# The Impact of the ‘U.S. Executive Order on AI’

Published 15 December 2023 - ID G00805298 - 29 min read

By Analyst(s): Lydia Clougherty Jones, Frances Karamouzis, Svetlana Sicular, Avivah Litan

Initiatives: Executive Leadership: Artificial Intelligence; Artificial Intelligence; Build Trust and Mature D&A Culture; Procurement and Strategic Sourcing Applications

The U.S. has issued an executive order that it calls “the most significant actions ever taken by any government to advance the field of AI safety.” Executive leaders should adjust leadership priorities, reconcile AI investment with redistributed risk, and prepare today for a more regulated tomorrow.

## Overview

### Impacts

- The U.S. Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence (the “EO”) sends a strong message to the private and public sector: AI is not a new technology, it’s a leadership responsibility.
- The EO, in effect, redistributes the risks of loss associated with responsible AI failures, which will affect current and future AI investments and the projected corresponding ROI.
- The EO has major impacts on U.S. federal agencies, and those who work with them. It will undoubtedly have a significant impact on the global AI vendor and solutions market, including to address the new standards for the development, functionality, use and output of AI.

### Recommendations

- Accelerate scenario and use planning by engaging in methodical approaches. Create an “AI biometric dashboard“ to derive an executive summary to monitor the vital signs of organizational alignment of AI to various scenarios and use cases proposed, piloted and funded by the organization.

- Adapt your AI strategy and AI investment priorities to address new or elevated risks of harm and loss from responsible AI failures. Start by identifying and quantifying the potential AI risks of harm most relevant to stakeholder priorities. Then adopt and promote “effort as a stand-alone enterprisewide best practice” to avoid responsible AI failures.

- Prepare for EO-driven regulatory compliance — and a safer AI future in general — by working with stakeholders (such as legal, finance, government relations and assurance) across the organization to implement a comprehensive AI trust, risk and security management program. Use Gartner’s AI trust, risk and security management (TRiSM) framework to identify technology solutions that support this endeavor.

## Strategic Planning Assumptions

By 2027, nearly 15% of new applications will be automatically generated by AI without a human in the loop, up from zero percent today.

By 2026, over 100 million people will engage robo colleagues (synthetic virtual colleagues) to contribute to enterprise work.

By 2025, insider risk will cause 50% of organizations to adopt formal programs to manage it, up from 10% today.

By 2028, more than 50% of enterprises that have built their own large language models (LLMs) from scratch will abandon their efforts due to costs, complexity and technical debt.

By 2025, the concentration of pretrained AI models among 1% of AI vendors will make responsible AI a societal concern.

By 2026, enterprises that apply AI TRiSM controls to their applications will consume up to 80% less inaccurate or illegitimate information.

## Introduction

*Prepare now for the far and wide impact of the EO on executive leadership*

U.S. president Joe Biden's EO underscores AI's promise of innovation and competitive advantage. [1] It specifically calls out for Americans' privacy and civil liberties' protection, equity and civil rights advancement, and consumers' and workers' support, thus reinforcing the U.S. "Blueprint for an AI Bill of Rights." [2]

The EO, and The Office of Management and Budget (OMB) Implementation Guidance, [3] establishes new and elevated standards for AI safety and security applicable to private and public sector organizations via directives to over 20 U.S. federal agencies. Deadlines for implementation span between 30 and 365 days from the 30 October 2023 date of the EO. The U.S. agencies charged with issuing guidance are acting quickly. Executive leaders should expect this accelerated pace to continue.

Through the EO, the U.S. is sending a clear signal that the impact of AI and GenAI is far more than a disruptive technology; it has far reaching consequences for every aspect of daily life, national and global economies, military matters and the future of the planet. As such, unlike many other areas of technology or disruptive forces where government organizations have a low reaction time, this is different and executive leadership will be tested accordingly.

*Executive leadership across the public and private sector will be tested within the EO prioritization of AI innovation alongside new mandatory guardrails against AI harms.*

The EO's applicability can reach U.S. and non U.S. entities across the public, private, academic, nongovernmental organization and consumer/citizen ecosystem. As such, the EO will have a broad impact on AI strategy and the ROI on AI investment. It will alter and redistribute the risk of loss from AI harms considered too costly for AI benefit or value. It will also change corporate, individual and machine behaviors arising from new regulatory frameworks, alongside industry self regulation we are already beginning to see take effect. Together, this creates opportunity for new market solutions, but the oversight of vendors by the government and the commercial sector will be exacting.

