

Hugging Face, Inc. 20 Jay Street, Suite 620, Brooklyn, NY 11201

Hugging Face Response to the Office of Management and Budget Request for Comments on Agency Use of Artificial Intelligence

Hugging Face commends the Office of Management and Business on its extensive work framing the responsible use of AI by federal agencies. The following comments are informed by our experiences as an open platform for state-of-the-art (SotA) AI systems, working to make AI accessible and broadly available to researchers for responsible development.

About Hugging Face

Hugging Face is a community-oriented company based in the U.S. and France working to democratize good Machine Learning (ML), and has become the most widely used platform for sharing and collaborating on ML systems. We are an open-source and open-science platform hosting machine learning models and datasets within an infrastructure that supports easily processing and analyzing them; conducting novel AI research; and providing educational resources, courses, and tooling to lower the barrier for all backgrounds to contribute to AI.

Summary

The draft memorandum outlines a strong framework for fostering responsible Al adoption by federal agencies. In a context of <u>commonly inflated claims</u> about the performance of Al systems and pervasive risks of exacerbating discrimination, we particularly appreciate the proposed requirements to not only conduct thorough impact assessments, but also to demonstrate that the adoption of Al does provide added value to the agencies.

In order to best support this mission, we recommend that the OMB and the proposed Al Governance bodies prioritize mechanisms that support the open sharing of skills, resources, and information; build internal capacity along as much of the Al development chain as possible, from pre-training data curation to model fine-tuning and deployment (including, as currently outlined in Section 4.b.iv, non-technical expertise); and develop further guidance on what constitutes a sufficient improvement to warrant Al adoption and how to choose an Al system to meet those goals.

For the purpose of sharing best practices, we note that the proposed Al Use Case inventory constitutes a particularly promising tool. We recommend that entries in the inventory include as much information as possible about the Al application and system, including data and model documentation according to established standards such as dataset sheets and model cards, the metrics used to assess whether Al adoption would have a positive impact for the agency's use case, the reasoning for determining whether the system was safety- or rights-impacting, and links to the specific versioned systems used, including any open-source or open-access components.



We also recommend that the OMB extend the scope of memorandum OMB M-16-21 on open source software to include Al systems. In particular, we believe that the testing datasets developed by agencies to meet the requirements laid out in the present draft memorandum have a significant role to play in advancing the state of the art in performance and risk evaluation of Al systems if they are shared broadly with researchers and stakeholders. By supporting agencies in sharing this data to all relevant stakeholders, including by providing guidelines for the use of synthetic data, data pseudonymization, and governance of public datasets, the OMB can accelerate the development of robust responsible Al practices that will benefit all of society.

Beyond evaluation datasets, we encourage the OMB to direct agencies to build capacity to develop and share their own full AI stacks, including by leveraging and adapting existing public, open-source, and open Machine Learning models and datasets. This approach can lead to more energy-efficient AI applications, more transparency into the regulatory compliance of the AI development chain, and better visibility into the handling of e.g. privacy and intellectual property rights along the development process (see for example the StarCoder open code generation
LLM). To that end, and in addition to the evaluation and test, we recommend that the AI strategy of federal agencies include to the extent possible the development and sharing of:

- pre-training datasets for base models (or foundation models) focused on their domain of activity
- the corresponding pre-trained (or foundation) models
- fine-tuning datasets and fine-tuned models tailored to the agencies' needs
- any software needed to process the inputs and outputs of AI systems

We provide further feedback on the draft memorandum below by answering the questions accompanying the Request for Comments. If a question is not highlighted, we do not have specific, actionable feedback.

Questions

2. What types of coordination mechanisms, either in the public or private sector, would be particularly effective for agencies to model in their establishment of an Al Governance Body? What are the benefits or drawbacks to having agencies establishing a new body to perform Al governance versus updating the scope of an existing group (for example, agency bodies focused on privacy, IT, or data)?

