As the development of artificial intelligence (AI) systems accelerates globally and the benefits and risks of their use become evident, calls for government regulation in the U.S. and abroad have accelerated. Two significant governmental developments occurred in the past month to respond to these calls. In an executive order issued at the end of October, President Joe Biden revealed a comprehensive set of guidelines and policy goals for the future of AI development and regulation. Less than a month later, the U.S., U.K., and more than a dozen other countries unveiled the first international agreement on AI safety and security. Though differing in scope and actionable initiatives, the two documents reflect an international acknowledgment of the global impact and risks posed by AI systems, as well as an urgency to create proactive policies for their regulation.

Key Takeaways

- President Biden issued Executive Order 14110 on "Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence" on October 30, 2023, with the goals of establishing standards for AI safety and security; protecting privacy, consumer, and worker rights; advancing equity; promoting global innovation and competition; and advancing American leadership around the world.
- The Order sets forth various policy goals, tasks, and guidance for federal agencies to implement in the next year.
- Federal agencies are directed to use their regulatory powers to monitor and mitigate risks, create guidelines to shape industry standards, develop uses for AI technology, and implement such technologies safely.
- On November 27, 2023, the U.S., U.K., and 16 other countries entered into a landmark international agreement on cybersecurity in AI, emphasizing a "secure by design" approach to AI systems development.
- The non-binding agreement consists of a set of guidelines addressing four key areas of the AI system development life cycle (secure design, secure development, secure deployment, and secure operation and maintenance) and outlines recommendations to reduce overall risk.
- The first of its kind, the agreement recognizes the importance of international collaboration in guiding AI development and establishing a cohesive framework for responsible AI practices.

White House Executive Order on Artificial Intelligence

On October 30, 2023, President Biden issued Executive Order 14110 on "Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence,"[1] taking a significant step toward shaping the future of AI[2] and its regulation. The Order, which reflects growing calls for federal guidance on AI from public and private stakeholders, focuses on establishing a framework for safe, secure, and trustworthy AI development, focusing on ethical innovation, national security, and global cooperation. The Order builds on the White House's October 2022 "Blueprint for an AI Bill of Rights"[3] and the National Institute of Standards & Technology's (NIST) January 2023 "Artificial Intelligence Risk Management Framework."[4]

The Order is broad in scope, covering a spectrum of industries and issues, including the establishment of new standards for AI safety and security; protection of privacy; advancement of equity and civil rights; support of consumers, patients, and employees; and promotion of innovation and competition.

Although the Order is primarily applicable to federal agencies, it reflects a vision and roadmap for AI regulation intended to guide both industry standards and future federal legislation.

1 Full text available at Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence.
2 The definition of "artificial intelligence," or "AI," is as set forth in 15 U.S.C. 9401(3): "a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations or decisions influencing real or virtual environments." The Order is therefore broad in scope, applying to any machine-based system that makes predictions, recommendations or decisions, not only generative AI.
3 Full text available at Blueprint for an AI Bill of Rights.
4 Full text available at Artificial Intelligence Risk Management Framework.

DECEMBER 13, 2023

CONTACT

SUSANNA BAGDASAROVA
SBagdasarova@babstcalland.com
412.394.5434

MARY H. BINKER
MBinker@babstcalland.com
412.394.6810

CHRISTIAN A. FARMAKIS
CFarmakis@babstcalland.com
412.394.5642

JUSTINE M. KASZNICA
JKasznica@babstcalland.com
412.394.6466

Pittsburgh, PA
Two Gateway Center
603 Stanwix Street
Sixth Floor
Pittsburgh, PA 15222
412.394.5400

BABSTCALLAND.COM

The Order sets out eight principles and priorities to guide policymaking on AI systems:

- AI must be safe and secure, requiring *robust, reliable, repeatable, and standardized evaluations* of AI systems, as well as mechanisms to test, understand, and mitigate risks.
- The U.S. should promote *responsible innovation, competition, and collaboration* through investments in AI-related education, training, development, research, and capacity as well as by opposing monopolies and unlawful collusion with respect to key assets.
- The responsible development and use of AI require *a commitment to supporting American workers* through job training and education, both to prevent AI systems from being deployed in ways that negatively impact employee rights and to use AI in ways that increase human productivity.
- AI policies must be consistent with the Biden administration's policy of *advancing equity and civil rights* and be structured to prevent deepening inequities, new types of harmful discrimination, and online and physical harms.
- The federal government must enforce existing consumer protection policies and enact appropriate safeguards against *fraud, bias, discrimination, and privacy infringement* to protect Americans who are increasingly using AI and AI-enabled products, particularly in critical fields such as healthcare, financial services, education, housing, law, and transportation.
- Policies and tools must be developed to protect Americans' *privacy and civil liberties* to ensure that personal data collection, use, and retention is done in a lawful and secure manner that mitigates privacy and confidentiality risks.
- The risks arising from the federal government's own use of AI must be mitigated, and it must increase its ability to internally *regulate, govern, and support responsible use of AI* including, but not limited to, the recruitment of AI professionals.
- The U.S. should be a global leader for *societal, economic, and technological progress*, and responsibly deploy technology through engagement with its international allies and partners to develop an AI governance framework and ensure that AI benefits the world rather than increasing or exacerbating existing harms and inequities.

Building on this foundation, Sections 4 through 11 of the Order each correspond to one of the eight guiding principles, setting out a host of practical policy goals, tasks and guidance for federal agencies to implement in the next year. The lengthy Order contains directives for nearly all 15 executive departments to use their regulatory powers to monitor and mitigate risks, develop uses for AI technology, and implement such technologies safely. Certain directives are highlighted below:

- The Order tasks NIST with establishing a series of guidelines for AI use and development, including (i) best practices to promote industry standards for safe, secure and trustworthy AI systems, (ii) a companion to the AI Risk Management Framework for generative AI, (iii) a companion to the Secure Software Development Framework for generative AI and dual-use foundation models,[5] (iv) AI auditing and evaluation guidelines with a focus on cybersecurity and biosecurity, and (v) procedures and processes for AI developers to conduct red-team testing[6] of dual-use foundation models.
- The Order imposes recordkeeping and reporting requirements on developers of dual-use foundation models, including reporting of red-team safety test results and other critical information on model training and physical and cybersecurity measures. Developers will also be required to report the acquisition, development, or possession of large-scale computing clusters, including their location and the total amount of computing power available in each. Infrastructure as a Service (IaaS) products tested or sold by foreign persons will also be subject to recordkeeping and reporting requirements.
- Various agencies with regulatory authority over critical industries are directed to assess and develop mitigation strategies for AI-related critical infrastructure vulnerabilities, including critical failures, physical attacks and cyberattacks.
- The Department of Commerce is tasked with creating guidance for content authentication and watermarking of AI-generated content in government communications, in order to increase transparency and public trust and encourage adoption of such standards by the private sector.
- The Department of Labor is instructed to create best practices for employers to mitigate AI risks and maximize AI benefits in the workforce, paying careful attention to the intersection of AI and worker protections.
- The State Department and Department of Commerce must establish international frameworks for AI regulation, and the White House plans to collaborate with international partners and organizations for global and consistent AI regulation. The initial results of such collaboration are evident in the international agreement recently entered into by the U.S., as discussed below.
- The Order also calls on Congress to enact federal data privacy legislation and establishes a White House Artificial Intelligence Council to coordinate the implementation of AI-related policies by executive agencies.

Sweeping in its scope, the Order seeks to be comprehensive and consistent in addressing topics and sectors most keenly affected by the development and use of AI systems. Such directives will inevitably impact federal procurement policy and requirements for government contractors, a historically powerful tool to develop industry standards, even without legislative action. Given the constant and evolving nature of AI development, governments around the world are struggling to legislatively address the risks posed by this technology at a speed that matches AI development. By presenting a coherent set of guidelines and practical and strategic goals, as well as implementation deadlines for federal agencies and departments (some as short as 90 to 270 days), the

5 Defined as "an AI model that is trained on broad data; generally uses self-supervision; contains at least tens of billions of parameters; is applicable across a wide range of contexts; and that 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..."