One example that requires the immediate attention of executive leaders in the private sector is the invocation of the U.S. Defense Production Act to require disclosures regarding acquisition, development or possession of a "potential large-scale computing cluster" to the U.S. government as follows:

> **Companies, individuals, or other organizations or entities that acquire, develop, or possess a potential large-scale computing cluster to report (to the Federal Government) any such acquisition, development, or possession, including the existence and location of these clusters and the amount of total computing power available in each cluster.**
>
> *— Source: Executive Order on the Safe, Secure and Trustworthy Development and Use of Artificial Intelligence, Section 4.2. Ensuring Safe and Reliable AI. (a) (ii) The White House*

Another disclosure requirement, that includes premarket safety testing requirements, applicable to companies, is as follows:

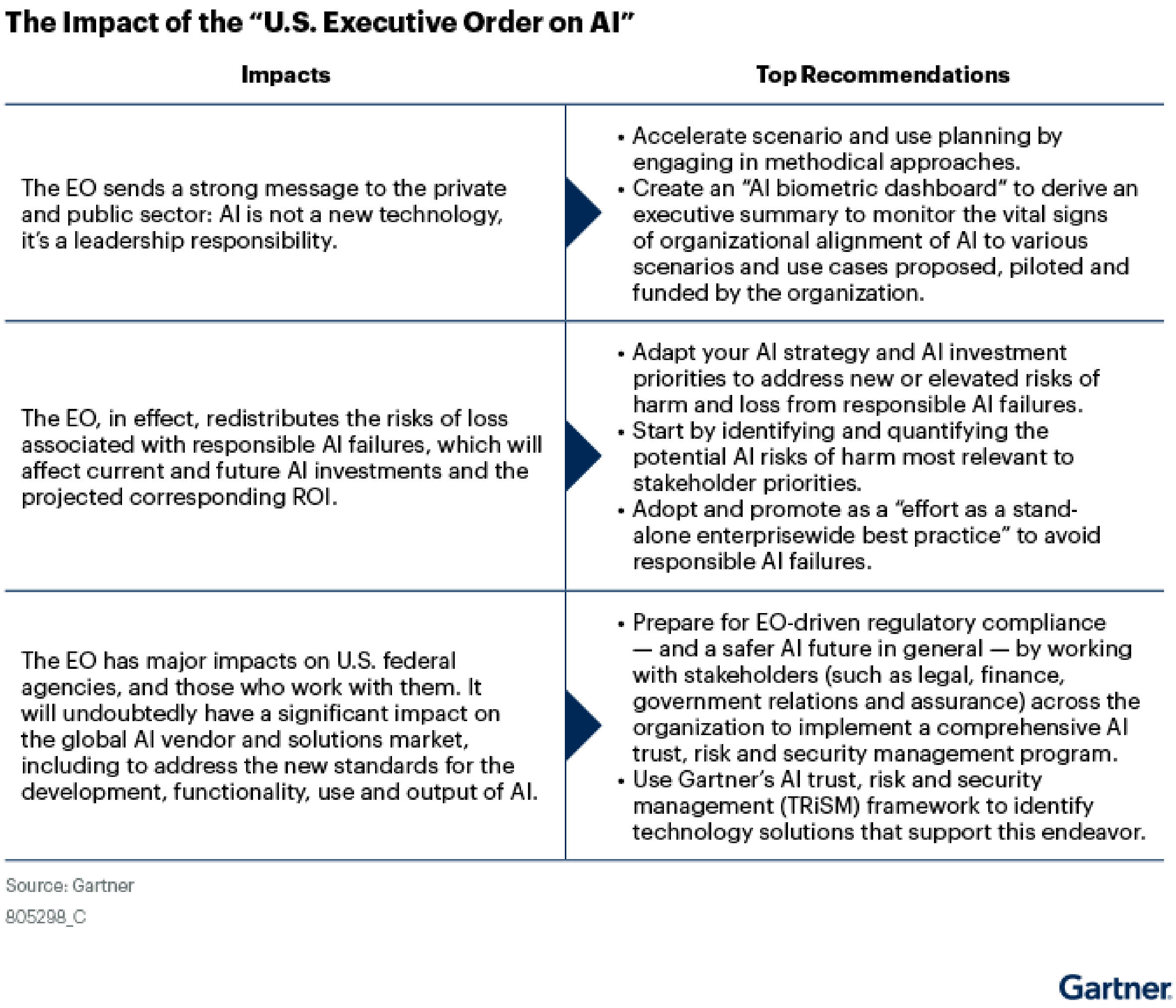
> **Companies developing or demonstrating an intent to develop potential dual-use foundation models provide training information, model weights reports and safety training results to the Federal Government, on an ongoing basis.**
>
> *— Source: Executive Order on the Safe, Secure and Trustworthy Development and Use of Artificial Intelligence, Section 4.2. Ensuring Safe and Reliable AI. (a) (i) The White House*

For the associated texts regarding these disclosure and safety testing requirements under the EO, see Note 1.

