DENTONS GLOBAL ADVISORS | ASG ALBRIGHT STONEBRIDGE GROUP

ASG ANALYSIS:
White House Lays Down Marker with Executive Order on AI, But Offers Few Concrete Measures

November 9, 2023

Key takeaways

- On October 30th, the Biden administration released the highly anticipated Executive Order (EO) on Safe, Secure, and Trustworthy Artificial Intelligence (AI). The order is sweeping in scope and adopts a whole-of-government approach to overseeing elements of AI model development and deployment, while laying out some initial frameworks for monitoring, and regulation, particularly of advanced or "frontier" models.
- The order builds on and consolidates many of the administration's previous actions on AI, including the *White House Voluntary Commitments,* the *Blueprint for an AI Bill of Rights*, and the U.S. national AI strategy. This time, however, the EO takes specific actions on several concerns that have been a key driver of the growing political urgency around AI issues in the U.S.
- The EO lays out a broad set of principles for federal agencies to develop guidance and offers a new set of specific measures designed to lay the groundwork for further government oversight of AI model development, including in future legislation.
- National security considerations are at the heart of the EO; with many of its provisions geared toward addressing concerns around biosecurity, cybersecurity, and the protection of critical infrastructure.
- Though limited in scope, the order imposes significant responsibilities on AI developers, including requiring companies working on advanced AI models to submit the results of their red teaming tests to the government. It further includes a Know-Your-Customer due diligence and reporting regime for companies working on advanced AI systems that could pose cybersecurity risks or provide access to computing power that could be used for very large AI training runs.
- At present, only a handful of large U.S. tech companies are developing AI systems powerful enough to fall under the scope of this part of the EO, but these performance thresholds are subject to additional review and could change down the road as the technology develops.

- But even as the order attempts to create guardrails around the use and application of the technology, it similarly includes provisions to foster its development. Consequently, the EO outlines support for smaller AI developers and researchers and calls for reforms to immigration policy to attract and retain AI talent.
- The order includes a major nod to the importance of multilateral action on advanced AI, and the contents of the order were discussed this week at the U.K. AI summit, attended by 28 countries, including China.
- Some important questions remain about the implementation. For example, it is unclear how many of the federal agencies will fund the new expanded mandates, or where they will find AI talent. The order also does not make recommendations on how the U.S. government should approach institutionalizing AI oversight, to ensure that the issue remains a priority through successive administrations.

Biden White House lays down clear marker on AI development

On October 30th, the Biden administration released its much-anticipated executive order on Artificial Intelligence (AI), marking a milestone in the government's efforts to develop effective guardrails on the rapidly evolving technology. The order came out two days before Vice President Kamala Harris and other high-level government officials such as Secretary of Commerce Gina Raimondo attended the U.K.'s AI Summit on Frontier AI Safety, allowing the Biden administration to showcase the order for world leaders in the U.K. and to gain support for its approach to AI governance.

The EO builds on, and is complementary to, many other AI governance documents issued by the administration and like-minded allies over the past year. These include an earlier executive order published in 2019 laying out the U.S.'s national AI strategy, the White House Voluntary Commitments, which 15 leading AI companies have endorsed since July, and the White House's Blueprint for an AI Bill of Rights, which the Biden administration released last year. The order is also consistent with developing multilateral efforts to establish principles and codes of conduct, such as the G7 Hiroshima process, which this week released its own voluntary code of conduct for companies.

The Biden administration is eager to demonstrate to a domestic political audience that it is acting on regulating AI, amid an increasing discussion about the risks of AI since the debut of the ChatGPT and other large language models (LLM) last year. With federal legislation on AI not likely to move forward in the last year of the Biden administration**,** the order includes some novel attempts to leverage existing statutes to place requirements on developers of AI. It also seeks to provide the government with increased visibility into which actors are developing large AI models that could have 'dual-use' applications, including when and how advanced AI systems are being trained in the cloud or other large computing clusters. Moreover, the order is also a clear statement of administration priorities in AI governance that will undergird U.S. engagement in multilateral fora on the issue.

National security concerns drive actions in the EO, but guidance is mostly focused on threat assessments for now

The EO is organized around eight guiding principles, including:

- creating new standards for AI safety and security;
- protecting user privacy;
- advancing equity and civil rights;
- protecting consumers, patients, and students;
- supporting workers;
- promoting innovation and competition;
- advancing U.S. leadership in AI technologies, and
- ensuring the responsible and effective government use of the technology

Across these principles, national security emerges as a central theme animating the Biden administration's approach to AI regulation. Among these concerns, threats to cybersecurity, biosecurity, and critical infrastructure sectors such as healthcare and finance emerge as core elements of the executive order**.** Over the past year, the rise of generative AI models and platforms based on them, such as ChatGPT, has prompted closer scrutiny of the novel ways in which next-generation or "frontier" models could pose national security threats. Examples of these concerns include the potential for threat actors to design malicious code for offensive cyber operations, develop novel pathogens for bioterrorism, and manipulate public opinion through mis/dis-information campaigns.