While AI Governance will require involvement from many existing groups in agencies, including but not limited to privacy, IT, data, and legal departments, approaching the new challenges posed by AI adoption holistically and efficiently will require coordination and skill sharing between different stakeholders, which is best managed by a new AI Governance Body. Such a body can also facilitate broader civil society involvement and standardization of documentation practices. As examples of the possible functions of such a body, we point to the role of the cross-areas working groups within the BigScience workshop on Large Language Models, and its proposed structure for a multi-stakeholder AI data governance organization for large scale ML. In particular, we stress the importance of the role of centralizing documentation and transparency guidelines and the adoption of interoperable data formats to allow for better skill sharing within and across agencies.

3. How can OMB best advance responsible Al innovation?

The OMB has a unique role to play in championing responsible Al innovation beyond the federal government by helping create a body of positive examples and in-context operationalizations of responsible Al principles. Other agencies and companies may then learn from those examples



and leverage the resources created in the process – especially if they are shared as open and open-source AI systems – to follow suit. These should include:

- Requiring documentation of key sections of the development pipeline demonstrating decisions that respect peoples' rights and abide by applicable laws. This included:
 - Data Development
 - Model Training
 - Model Evaluation
 - System Evaluation (when models are implemented as part of a larger system)
 - Actions for continued Monitoring
- Building resources and knowledge about how to assess, evaluate, document, and audit Al systems
 - In particular, focus on building capacity to audit specific parts of Al development, including training data and procedure
 - Build a robust reporting system including labor complaints and whistleblower protections
- Leveraging shared experience between agencies to build best practices informed by the practicalities of deployment
- Highlight that AI should in many cases *NOT* be used, and/or may not be an
 improvement over current/alternative systems, in particular by providing transparency
 through the AI Use Case inventory about the process for weighing the risks and benefits
 of AI adoption.
- 4. With adequate safeguards in place, how should agencies take advantage of generative AI to improve agency missions or business operations?

Generative AI, when used for the purposes it is best suited for, can be a useful tool to explore AI in a new application setting, increase productivity in settings such as software engineering, and support applications that are less reliant on accuracy or grounding in factual information – especially when it is developed with <u>sufficient transparency</u>, <u>consideration for the rights of data subject</u>, and <u>attribution</u> of the model outputs. In particular, generative AI systems that are designed to handle a variety of tasks or work in so-called zero-shot settings can significantly accelerate the development of a "proof-of-concept" AI system to explore whether AI can bring value to an agency in a particular setting; possibly to then be <u>replaced by a more efficient discriminative system</u>.

Generative AI also introduces unique concerns, including <u>increased financial and environmental</u> costs when it is used in settings where a discriminative AI alternative could suffice, an increased cybersecurity attack surface and <u>range of vulnerabilities</u>, and a dependence on web-scale and larger training datasets that exacerbates social issues, including biases and discrimination.

The trade-offs mentioned also depend on the modality considered. Code generation models such as the openly developed <u>StarCoder</u> (see the project's <u>Governance Card</u> for more detail), have already shown their value in accelerating the process of software development, and are less subject to (but not exempt from) <u>exacerbating social biases along the development chain</u>. Image generation models, on the other hand, are still very likely to <u>reproduce and exaggerate stereotypes</u>, which makes them a poor fit for generating illustrations of people in government communications.

5. Are there use cases for presumed safety-impacting and rights-impacting AI (Section 5 (b)) that should be included, removed, or revised? If so, why?



We note that neither privacy nor handling of medical processes are currently included in the safety section. Mishandling of medically relevant decisions by AI systems can lead to immediate personal risks – especially given the prevalence of social biases in medical datasets, and leaks of private information can expose citizens to scamming, phishing, and identity theft. We would recommend explicitly adding both those considerations in (Section 5 (b.i)). The latter should also extend to systems that are used to infer or predict private information or personal characteristics.