6 Defined as "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…[it] 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."

Order proactively seeks to mitigate risks before the potential harms of AI technology become entrenched. Federal agencies are already working on executing their directives under the Order, with the White House Office of Management and Budget releasing a draft policy[7] on Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence on November 1 and the U.S. Cybersecurity and Infrastructure Security Agency (CISA) releasing its implementation plan, "Roadmap for Artificial Intelligence"[8] on November 14.

International Agreement on Guidelines for Secure AI System Development

On November 27, 2023, the U.S., U.K., and 16 other countries[9] entered into a landmark international agreement on cybersecurity in AI systems.[10] The 20-page document titled "Guidelines for Secure AI System Development" is a non-binding series of general guidelines and recommendations emphasizing a "secure by design" approach to AI systems development. The Guidelines were collaboratively developed by CISA, the U.K. National Cyber Security Center, and cybersecurity agencies from each signatory nation, in partnership with industry experts, including Amazon, Google and OpenAI, and other organizations such as the Software Engineering Institute at Carnegie Mellon University. The Guidelines reaffirm an increasing understanding that international collaboration in guiding AI development and the creation of a cohesive framework for responsible AI practices are necessary for international cybersecurity.

The release of the Guidelines follows the AI Safety Summit hosted by the U.K. in early November, which was attended by Vice President Kamala Harris, as well as other international leaders, technology executives, and industry experts. The Summit established the Bletchley Declaration on AI Safety.[11] Signed by leaders of 29 countries, the Bletchley Declaration affirms a joint commitment to developing AI safely and responsibly and acknowledges a shared responsibility in addressing the risks posed by AI. Although China was a signatory to the Bletchley Declaration, it is notably absent from the Guidelines themselves.

The Guidelines identify four key areas of the AI system development life cycle: secure design, secure development, secure deployment, and secure operation and maintenance. AI developers are to consider the Guidelines along with industry best practices in the areas of cybersecurity, risk management, and incident response. The use of both the Guidelines and best practices is the international community's effort to ensure AI systems "*function as intended, are available when needed, and work without revealing sensitive data to unauthorized parties.*" Cybersecurity is therefore both a pre-condition to AI system safety and also a necessary component of the development process from beginning to end. This approach places the responsibility for downstream security outcomes on the providers of AI components.

Each of the four key areas are supported by a series of "*considerations and mitigations*" intended to reduce overall risk in AI system development. These considerations include monitoring of AI systems for potential threats and abuse (including hacking), securing of the AI supply chain and vetting of hardware and software components developers, identification of AI vulnerabilities and data tampering opportunities, development of robust incident management procedures, the release of AI products to users only after security testing, and collaboration across the global AI-ecosystem to share best practices across industry, academia, and government.

The significance of the Guidelines is summarized by CISA Director Jen Easterly:

> "*The release of the Guidelines for Secure AI System Development marks a key milestone in our collective commitment—by governments across the world—to ensure the development and deployment of artificial intelligence capabilities that are secure by design...The domestic and international unity in advancing secure by design principles and cultivating a resilient foundation for the safe development of AI systems worldwide could not come at a more important time in our shared technology revolution. This joint effort reaffirms our mission to protect critical infrastructure and reinforces the importance of international partnership in securing our digital future.*"[12]

This landmark agreement seemingly marks the first of many international commitments to development of consistent AI-systems standards, providing the basic cybersecurity component of an international framework on AI.

---

7 Full text available at Proposed Memorandum for the Heads of Executive Departments and Agencies.
8 Full text available at Roadmap for Artificial Intelligence.
9 Australia, Canada, Chile, Czechia, Estonia, France, Germany, Israel, Italy, Japan, New Zealand, Nigeria, Norway, Poland, the Republic of Korea, and Singapore.
10 Full text available at Guidelines For Secure AI System Development.
11 Full text available at The Bletchley Declaration.
12 See DHS/CISA and UK NCSC Release Joint Guidelines for Secure AI System Development.