While many of these concerns are not new, conversations around the emergence of dual-use AI systems have changed the urgency of the policy conversation in Washington D.C., prompting different stakeholders to advance a range of policy proposals to help regulate AI, from increasing funding for the National Institute of Standards and Technology (NIST) to creating an "international agency" to police AI development and deployment, among others.

The EO, however, largely stays away from setting out granular prescriptions for how to tackle AI risks as they relate to national security. Rather (except for some cybersecurity provisions discussed below), the order directs federal agencies to undertake assessments of the risks of AI in critical sectors to understand the evolving landscape and provide recommendations for *future* regulatory actions. For example, the EO calls on agencies to take the following actions, among others:

a) [Assess] potential risks related to the use of AI in critical infrastructure sectors involved, including ways in which deploying AI may make critical infrastructure systems more vulnerable to critical failures, physical attacks, and cyber-attacks, and shall consider ways to mitigate these vulnerabilities.
b) [Assess] the ways in which AI can increase biosecurity risks, including risks from generative AI models trained on biological data, and makes recommendations on how to mitigate these risks.
c) [Consider] the national security implications of the use of data and datasets, especially those associated with pathogens and genomics studies, that the United States Government hosts, generates, funds the creation of, or otherwise owns, for the training of generative AI models, and makes recommendations on how to mitigate the risks related to the use of these data and datasets.

d) [Establish] criteria and mechanisms for ongoing identification of biological sequences that could be used in a manner that would pose a risk to the national security of the United States.
e) [Evaluate] AI model capabilities to present CBRN threats — for the sole purpose of guarding against those threats — as well as options for minimizing the risks of AI model misuse to generate or exacerbate those threats.

The lack of *specific* guidance on how to deal with emerging national security threats has provoked criticism about the order lacking teeth, but this approach is pragmatic**.** While conversations about AI and national security have intensified**,** questions remain about what specific policies will be effective in mitigating these risks and the potential costs that these policies may have in restricting AI-driven innovation. Moreover, the risks of next-generation AI systems are not fully understood, with AI developers themselves unsure of how their systems could be co-opted by threat actors. Against this backdrop, the administration's wait-and-see approach is prudent; allowing federal agencies to develop sector-specific risk assessments before laying out guardrails.

EO imposes stricter requirements for industry, but only a handful of players fall in scope

The only prescriptive guidance and legal requirements set out in the EO are for AI developers, but even these actions are limited to those developing the next generation of AI models. The EO devotes significant attention to these systems, which it refers to as "dual-use foundation models" and defines as "AI model[s] that [are] trained on broad data; generally [use] self-supervision; [contain] at least tens of billions of parameters; [are] applicable across a wide range of contexts; and that [exhibit], or could be easily modified to exhibit, high levels of performance at tasks t*hat [one again] pose a serious risk to security, national economic security, national public health or safety, or any combination of those matters.*"

The order leverages the United States Defense Production Act (DPA) to require companies to submit the results of their "red teaming" security tests to the government prior to public release. At the same time, the order requires companies developing these AI models to provide the government with regular updates about model training, performance, and security vulnerabilities. These reporting requirements, however, only apply to models trained above a certain and very high threshold of computing power, initially set at 100 yottaFLOPS (10^{26} FLOPS). This threshold is likely to be hit by a small number of companies in 2024 that are developing the next generation of advanced AI models and platforms, including the forthcoming versions of OpenAI's GPT and Meta's LLaMA foundational models, which would clearly be considered dual-use frontier models by this definition.

The threshold is roughly five times that required for the development of GPT-4, with one estimate putting the compute cost for this type of model at $250 million, limiting the number of companies that would be using this capability. U.S. officials see this provision as the start of a notification process for the government to gain visibility on who is training advanced models, and the threshold could be moved up as the technology advances. It is noteworthy that this is just a disclosure requirement, not a licensing one.

The requirements for government notification about training of specific AI models amounts to a know-your-customer (KYC) requirement. This also includes a requisite to report foreign entities using Infrastructure as a Service (IaaS) services from U.S. cloud hyper scalers to train dual-use models. These types of services allow customers access to advanced hardware, such as GPUs, for training AI models in the cloud. Though the order does not mention China specifically, it is clear that administration officials are concerned about the potential for Chinese companies to access export-controlled hardware such as advanced GPUs via the cloud.