Executive leaders in the U.S. government also have a lot of work ahead, including navigating their leadership alongside the new mandate that every federal agency, excluding intelligence agencies, within 60 days from 1 November 2023, appoint a chief artificial intelligence officer to lead the implementation efforts of the EO. For a summary of the consolidated actions directed by the White House to all U.S. agencies, excluding intelligence agencies, see Appendix I: Consolidated Table of Actions in Proposed Memorandum for the Heads of Executive Departments and Agencies (whitehouse.gov).

This research discusses how the EO on AI impacts executive leadership, along with recommendations for immediate action as shown in Figure 1.

Figure 1: The Impact of the "U.S. Executive Order on AI"



**The Impact of the "U.S. Executive Order on AI"**

| Impacts | Top Recommendations |
|---|---|
| The EO sends a strong message to the private and public sector: AI is not a new technology, it's a leadership responsibility. | • Accelerate scenario and use planning by engaging in methodical approaches.<br>• Create an "AI biometric dashboard" to derive an executive summary to monitor the vital signs of organizational alignment of AI to various scenarios and use cases proposed, piloted and funded by the organization. |
| The EO, in effect, redistributes the risks of loss associated with responsible AI failures, which will affect current and future AI investments and the projected corresponding ROI. | • Adapt your AI strategy and AI investment priorities to address new or elevated risks of harm and loss from responsible AI failures.<br>• Start by identifying and quantifying the potential AI risks of harm most relevant to stakeholder priorities.<br>• Adopt and promote as a "effort as a stand-alone enterprisewide best practice" to avoid responsible AI failures. |
| The EO has major impacts on U.S. federal agencies, and those who work with them. It will undoubtedly have a significant impact on the global AI vendor and solutions market, including to address the new standards for the development, functionality, use and output of AI. | • Prepare for EO-driven regulatory compliance — and a safer AI future in general — by working with stakeholders (such as legal, finance, government relations and assurance) across the organization to implement a comprehensive AI trust, risk and security management program.<br>• Use Gartner's AI trust, risk and security management (TRiSM) framework to identify technology solutions that support this endeavor. |

Source: Gartner

805298_C

Gartner

## Impacts and Recommendations

# AI is Not a New Technology, It's a Leadership Responsibility

*The EO reflects a new "inflection point," The proof points are in the execution by the combination of government agencies, academia and the business sector. Not necessarily in that order. It's likely the reverse order.*

## EO Enumerated Priorities and Principles

The EO reflects a strong and visible signal by the U.S. that "AI" is not a new technology, it's a leadership responsibility. It's the responsibility of business, government, academia and citizens at large. It sets forth specific priorities and principles that can be grouped into the following four categories:

- **Overall safety, security and trustworthy focus:**
  - AI must be safe and secure.
  - Citizens and consumers must be protected.
  - Privacy and civil liberties must be protected.
  - Equity and civil rights must be advanced.
- **Competitive advantage — Industry:**
  - Organizations must comply with a number of guidelines, guardrails, transparency and U.S. government oversight in the areas of AI model development and use.
  - Enterprises contracted by the U.S. federal government for AI development and government use of AI must watermark the government's AI/synthetic outputs.
  - The AI marketplace for AI and related technologies should be fair, open and competitive. Unlawful collusion/risks from dominant firms' use of key assets (e.g., semiconductors, computing power, cloud storage and data) to disadvantage competitors must be stopped.

- **Competitive advantage — Global economic U.S. power:**
  - Support must be extended to American workers.
  - The U.S. should lead the way to global societal, economic and technological progress.
- **US Government — Internal operating model and execution:**
  - The U.S. federal government's own use of AI must be managed.
  - Risks from the federal government's own use of AI must be managed, including taking steps to attract, retain, and develop public service-oriented AI professionals, including from underserved communities.
  - U.S. federal government agencies must establish a critical operating model for execution and governance, appoint a designated chief artificial intelligence officer and immediately commence analysis and guidelines for specific areas of scope.

