Better Safe Than Sorry: OMB Releases Memorandum on Managing AI Risks in the Federal Government

Article By:

Townsend L. Bourne

Daniel J. Alvarado

On March 28, 2024, the Office of Management and Budget ("OMB") issued Memorandum M-24-10, *Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence* (the "Memo"). This is the final version of a draft memorandum OMB released for public comment on November 1, 2023. The Memo primarily focuses on agency use of AI and outlines minimum practices for managing risks associated with the use of AI in the federal government. The Memo also provides recommendations for managing AI risks in federal procurement of AI that industry should keep in mind, specifically entities developing AI tools to sell to the federal government.

Responsible Development, Testing, and Operation of AI

The federal government plans to leverage the positive aspects of AI, but acknowledges the development, deployment, and use of AI must be done responsibly. The burden of managing AI risks falls on federal agencies and, specifically, their Chief AI Officer, which is a newly created position that each agency must designate within 60 days of the issuance of the Memo. Each individual agency Chief AI Officer is charged with coordinating agency use of AI and promoting AI innovation, while also managing the risks associated with the use of AI.

To advance responsible AI innovation, the Memo charges agencies with releasing publicly on their websites a strategy for removing barriers to the responsible use of AI. This strategy must include (a) a current assessment of the agency's AI maturity and AI maturity goals, (b) the agency's plans to effectively govern its use of AI, (c) a plan for establishing organizational and governance processes as well as developing the necessary infrastructure to manage risks, and (d) plans for future AI investment, among other items. Although collaboration and sharing between agencies of AI code, models, and data is emphasized in the Memo, this is subject to applicable law, contractual obligations, and national security risks.

The Memo also outlines minimum practices applicable to new and existing AI that is developed, used, or procured by or on behalf of agencies. These minimum practices are similar to the authorization and continuous monitoring frameworks under the Federal Risk and Authorization

Management Program (FedRAMP). The following minimum practices are applicable to both “Safety-Impacting AI”[1] and “Rights-Impacting AI”[2]:

- **Complete an AI impact assessment:** The assessment must document the intended purposed for the AI and its expected benefit, the potential risks of using AI and any mitigation efforts, and the quality and appropriateness of the relevant data used in designing, developing, training, testing, and operating the AI. Agencies will require vendors to provide sufficient descriptive information for the assessment.
- **Test the AI for performance in a real-world context**: Agencies are required to conduct adequate testing of the AI to ensure it will work in its intended real-world context.
- **Independently evaluate the AI**: Agencies must independently review the AI documentation to ensure the AI is working appropriately and as intended, and that the expected benefits outweigh potential risks. Once the independent evaluation is complete, the agency must incorporate the evaluation into an applicable release or oversight process, such as the Authorization to Operate process.
- **Conduct ongoing monitoring**: Once the AI has been deployed, agencies must institute ongoing procedures for monitoring the AI, including degradation of functionality and impacts on rights and safety. Agencies also must monitor and defend the AI from AI-specific exploits.
- **Regularly evaluate risks from the use of AI**: As part of the ongoing monitoring process, agencies must conduct periodic human reviews at least on an annual basis, or after significant modifications to the AI.
- **Mitigate emerging risks to rights and safety**: If an agency identifies new or significantly altered risks to rights and safety during the continuous monitoring processes, the agency must take steps to mitigate those risks. If the risks exceed the acceptable level for the particular AI and the mitigation steps do not sufficiently reduce risk, the agency must stop using the AI as soon as practicable.
- **Ensure adequate human training and assessment**: Agencies must ensure the human operators are adequately and sufficiently trained on a periodic basis as determined by the agency.
- **Provide additional human oversight, intervention, and accountability as part of decisions or actions that could result in a significant impact on rights or safety:** If the AI has Rights-Impacting and/or Safety-Impacting uses, agencies must identify decisions or actions that the AI is not permitted to make or take without human oversight, intervention, and accountability.
- **Provide public notice and plain-language documentation**: Agencies must include their AI use cases in the public use case inventory and provide accessible documentation in plain language to serve as public notice of the AI to its users and the general public.

There are two exclusions from the minimum practices: when using AI solely to (1) evaluate a potential vendor, commercial capability, or freely available AI capability that is not otherwise used in agency operations, exclusively for the purpose of making a procurement or acquisition decision, or (2) achieve an agency’s conformity with the requirements of managing risks from the use of AI. The Memo also includes additional minimum practices applicable to Rights-Impacting AI focused on equity, fairness, and mitigating algorithmic discrimination when it is present.