The provisions laid out in the EO for AI developers do not come as a major surprise given recent conversations and growing concerns in Washington. The White House's Voluntary Commitments also asked companies to do the same. However, the new KYC requirements may pose challenges for some cloud services providers**,** as it may be difficult to judge when a foreign person or entity uses an IaaS provider "to train a large AI model with potential capabilities that could be used in malicious cyber-enabled activity."

Overall, the EO's approach to model developers is balanced and pragmatic. The order does not require companies to obtain licenses for AI models or to disclose their training methods, proposals that have once been considered but have been met with intense opposition from industry. Still, the EO is likely to kick off a more intense debate in Washington and industry about whether a notification process is sufficient and how it can be implemented in ways that do not put an undue burden on companies. In addition, the question of whether such an approach could or should be multilateralized is also up for debate.

This is a complex issue, as U.S. firms dominate the development of large language models, along with Chinese companies, and the question of which countries the Department of Commerce would share information about the training of advanced models is likely to become a major issue. In addition, as part of a new agreement coming out of the U.K. AI Summit, several large U.S. AI companies have agreed to let the U.S., U.K., and Singapore governments test their models, a process that would require multilateral sharing of information about advanced models.

Order expands requirements on U.S. government agencies to assess the use of AI and develop standards, but staffing and funding remain a question

Finally, the EO also tasks government agencies to consider their own use of AI systems; to comply with AI ethics frameworks and to leverage their authority to prevent the misuse of AI systems in areas like education, hiring, and housing. Privacy features prominently as a core principle in the EO, with the administration encouraging agencies to set guidelines on how they collect, use, and share personal data, as well as to support and implement measures to strengthen privacy-preserving research and technologies.

Similarly, it directs agencies to complement AI safety efforts by introducing new testing, standards, and other tools. To that end, it directs the National Institute of Standards and Safety (NIST) to develop red-teaming standards and a companion resource for the AI risk management framework for generative AI and the Department of Commerce to issue guidance for watermarking AI-generated content. It remains unclear, however, the extent to which these tools and standards will apply to the AI developers mentioned above. For example, it is not clear

whether companies falling in the scope of the order will be required to adhere to NIST's red-teaming standards, or whether these standards are voluntary, and intended more for use in the federal government.

Either way, the reliance on federal agencies and government institutions to create new standards and risk management tools is in keeping with the U.S.'s overall approach to AI governance, which since the Obama administration has relied on the executive branch and independent agencies to do the heavy lifting on AI regulation. Under the new mandate, however, federal agencies will likely have to staff new positions, including Chief AI Officers. This process is already underway in some agencies, but not in others. While AI leadership positions are emphasized, many leading Departments and Agencies are reinvigorating their tech/digital/data teams and approaches to procurement, hiring, and other fundamental aspects of modernization as a result, as well.

Key questions remain about the ability of different agencies to fulfill their mandates laid out in the order, at least in the near term. The lack of AI talent remains a significant challenge across all levels of government. To some extent, the order addresses these challenges by calling on the heads of agencies to identify priority mission areas for hiring AI talent and accelerate hiring pathways to ensure adequate implementation of the order. It further recommends loosening immigration restrictions and streamlining the visa process for international workers with specializations in AI. But these efforts are nascent, and it remains unclear whether individual agencies have the capacity and political will to staff up new positions in a short period of time or whether shifts in immigration policy are realistic ahead of the next presidential election.

Delivering on some of the EO's directives will also depend on a number of other factors, including the availability of funds, the ability of the White House to continue driving the process within the executive branch as the campaign season heats up, as well as its ability to work with Congress to put elements of the EO on a sound and enduring legal basis. Furthermore, the administration remains in ongoing discussions about how best to institutionalize a governance function for advanced AI models within the executive branch–-either via empowering existing agencies or forming a new organization--that would ensure the current effort extends beyond next year's election. There is general agreement that the current White House-led effort should be put on a more sustainable footing, but much disagreement about exactly how to operationalize this approach.

Multilateral cooperation key element of order

The timing of the order's release before the U.K. AI Summit this week at Bletchley Park was not a coincidence. The administration is eager to keep pace with multilateral efforts underway in the G7 and specific countries such as Canada, which have a significant commercial presence in the AI sector. U.K. Prime Minister Rishi Sunak is positioning the U.K. to play a convening role in bringing together key players, including China, to develop a global framework around AI governance. Other efforts are being headed by a new AI Advisory Body at the United Nations. There is now general agreement on basic principles and a focus on frontier AI, as evidenced by the communique coming out of Day 1 of the U.K. Summit, which was endorsed by 28 countries, including China and the EU. The final communique contained less detail than an initial draft communique circulated ahead of the summit, suggesting some of the text – for example which referred to the OECD and UN, was generalized to gain consensus.