With the U.S. being a bellwether for economic and innovation, the EO and related announcements serve several different purposes. Accordingly, for executive leaders, Gartner recommends taking the following steps:

**Action items:**

- **Mandate organizational alignment:** this EO is doing more than just hypothesizing or "analyzing." By requiring nearly every government agency to examine how and where AI is relevant to their jurisdictions of policy-setting and regulation, the U.S. is taking a major step in advancing a sectoral approach to AI governance.
- **Recognize and require an iterative, dynamic and ongoing process (in perpetuity)** where the organization ensures the formal structures, operating model and culture and ethos is aligned to the priorities and principles of EO. *This is not a nice to have or wait and see moment.* This is an inflection point as the posture and gait of organizations going through crawl, walk, run stages will shape the foundation. The EO is a call to action for both academia and the business sector. Enterprises of all sizes and shapes need to take action rather than just take a "wait and see" approach.

- **Create an “AI biometric dashboard“ for the organization:** In order to continually monitor the critical parameters of organizational alignment to AI in various scenarios and use cases, executive leaders need to derive an executive summary for leadership. Tally the scenarios you expect for AI use cases and applications in your organizations and provide regular updates to executives and other key stakeholders.

## Risk of Loss Redistribution

*To ensure change in model and human behaviors that can foster responsible AI innovation, development and use, the EO, in effect, redistributes risks of loss by shifting the risk of bearing the financial and nonfinancial costs of harm from responsible AI failures.*

The EO drives heightened expectations for a more safe, secure, trustworthy development and use of AI, alongside introducing a corresponding redistribution of risk of loss and harm from responsible AI failures.

Executive leaders should focus on what new risks of harm or loss may be introduced within their development or use of AI, including GenAI, and adapt their efforts to reduce the likelihood of resulting AI harms.

*The EO places AI and LLMs alongside products with high potential for intellectual property (IP) value but often considered too unsafe for human consumption — like pharmaceuticals, medical devices, cosmetics and foods — without U.S. oversight or approval prior to entering the commercial market.*

### Risk of Loss of Return on AI Investment

The EO requires private companies, including model creators and developers, to preemptively test their LLMs for specific safety concerns; it also specifies “red-teaming” as the testing methodology. Effective 30 January 2024, companies must report the resulting documentation of safety testing practices and results to the U.S. federal government.

Whether your organization is subject to this requirement turns on a primary objective trigger of whether the AI model was trained on broad data, generally uses self-supervision and contains at least tens of billions of parameters. For the associated text from the EO regarding this safety testing requirement, see Note 1.

For companies that have the resources required to train and run these LLMs, they will absorb a new risk of loss, including from the risk of model production delay or modification costs due to safety testing challenges and related issues. These alone can undermine financial investment or otherwise delay enterprise returns from AI investments.

Action item:

- Evaluate and compare LLMs that can drive maximum alignment with responsible AI outputs, while generating value/harm reduction for employees, customers and national security. Use Quick Answer: How Do I Compare LLMs? to get started.

### Risk of Loss of IP Rights

For companies and individuals, AI technology presents innovation opportunities for new IP assets. However, as with any potentially valuable asset, there are corresponding risks of loss that can undermine the initial and ongoing financial and nonfinancial investment.

*AI, including GenAI, can be generated by, and can generate itself, IP assets. But creation of an IP asset does not equate with IP protection of the asset.*

While the financial investments in an invention asset can be protected against competition through the U.S. patent protection process, the real value of a patent is market exclusivity, which somewhat ironically is undermined by the powerful creative capabilities of the AI, including GenAI, technologies themselves.

This also muddies the waters regarding who has the proprietary right to reuse or monetize IP assets generated by AI and GenAI, which necessarily recalculates potential risks of IP infringement liability.

In either case, commercial enterprises incur a “difficult to calculate” potential risk of loss because of either yet-to-be-determined legal interpretations of permissible fair uses of the newly created content, or known and unknown unauthorized uses by third parties.

The EO directs the Departments of Homeland Security and the Department of Justice to develop training, guidance, and other resources to combat and mitigate AI-related IP risks, including AI-related IP theft and other IP risks. At this time, though, this directive does not apply to the private sector. If the private sector does not self-regulate here, a larger portion of the risk of loss of IP could fall onto the private sector, compared to the public sector.

Action items:

- Work with your general counsel to determine applicability of the EO to your business, including the nuances for global companies or those located outside of the U.S.
- Work with your innovation teams to prioritize AI use cases with the most opportunity for increased clear and established IP rights while mitigating the risk of unauthorized use or theft of your training data or model output.
- Regarding GenAI models, be mindful that current legal precedent does not adequately specify ownership of content created by GenAI models.
- Given the newness of the technology, be aware that application of copyright and IP law will remain unclear. While there are emerging and evolving standards for training AI models and authenticating content, these do not yet include standards for identifying copyrighted material. For more guidance, see 2024 Legal, Compliance and Privacy Hot Spots.

