk" through a collaborative, robust stakeholder process.

Policymakers should focus their efforts on risks that may arise during the development and deployment of AI models and tools, rather than seeking to define the technology itself as inherently risky. Understanding the types and degrees of risk posed by AI can be complicated. AI risks are shaped by many factors, including how a system is used; the environment in which it is deployed; the type of data processed or used to create models and tools; interactions with other AI systems; and the user characteristics, such as level of experience with or training on AI. Ultimately, the risks associated with a particular AI model or tool will be higher or lower depending on the specific use case.

Ultimately, the risks associated with a particular AI model or tool will be higher or lower **depending on the specific use case**.

The most important considerations for understanding the nature and degree of AI risk — and for distinguishing truly high-risk use cases for regulatory purposes — are the severity, scale and likelihood of potential harm to consumers or society, and whether the harmful impact could be effectively remediated or reversed. Regulatory guardrails for high-risk use cases may be appropriate depending on the purpose for using AI, whether the AI is the principal basis for determining consequential decisions and the extent to which AI systems are already constrained by substantial guardrails or other effective controls.

New regulatory and legislative proposals should be risk-based, context-driven, flexible and crafted based on stakeholder engagement and feedback, allowing for adaptability to evolving technologies, use cases and degrees of risk. Regulatory guardrails should focus on reducing risk and mitigating harm from high-risk use cases in a nuanced manner, without adding unnecessary regulatory burdens on beneficial, less risky use cases that could inhibit innovation that benefits the economy and consumers.

Policymakers must avoid designating broad categories of AI as high-risk and, instead, focus on specific characteristics that would make a use case truly high-risk. For example, AI use in a critical infrastructure sector should not be defined as a high-risk without regard for the actual impact to an asset.

To create effective regulatory guardrails to manage the risks associated with AI, policymakers should work in partnership with industry and apply an iterative approach to regulation to address emerging risks and contexts as they arise. Businesses need certainty to determine what is and is not considered high-risk as they continue to develop and deploy AI technology. To that end, policymakers should define high-risk uses through a collaborative, robust stakeholder process while building on existing guidance and oversight, such as the AI Risk Management Framework (RMF) from the National Institute of Standards and Technology (NIST).

## Addressing AI risk using existing approaches

**Business Roundtable Recommendation:**
Policymakers should align legislative and regulatory proposals with existing, effective domestic and international policies and industry risk management strategies to promote a harmonized approach and avoid introducing uncertainty and conflicting compliance requirements.

Existing laws, regulations and risk management frameworks, together with voluntary international standards, already address many of the potential risks associated with AI systems and uses through technology-neutral approaches. For example:

- The Fair Credit Reporting Act, the Fair and Accurate Credit Transactions Act, the Equal Credit Opportunity Act, the Dodd-Frank Wall Street Reform and the Consumer Protection Act constrain the use of AI to make decisions affecting consumers, such as whether to grant a loan.
- The Civil Rights Act and the Fair Housing Act address discriminatory housing practices and decisions, regardless of whether AI is incorporated into the decision-making process.
- The Food and Drug Administration has established a robust and longstanding regulatory regime to regulate AI.
- The U.S. Equal Employment Opportunity Commission (EEOC)'s guidance around employment and hiring policies and practices underscore that existing law prohibits AI-enabled discrimination throughout the hiring and employment lifecycle.

In addition, regulators including the EEOC, Consumer Financial Protection Bureau, Department of Justice and Federal Trade Commission have made clear in public statements that their requirements apply regardless of the technology used to make the decision.

Beyond existing regulations, the NIST AI RMF establishes voluntary, flexible and practical AI risk management considerations and practices. To facilitate global interoperability and trade, a growing body of international AI standards work provides common approaches for terminology, concepts, governance and trustworthiness.

To ensure that risk management for AI technologies is effective, any new laws and regulations to address gaps must avoid contradicting, duplicating or fragmenting established rules and requirements. To the extent that existing regulations are not preempted, policymakers across regulating agencies may need to update guidance to ensure that existing requirements align and do not impose additional undue compliance burdens.

Any new laws and regulations to address gaps **must avoid contradicting, duplicating or fragmenting** established rules and requirements.

By aligning common principles, definitions and standards, policymakers can promote a harmonized approach that accommodates diverse applications of AI and preserves the ability of U.S. companies to innovate in the global economy while safeguarding against potential risks and uncertainty.

## Testing and evaluating AI risk

**Business Roundtable Recommendation:**
Policymakers should identify clear, measurable strategies for evaluating and addressing AI risks that will equip developers and deployers with the necessary information to safely, securely and confidently use AI.

Continuous testing and evaluation of AI risk can enable effective guardrails to protect individuals, organizations and ecosystems from potential harms arising from high-risk AI systems. Existing risk management approaches, such as privacy and security assessments, can be adapted to test and evaluate a wide variety of risks associated with AI systems. Multi-stakeholder technical forums — such as NIST and its U.S. AI Safety Institute and standards development organizations — should continue leading these efforts to design strategies and adapt existing approaches and international standards, as appropriate. Policymakers should work closely with industry stakeholders to identify the most appropriate risk management strategies across different contexts and use cases. Evaluation should be conducted by experts such as developers, deployers and other stakeholders, utilizing these shared approaches with modifications based on context, as needed.

Several tools that are already widely used by industry are effective for addressing AI risk. The NIST AI RMF provides a flexible, stakeholder-informed way to measure risk based on the context in which AI is used. Its "Measure" function outlines the kinds of quantitative and qualitative tools that might be used to understand AI risk and impact. NIST should continue to evolve its guidance to further develop quantitative risk measures to provide more robust, domain-specific guidance. Algorithmic impact assessments are another approach for evaluating the underlying components of an AI system, focusing on critical aspects such as bias, explainability, interpretability, accuracy, validity and fairness.

Another risk management approach is "red-teaming." The Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence defines red-teaming as a structured testing effort to find flaws and vulnerabilities in AI systems. These tests often leverage adversarial methods to identify harmful or discriminatory outputs from an AI system, unforeseen or undesirable system behaviors, limitations or potential risks associated with the misuse of the system.

Business Roundtable supports the use of these tools in appropriate contexts to address risk in AI systems. Many of our members already use algorithmic assessments and internal red-teaming to identify and address potential harmful risk from AI systems.

Policymakers should also consider how AI risk is introduced and varies substantially at different points. Specific use case, data and user context often results in different risk profiles and mitigation opportunities for the same AI system or tool. Developers of AI models and tools face and manage risk differently than intermediate deployers or enterprise users of said tools. Intermediaries will not have access to the underlying data used to train a model that was built by a third-party developer and cannot provide that training data to downstream customers or to regulators. They should, however, bear responsibility for providing any risk assessments that they conduct, or data sets that they add, to an underlying model to the extent similar disclosures are required of upstream developers and downstream deployers.

Specific use case, data and user context often results in **different risk profiles and mitigation opportunities** for the same AI system or tool.

## Conclusion

As AI development and deployment continues, it is crucial that government and industry collaborate on managing risk. AI has the potential to be transformative for the American economy. An effective regulatory framework will be critical to building public trust, guarding against risks and fostering innovation.