The somewhat thornier issue of codes of conduct for companies will be critical to see progress on in the coming months, with the G7 last week releasing a new code of conduct that is largely in line with the White House Voluntary Commitments. In addition, associated with the order is the establishment of the AI Safety Institute (USAISI) under NIST, and a related Consortium, to work with industry on joint research and development. The USAISI will also work with the U.K. AI Safety Institute to ensure both countries are on the same page on the governance of frontier AI models going forward.

The goal is to create resources within government where current capabilities are lacking, as most testing happens in-house within companies. U.S. officials are particularly eager to add value in the testing process by focusing on the national security related concerns about advanced model development and deployment. Secretary Raimondo, in her plenary address at the U.K. AI summit, called out cooperation with the U.K. AI Safety Institute as a priority. In addition, a number of leading AI companies released their internal safety policies prior to the Summit, but Chinese companies did not, and this will be something to watch going forward—whether Beijing sees this as a best practice and allows its firms to release safety policies.

A key issue will remain the status of China, whose participation in the U.K. Summit was opposed by many in the U.S. and EU. Opponents of involving Beijing would have preferred for "like-minded" democracies to establish a framework first, before engaging with China on the issue. For its part, Beijing is eager to engage in dialogue, not wanting to be left out of global deliberations on setting some guardrails around generative and frontier AI. At the same time, Chinese officials are highly critical of U.S. efforts to restrict Chinese company access to critical hardware, advanced GPUs, that are widely used to train large language models. Last month, Beijing released its own blueprint, called the Global AI Governance Initiative, which includes some language that is compatible with Western documents issued over the past year, but also contains objections to efforts to control technology, and positions Beijing as an advocate for the Global South on AI governance.

Looking ahead: next steps

Along with directing agencies to take specific steps including establishing new roles and standards for red-teaming, KYC, and reporting obligations for companies working with some powerful AI models, the order called for a number of studies within 30, 60, 90, 120, 150, 180, and 270 days. There will be a steady cadence of such reports released over the next year, which companies should watch for. The White House will also be collecting comments on the reporting requirements outlined in the document and socializing the more controversial elements with key players.

The White House and Department of Commerce will also be working to align the process of implementing elements of the EO with plurilateral and multilateral efforts such as the G7 Hiroshima process and the U.K. AI Safety Summit process. Already U.K. AI Safety Summit organizers have announced their next meeting in South Korea six months from now, followed by another round of engagements in France in a year. The Summit also saw the announcement of a new expert group to oversee a global research effort headed by leading AI/ML scientist Yoshua Bengio, which will include researchers from China, and will generate a report on the state of the technology before the next summit in South Korea.

The U.S. and China are also likely to establish an AI working group following APEC, focused primarily on the use of AI in weapons systems. The working group could also provide a platform for discussions around broader AI governance issues, including principles coming out of the U.K. Summit, and various voluntary commitments and codes of conduct. Beijing is likely to eventually support its companies signing on to a code of conduct, as long as it is not clearly U.S. or G7 branded. In this regard, U.S. government export controls on GPUs are working strongly against any attempts to get Beijing to consider allowing Chinese companies to fully participate in any initiatives involving the U.S. or other allied governments. This threatens to undermine Prime Minister Sunak's efforts to include the second most important country in what U.K. government officials are touting as "truly global consensus." The run-up to the South Korea summit will be critical to determining both the future direction of China's participation in global AI governance initiatives and U.S. government efforts, embodied in the EO, towards building a regulatory framework around frontier AI models will forward.

Finally, U.S. officials have characterized the EO as just the beginning of a long process to build out U.S. government capabilities in the AI governance space. Many areas remained unaddressed by the executive order and will be a major part of the debate around AI governance going forward. For example, the open sourcing of advanced AI models is a major concern for governments, an issue that the EO does not address. The open-source issue is, however, tied to concerns about the potential for malicious actors to leverage advanced models in areas that could impact national security, so further executive and eventually legislative actions will likely attempt to tackle this issue in the future. For now, the administration has laid down a significant marker, stressing that AI governance is a critical priority, and the U.S. government will mobilize efforts in the coming month to ensure that AI develops in a safe and secure manner.

About ASG

Albright Stonebridge Group (ASG), part of Dentons Global Advisors, is the premier global strategy and commercial diplomacy firm. We help clients understand and successfully navigate the intersection of public, private, and social sectors in international markets. ASG's worldwide team has served clients in more than 120 countries.

ASG's Tech Policy Practice has extensive experience helping clients navigate markets gloablly. For questions or to arrange a follow-up conversation please contact Paul Triolo or Anarkalee Perera.