### Risk of Loss from Human Harms

The EO calls out several specific human harms from AI and GenAI interaction and use. Those include fraud, discrimination and threats to privacy. The EO calls out several specific agencies (and related industries) that need to prioritize AI and GenAI use cases for generally acceptable implementation. Those include healthcare, the transportation sector, education and communication networks.

To mitigate those harms, within and outside of those industries, the EO suggests that organizations conduct more proactive due diligence on and monitoring third-party AI services for greater transparency and explaining use of models.

The EO also requires that "companies, individuals, or other organizations or entities that acquire, develop or possess a potential large-scale computing cluster report any such acquisition, development or possession. Including the existence and location of these clusters and the amount of total computing power available in each cluster."

There are different definitions of "large-scale computing cluster" depending upon certain variables, including time periods from the 30 October 2023 date of the EO. Accordingly, the impact of these requirements is one to watch. For the associated text from the EO regarding this disclosure requirement, see Note 1.

For U.S. federal agencies, the EO provides mitigation guidance by prioritizing the identification of approved AI use cases. The efforts, though, have to be continuous since the training data and respective outputs are dynamic. This requires a level of discipline that may be difficult to implement at scale.

So while the EO focus on mitigating harms may include uncertainty and corresponding risks of loss to short-term returns on AI investment, by focusing on the long-term AI and company performance, along with increases in good will, the risk of human harms may (appropriately) be shifted away from those who engaged in efforts to identify and mitigate human harms to those who did not.

Action items:

- If you are connected to large-scale computer cluster activities, monitor the upcoming guidance from U.S. agencies as they continue to define the parameters of these activities and clarify the triggering events for disclosure requirements.
- Focus on the AI use cases that incorporate early in the model process harm reduction functionality and observation of the technology outputs that can be reviewed by a human in the loop. The primary use cases here should focus on individuals in their capacity as citizen, consumer, employee, patient, student and other similar roles that include the rights and liberties called out in the EO.
- Assess your own and third-party AI products on a continuous basis to improve risk mitigation of unintended harmful consequences to humans and the communities in which they interact with AI.

### Increase Identification and Mitigation of Risk of Loss

*Invest in "effort as a stand-alone best practice" to improve the safety, security and trustworthiness of AI and GenAI.*

Effort as a stand-alone practice to mitigate risk of loss, regardless of result, can positively impact financial and non-financial returns of AI and GenAI investments.

Harvard Business Review research shows that business partners and customers are more forgiving of negative outcomes or losses when enterprises "try" to avoid them. [4] Public sentiment dictating the scale of revenue loss is often based more on "what did you do to try to avoid this outcome" than the actual outcome, often, no matter how undesirable.

Conversely, insurers, for example, may deny coverage for business interruption caused by AI failures if enterprises did not make enough effort to mitigate the loss. When regulators investigate alleged noncompliance, evidence of a *lack of effort* to attempt compliance, regardless of the outcome, can escalate financial penalties.

With respect to IP infringement liability — now top of mind given ChatGPT's rapid acceleration into the commercial and personal mainstream market — risk of loss is heightened as courts may award a penalty of "triple damages" depending upon the effort to mitigate the risk of harm to third parties, including copyright holders.

Action items:

- Promote "effort as a stand-alone best practice" by using AI and GenAI use case narratives to show that even the least amount of effort or change in process can lead to the most amount of loss avoidance, leading to increased responsible AI rather than the AI failures. For additional guidance, see 3 Proactive Ways CDAOs Can Derisk Enterprise Value.

- Update your policies to isolate risks specific to different domains, creating a fast-path to innovation and value for low-risk data/practices and a heightened level of governance for higher risk components. See Overcoming Data Quality Risks When Using Semistructured and Unstructured Data for AI/ML Models.

## Prepare Today for More Regulated AI Tomorrow

> **The U.S. executive order on AI aims to reflect the values of "the people who build it, the people who use it, and the data upon which it is built."**
>
> *— Source: Executive Order on the Safe, Secure and Trustworthy Development and Use of Artificial Intelligence, The White House*

The EO has major impacts on U.S. federal agencies, and those from the private sector and academia who work with them. It will undoubtedly have a significant impact on the global AI vendor and solutions market, including those that serve the U.S. government. Those affected will have to abide by new standards and rules that govern the development, functionality, use and output of AI in several different domains.