For (Section 5 (b.ii)), we also recommend considering at least some of the rights listed in the recently published <u>Taxonomy of Human Rights Risks Connected to Generative Al</u> published by the UN OHCHR, including but not limited to rights to employment, freedom from Physical and Psychological Harm, and rights to Culture, Art and Science. <u>This recent editorial</u> also delves deeper into the rights at play and outlines how different steps in the development process have bearing on specific rights.

6. Do the minimum practices identified for safety-impacting and rights-impacting AI set an appropriate baseline that is applicable across all agencies and all such uses of AI? How can the minimum practices be improved, recognizing that agencies will need to apply context-specific risk mitigations in addition to what is listed?

The minimum practices identified constitute a strong baseline on which to develop responsible Al practices across a variety of potentially safety-impacting and rights-impacting Al applications. We strongly believe that these minimum practices should be encouraged for **all Al systems** even when not strictly required, to help develop robust best practices and mitigate risks that some systems may be mistakenly omitted.

We also recommend taking the following additions:

- For the sections on Completing an Al impact assessment and ensuring that the Al
 will advance equity, dignity, and fairness, we recommend explicitly mentioning due
 diligence background research that considers historical discrimination in the domain. Al
 applications tend to exacerbate existing problems.
- In addition for the existing requirements to conduct monitoring, mitigate emerging risks, and prioritize incremental adoption, agencies should, whenever possible, have a plan in place for rolling back AI systems as needed without disrupting ongoing processes,
- Finally, we stress the value of documentation in pursuit of the goals outlined in this
 memorandum. We recommend that data documentation and model cards be required
 for Al systems with safety-impacting and rights-impacting potential, along with system
 cards and governance cards as needed.
- 7. What types of materials or resources would be most valuable to help agencies, as appropriate, incorporate the requirements and recommendations of this memorandum into relevant contracts?

The following categories of guidance will be particularly helpful to agencies:

- Specific questions to answer to inform the processes outlined in this document, especially with respect to pro-actively assessing the benefits of AI adoption and risks to rights, fairness and non-discrimination, and safety and how to weigh those to make an adoption decision
- Standards on documentation content and quality, especially as regards information to be included in the AI Use Case inventory
- Providing a repository of positive examples of implementation of the requirements to help showcase successful operationalization



We note in particular that given the fast evolution of the technology and breadth of possible adoption, this guidance should be designed to evolve with time and feedback from federal agencies. We recommend in particular dedicating appropriate resources to maintaining and updating this information.

8. What kind of information should be made public about agencies' use of Al in their annual use case inventory?

We recommend the following information for entries in the AI Use Case inventory:

- Brief description of the Al system, including its deployment context, expected inputs and outputs, and expected behaviors, and appropriate context of use
- Description of the metrics used to assess what value the AI system is bringing to the agency's use case
- Whether the system is deemed to be rights-impacting or safety-impacting, with justification
- Description of the metrics tracked to assess performance and potential rights and safety risks, especially with regard to non-discrimination
- The expected volume of interactions with the system, *e.g.* number of inputs processed per time period or number of users within the agency
- A list of the AI components used in the AI system, including:
 - ML models, including base pre-trained model and fine-tuned models, with version when applicable
 - ML training datasets, including pre-training and fine-tuning datasets, with version when applicable
- Model and data documentation for all components according to established standards
 - For models, this should include appropriate contexts of use and evaluation in light of both those uses and different types of users and affected groups
 - Example model cards to consider: BLOOM, OBELICS, StarCoder
 - For data, this should include handling of consent, licensing, analysis of possible source of biases, etc.
 - see also: https://hf.co/blog/yjernite/data-transparency
- Description of the governance processes relevant to the Al system, for example:
 - o <u>Data governance in BigScience</u> (BLOOM model)
 - o Governance card for BigCode (StarCoder model)

All items outlined above have a significant role to play in ensuring the proper governance of Al systems. By centralizing this information across uses of Al in federal agencies, the Al Use Case inventory will both be a useful tool for supporting this governance and an invaluable resource to support new research into best practices for Al adoption across society.