Managing Risks in Federal Procurement of AI

OMB also provides recommendations for agencies to responsibly procure AI and supplement their required risk management practices for Rights-Impacting AI and Safety-Impacting AI. Consistent with the theme of the Memo, the recommendations focus on the methods to ensure the responsible procurement of AI as well as promoting competition and ensuring the Government retains sufficient rights to data used in the design, development, testing, and operation of the AI. The following list summarizes the recommendations for managing risks in federal procurement of AI:

- **Aligning with the Law**: Procurement of AI should be consistent with applicable laws, regulations, and policies, with a particular emphasis on those addressing privacy, confidentiality, IP, cybersecurity, human and civil rights, and civil liberties.
- **Transparency and Performance Improvement**: The Memo emphasizes agencies should seek to obtain adequate documentation regarding the procured AI's capabilities, known limitations, and the provenance of the data used to train and operate the AI. The Memo also emphasizes the importance of continuous monitoring activities post-award to mitigate risk and incentivize continuous improvement of the procured AI.
- **Promoting Competition in Procurement of AI**: The Memo emphasizes the need to promote interoperability of procured AI to ensure agencies do not improperly entrench incumbents or permit vendors to favor their own products.
- **Maximizing the Value of Data for AI**: Agencies are encouraged to ensure their contracts retain rights to data and any improvements to the data to ensure the agency's continued design, development, testing, and operation of AI. Agencies also should consider including contract provisions regarding the protection of Federal information used by vendors in the development and operation of AI products and services for the government, including protection from unauthorized disclosure and prohibiting vendors from subsequently using the data to train or improve the functionality of the vendor's commercial offerings without express permission from the agency.
- **Overfitting to Known Test Data**: Agencies should take appropriate measures to ensure AI developers or vendors are not directly relying on test data to train their AI systems because this could result in the system appearing more accurate in tests than it would be in real-world applications.
- **Responsible Procurement of AI for Biometric Identification**: The Memo encourages agencies to take special care when procuring AI for biometric identification. This includes assessing and addressing risks that the data used may not be lawfully collected or used without appropriate consent, embeds unwanted bias, or was collected without validation of the included identities. Agencies should request supporting documentation from vendors and contractors related to test results to validate the accuracy of the AI.
- **Responsibly Procuring Generative AI**: When procuring generative AI tools, particularly dual-use foundational models, the Memo encourages agencies to require adequate testing and safeguards, require AI red-team testing against risks during testing and evaluation, and require appropriate labeling of content generated or modified by AI. Agencies also are encouraged to incorporate relevant National Institute of Standards and Technology ("NIST") standards such as the AI Risk Management Framework (and the forthcoming AI Risk Management Framework for Generative AI).
- **Assessing for Environmental Efficiency and Sustainability**: Agencies are encouraged to consider the environmental impact of computationally intensive AI services (e.g., those reliant on dual-use foundation models), including methods to improve efficiency and sustainability of AI.

OMB issued a Request for Information on the responsible procurement of AI and will incorporate information received in response to the RFI to develop procurement practices for AI systems and services consistent with the recommendations provided in the Memo. It will be important for contractors in this space to stay up-to-date on government guidance and practices for managing the risks associated with the use of AI in order to be able to provide solutions that meet government requirements and ensure the safe and responsible design, development, and use of AI.

FOOTNOTES

[1] This term refers to “AI whose output produces an action or serves as a principal basis for a decision that has the potential to significantly impact the safety of: (1) Human life or well-being, including loss of life, serious injury, bodily harm, biological or chemical harms, occupational hazards, harassment or abuse, or mental health, including both individual and community aspects of these harms; (2) Climate or environment, including irreversible or significant environmental damage; (3) Critical infrastructure, including the critical infrastructure sectors defined in Presidential Policy Directive 21 or any successor directive and the infrastructure for voting and protecting the integrity of elections; or, Strategic assets or resources, including high-value property and information marked as sensitive or classified by the Federal Government.

[2] This term refers to “AI whose output serves as a principal basis for a decision or action concerning a specific individual or entity that has a legal, material, binding, or similarly significant effect on that individual’s or entity’s: 1. Civil rights, civil liberties, or privacy, including but not limited to freedom of speech, voting, human autonomy, and protections from discrimination, excessive punishment, and unlawful surveillance; 2. Equal opportunities, including equitable access to education, housing, insurance, credit, employment, and other programs where civil rights and equal opportunity protections apply; or 3. Access to or the ability to apply for critical government resources or services, including healthcare, financial services, public housing, social services, transportation, and essential goods and services.

Listen to this post

National Law Review, Volume XIV, Number 120

Source URL:https://natlawreview.com/article/better-safe-sorry-omb-releases-memorandum-managing-ai-risks-federal-government