### AI Safety Testing

The EO is especially clear and helpful in the first section that requires companies that develop the “most powerful” AI systems to share safety test results and other related information with the U.S. government. The U.S. president can indeed order this sharing under the Defense Production Act and, in some cases, may give the president the power to impact production of the models. The mandate that the National Institute of Standards and Technology (NIST) set standards for security and safety standards is a critical piece of this action item and is most welcome.

Action item:

- Stay aware (and if possible ahead) of AI system safety test results and use AI models and systems that meet the new U.S. federal safety standards that will emerge and evolve over time.

### Biological Hazards

The EO orders the creation of standards to make sure AI is not used to engineer harmful biological organisms — like deadly viruses or medicines.

The EO will enforce this provision only by using the emerging standards as a baseline for federal funding of life-science projects. Given the EO focus on industry cooperation, we can expect dialogue around whether these emerging standards may go further and apply to a wider class of model developers and model consumers. Or whether these standards would be applicable to private capital or any non-federal government funding bodies and sources (like venture capital). Executive leaders should watch for how these standards will be enforced and what the penalties are for non-compliance.

Action item:

- Evaluate how this order will impact your organization, if you are involved in public/private or other life-science projects and whether or not they are federally funded.

*AI Content Authentication and Watermarking Guidance*

This guidance applies to federal government agencies that create content, and as such also applies to technology providers servicing related government functions.

The order to develop AI content authentication and watermarking guidance is needed to stem the onslaught of misinformation and disinformation proliferating through information networks.

Under the EO, the Department of Commerce will be "establishing standards and best practices for detecting AI-generated content and authenticating official content." It notes that "federal agencies will use these tools to make it easy for Americans to know that the communications they receive from the U.S. government are authentic — and set an example for the private sector and governments around the world." [5]

The executive order does not appear to require industry players or government agencies to use this technology, in part because of the limitations of the nascent technologies. However, the agencies tasked with developing a regulatory framework could encourage them to do this. And the White House has signaled its intention to do so.

Still in order to stem the proliferation of AI-generated misinformation and disinformation, the EO underscores that Americans may need new mechanisms (including enforcement of laws that apply to major social media networks, news outlets and other platforms) before these tools, now or when fully developed, can make a notable difference.

**Action item:**

- Stay up to date on emerging standards and guidance in AI content authentication and start using them as soon as you can.

*Privacy*

The EO calls for a federal privacy law, which depends on Congress's ability to pass a comprehensive federal data privacy legislation. Despite efforts, this has not happened for decades so it is unlikely to happen in the next three years. We expect that data privacy and protection standards continue to evolve.

**Action item:**

- Monitor development of data privacy standards and plan to implement them as best practice, whether or not you are in a federal agency.

### Still More Work to Do

Overall, the scope of the EO is comprehensive when it comes to AI safety and security. But the devil is in the details — who does it apply to, how will it be enforced? How are some of these terms defined (e.g., bias, fairness, responsible AI?).

Congress needs to act and new regulations must be established. The EO is a welcome start but more work is already being done, and more is on the way. In particular, the following may cause some confusion until several issues are clarified:

- Who sets the definition for, and what constitutes the, “most powerful” AI systems now and in the future? (see Note 1 for part of the answer).
- How does this apply to open source AI models?
- How will content authentication standards be enforced across social media platforms and other popular consumer venues?
- Which sectors are in scope when it comes to compliance?
- How will the proposed standards be enforced?

Still organizations can benefit from the EO and use its outputs to get a head start on making their own AI applications and systems more safe and secure, well before the regulations apply to them directly.

---

*Get Organized: AI trust, risk and security management is a team effort*

---

Executive leaders must work with stakeholders (including legal, finance, government relations, etc.) across their organization to implement a comprehensive AI trust, risk and security management program. Optimally, a dedicated AI risk management office should be established to work across organizational lines; for example legal, compliance, privacy, security, data and analytics, business functions, AI model development and more.

But when a dedicated office is not feasible, a fusion team with clear responsibilities must regularly monitor and adhere to established AI trust, risk and security management priorities. A 2021 Gartner "AI in Organizations" survey showed that organizations who organize as such achieve improved business results from AI projects and move more projects into production (see AI in Organizations Survey Results: Managing AI Risk Leads to Positive Business Outcomes).

Regulated industries in particular are under pressure to establish such programs, in order to comply with upcoming regulations. The EO will surely lead to regulations and laws that apply well beyond the federal government sector. Organizations should prepare now for compliance, and more importantly for a safer, more secure and more reliable AI future.

**Action item:**

- Organizations can use Gartner's AI TRiSM framework to identify technology solutions that support this endeavor. See Top Strategic Technology Trends for 2024: AI Trust, Risk and Security Management and Innovation Guide for Generative AI in Trust, Risk and Security Management.

---

### *Executive leadership starts now*

---

The EO provides a potential glimpse as to the future vision, guidelines and standards for AI that are yet to be developed. It also sharpens the priorities for open AI.

**Action items:**

Start now, for safe, secure and trustworthy AI by:

- **Identifying the right players responsible for compliance with the EO's direction.** The typical participants in AI-related decisions include representatives of legal, compliance, risk management, privacy, security, procurement, DEI, D&A leaders, and SMEs (especially in the customer service areas). Most of these people are new to AI and they should acquire AI literacy. To meet the challenges of the EO, you can organize to include AI governance, AI ethics and responsible AI committees.

- **Making sure the decisions about AI are transparent.** Transparency of AI-related decisions, such as why AI use cases, tools and approaches were selected and what aspects of decisions were discussed and agreed upon is a good immediate step — this involves documenting choices and validation procedures for each AI initiative or use case (see 4 AI Governance Actions to Make a Swift Business Impact).

- **Ensuring a feedback loop for full transparency,** where AI users or individuals affected by AI have an easy means to provide feedback — this will allow the capture of nuances and exceptions early on and use feedback to improve AI solutions. Even if an AI system design includes all the safe, secure and trustworthy AI considerations, the AI outcomes cannot be anticipated deterministically. Exception handling is a mechanism aimed at capturing what has not been accounted for within or between elements; but that is inevitable as a result of previously unknowable behavior.

## Acronym Key and Glossary Terms

| | |
|---|---|
| US Executive Order | A U.S. executive order is a signed, written, and published directive from the president of the United States that manages operations of the federal government, which by extension manages the private and other sectors. Executive orders are not legislation; they require no approval from Congress, and Congress cannot simply overturn them. Only a sitting U.S. president may overturn an existing executive order by issuing another executive order to that effect. |
| Defense Production Act of 1950 | The Defense Production Act of 1950 gives the president the authority to expedite and expand industrial production in order to promote national defense.6,7 |
| dual-use foundation models(as defined in the EO) | The term "dual-use foundation model" as defined in the EO means an AI model that: Is trained on broad data; generally uses self-supervisionContains at least tens of billions of parametersIs applicable across a wide range of contextsThat exhibits, or could be easily modified to exhibit, high levels of performance at tasks that pose a serious risk to security, national economic security, national public health or safety, or any combination of those matters, such as by:Substantially lowering the barrier of entry for non-experts to design, synthesize, acquire, or use chemical, biological, radiological, or nuclear (CBRN) weapons.Enabling powerful offensive cyber operations through automated vulnerability discovery and exploitation against a wide range of potential targets of cyber attacks.Permitting the evasion of human control or oversight through means of deception or obfuscation.1Models meet this definition even if they are provided to end users, with technical safeguards that attempt to prevent users from taking advantage of the relevant unsafe capabilities.1 |
| AI red teaming (as defined in the EO) | The term "AI red-teaming" as defined in the EO means a structured testing effort to find flaws and vulnerabilities in an AI system, often in a controlled environment and in collaboration with developers of AI. Artificial intelligence red-teaming is most often performed by dedicated "red teams" that adopt adversarial methods to identify flaws and vulnerabilities, such as harmful or discriminatory outputs from an AI system, unforeseen or undesirable system behaviors, limitations or potential risks associated with the misuse of the system.1 |
| Watermarking (as defined in the EO) | The term "watermarking" as defined in the EO means the act of embedding information, which is typically difficult to remove, into outputs created by AI. Including outputs such as photos, videos, audio clips or text for the purposes of verifying the authenticity |

| | |
|---|---|
| | of the output or the identity or characteristics of its provenance, modifications, or conveyance.1 |
| AI TRiSM | AI trust, risk and security management (AI TRiSM) ensures AI model governance, trustworthiness, fairness, reliability, robustness, efficacy and data protection. This includes solutions and techniques for model interpretability and explainability, AI data protection, model operations and adversarial attack resistance (see Definition of AI TRiSM — Gartner Information Technology Glossary). |

## Evidence

[1] Executive Order on the Safe, Secure and Trustworthy Development and Use of Artificial Intelligence | The White House, The White House.

[2] Blueprint for an AI Bill of Rights, The White House.

[3] OMB Releases Implementation Guidance Following President Biden's Executive Order on Artificial Intelligence, The White House.

[4] Research: Public Opinion Is Not Enough to Hold Companies Accountable, Harvard Business Review.

[5] FACT SHEET: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence, The White House.

[6] 50 USC Ch. 55: Defense Production, Office of the Law Revision Counsel, United States Code.

[7] The Defense Production Act of 1950: History, Authorities, and Considerations for Congress, Congressional Research Service.

Safe, Secure and Trustworthy Development and Use of Artificial Intelligence, Federal Register.

Department of Commerce to Undertake Key Responsibilities in Historic Artificial Intelligence Executive Order, U.S. Department of Commerce.

At the Direction of President Biden, Department of Commerce to Establish U.S. Artificial Intelligence Safety Institute to Lead Efforts on AI Safety, U.S. Department of Commerce.

Proposed Memorandum for the Heads of Executive Departments and Agencies, Executive Office of the President, Office of Management and Budget Washington, D. C. 20503.

Draft OMB Memo Details 10 New Requirements to Manage AI, Federal News Network.

The White House Is Preparing for an AI-Dominated Future, The Atlantic.

More Tech Companies Take White House AI Safety Pledge, Axios.

## Note 1: Associated Texts Regarding Certain Disclosure and the Safety Testing Requirements Under the EO

The italicized text here has been presented as it appears verbatim in the EO from which it was sourced as indicated below.

The EO reads: *"... the Secretary of Commerce shall require (i) Companies developing or demonstrating an intent to develop potential dual-use foundation models to provide the federal government, on an ongoing basis, with information, reports or records regarding the following:*

*(A) Any ongoing or planned activities related to training, developing or producing dual-use foundation models, including the physical and cybersecurity protections taken to assure the integrity of that training process against sophisticated threats.*

*(B) The ownership and possession of the model weights of any dual-use foundation models, and the physical and cybersecurity measures taken to protect those model weights.*

*(C)The results of any developed dual-use foundation model's performance in relevant AI red-team testing based on guidance developed by NIST pursuant to subsection 4.1(a)(ii) of this section, and a description of any associated measures the company has taken to meet safety objectives, such as mitigations to improve performance on these red-team tests and strengthen overall model security. Prior to the development of guidance on red-team testing standards by NIST pursuant to subsection 4.1(a)(ii) of this section, this description shall include the results of any red-team testing that the company has conducted relating to lowering the barrier to entry for the development, acquisition, and use of biological weapons by non-state actors; the discovery of software vulnerabilities and development of associated exploits; the use of software or tools to influence real or virtual events; the possibility for self-replication or propagation; and associated measures to meet safety objectives.* See the EO section 4.2(a)(i)."

Noting that the set of technical conditions for models and computing clusters that would be subject to the reporting requirements of change over time, see EO Section 4.2(b), which states: "*The Secretary of Commerce, in consultation with the Secretary of State, the Secretary of Defense, the Secretary of Energy, and the Director of National Intelligence, shall define, and thereafter update as needed on a regular basis, the set of technical conditions for models and computing clusters that would be subject to the reporting requirements of subsection 4.2(a) of this section*.

*… Until such technical conditions are defined, the Secretary shall require compliance with these reporting requirements for:*

*(i) Any model that was trained using a quantity of computing power greater than 10 $^{26}$ integer or floating-point operations, or using primarily biological sequence data and using a quantity of computing power greater than 10 $^{23}$ integer or floating-point operations.*

*(ii) Any computing cluster that has a set of machines physically colocated in a single data center, transitively connected by data center networking of over 100 Gbit/s, and having a theoretical maximum computing capacity of 10 $^{20}$ integer or floating-point operations per second for training AI.*

*(ii) Companies, individuals, or other organizations or entities that acquire, develop, or possess a potential large-scale computing cluster to report any such acquisition, development, or possession, including the existence and location of these clusters and the amount of total computing power available in each cluster."* See the EO section 4.2(a)(ii)."

## Recommended by the Authors

Some documents may not be available as part of your current Gartner subscription.

A Comprehensive Guide to Responsible AI

Top Strategic Technology Trends for 2024: AI Trust, Risk and Security Management

Activate Responsible AI Principles Using Human-Centered Design Techniques

Introduce AI Observability to Supervise Generative AI

How to Keep AI From Getting Out of Control

Emerging Tech Roundup: Navigating Confusion Around Synthetic Data

Survey Analysis: An AI-First Strategy Leads to Increasing Returns

Research Roundup: Realizing Value From Artificial Intelligence (AI)