Request for Information:
Artificial Intelligence Risk Management Framework
86 Fed. Reg 40810 (July 29, 2021)
Docket # 210726-0151

August 30, 2021

Overview

Google welcomes the opportunity to provide comments in response to the National Institute of Standards and Technology's (NIST) Request for Information (RFI) on the Artificial Intelligence Risk Management Framework (AI RMF or Framework).

We have long-championed AI technology. Our company is at the forefront of AI development, and we have seen firsthand how AI can enable massive increases in performance and functionality. AI has the potential to deliver great benefits for economies and society —from improving energy efficiency and more accurately detecting disease, to increasing the productivity of businesses of all sizes. Harnessed appropriately, AI can also support more fair, safe, inclusive, and informed decision-making.

Google is optimistic about the incredible potential for AI and other advanced technologies to empower people, widely benefit current and future generations, and work for the common good. With that said, we recognize that these innovative technologies also raise important questions and challenges that will need to be addressed clearly, thoughtfully, and affirmatively for the AI ecosystem to thrive.

As one of the leaders in the field, we acknowledge that Google has an obligation to develop and apply AI thoughtfully and responsibly, and to support others to do the same. This is part of the reason that in 2018 we published our own AI Principles to help guide our ethical development and use of AI.[1]

We believe that self- and co-regulatory approaches remain the most effective and practical way to prevent and address a number of AI-related problems within the

[1] https://ai.google/principles/

boundaries already set by sector-specific regulation. By relying on expertise from a wide variety of industry and civil society perspectives, these frameworks can remain flexible and nimble in a way that static regulation cannot, evolving over time as the technologies innovate and change.

Google supports NIST's approach and goals for developing the AI RMF and agrees with the principles and attributes identified in the RFI. Our responses to each of NIST's specific requests for information are provided below.

Response to RFI

1. ***The greatest challenges in improving how AI actors manage AI-related risks—where "manage" means identify, assess, prioritize, respond to, or communicate those risks.***

Our understanding of AI technology— as well as its benefits, potential risks, and available mitigation options—is constantly evolving. Given the immense range of AI applications across a diverse set of sectors, the risks and impacts of AI technology can also vary significantly by application. There are few widely accepted metrics or benchmarks for measuring and comparing the benefits and risks of AI systems, and even when risks can be identified and measured, they can rarely be completely eliminated.

One of the biggest challenges in improving how AI actors manage AI-related risks is finding the right balance between putting in place established guidelines and responsible practices that govern the development and use of the AI technology, while still allowing the significant flexibility necessary to adapt to evolving scenarios and generate creative solutions.

The Framework must consider the social and economic context in which AI systems are deployed and focus on risks that can be effectively estimated and mitigated. Notably, it may be that some AI applications considered "high-risk" are also "high value" to society. For instance, AI has tremendous potential to advance health care, including new tools to identify, prevent, and treat serious disease, and it must be held to extremely high standards of safety, reliability, and fairness.

It is also important to acknowledge the opportunity costs of *not using* AI in a specific situation, or of intentionally developing AI without particular capabilities. The risks and

benefits of AI systems should be weighed against existing (non AI) approaches, including human judgement. If an imperfect AI system is shown to perform better than the *status quo* at a crucial life-saving task, for example, it may be irresponsible to not use the AI system. Where the alternative of not using AI poses greater risk than the risk posed by deploying an AI system, AI actors should continue to be supported for AI's net beneficial use.

Together, industry, academia, and civil society, along with others, will play a critical role in providing balanced, fact-based analyses of the opportunities and challenges presented by AI, reflecting views across diverse disciplines, perspectives, and walks of life. Although there remain plenty of questions and challenges related to AI and managing risk, given its remarkable promise across society, our biggest risk would be not encouraging the responsible use of AI to help us address some of the world's greatest challenges.

2. ***How organizations currently define and manage characteristics of AI trustworthiness and whether there are important characteristics which should be considered in the Framework besides: Accuracy, explainability and interpretability, reliability, privacy, robustness, safety, security (resilience), and mitigation of harmful bias, or harmful outcomes from misuse of the AI.***

Google agrees that public trust is best achieved if AI technology is developed responsibly and transparently, and supports NIST's aim of cultivating public trust in AI throughout the design, development, use, and evaluation lifecycle. Our AI Principles (discussed further in the response to request 3 below) outline seven characteristics that are instilled in allour AI projects, as well as four applications that we will never pursue.[2] We have also developed tools, techniques, and infrastructure to enable AI developers inside and outside Google to implement our Principles.[3]

Trust in AI systems only partially reflects the properties of the systems themselves. Brand trust, media coverage, and fears related to job disruptions also play a significant role in establishing public confidence and trust in AI.

The potential benefits of AI technology cannot be fully realized if its development is held back by unfounded fears and misunderstandings. Google therefore encourages

[2] https://ai.google/principles/
[3] https://ai.google/responsibilities/responsible-ai-practices/

the building of a Framework that will help create trust and guide responsible development and use of this widely applicable technology.

3. ***How organizations currently define and manage principles of AI trustworthiness and whether there are important principles which should be considered in the Framework besides: Transparency, fairness, and accountability.***

Responsible development of AI presents new challenges and critical questions for us all. In 2018, we published our own AI Principles to help guide our ethical development and use of AI, and we also established internal review processes to help us mitigate unfair bias, test rigorously for safety, and design with privacy top of mind.[4] Our principles also specify areas where we will not design or deploy AI, such as to support mass surveillance or violate human rights.[5]

Specifically, in addition to principles of transparency, fairness, and accountability, Google's AI principles include:

- Social benefit: The use and development of AI technology should be pursued where the overall benefits of the technology (social and economic) substantially outweigh any foreseeable risks and drawbacks.
- Safety and security: Strong safety and security practices should be incorporated into AI development and use, and AI actors should seek to avoid unintended results or misuse or abuse that creates risks of harm.
- Privacy: AI actors should incorporate privacy principles into the development and use of AI technology. Such privacy principles include opportunity for notice and consent, architectures with privacy safeguards, and providing appropriate transparency and control over the use of data.
- Scientific excellence: The highest standards of scientific excellence should be used as AI technology is researched and developed. Scientifically rigorous and multidisciplinary approaches are encouraged for AI research. Google also upholds this standard by publishing educational materials, best practices, and research so others can develop useful, thoughtful, and responsible AI applications.[6]

[4] https://ai.google/principles/
[5] https://ai.google/responsibilities/responsible-ai-practices/
[6] https://ai.google/responsibilities/responsible-ai-practices/

- Access and availability: AI technology has transformative potential for the common good. To that end, access to AI technology and research should be democratized, rather than reserved for those with the most resources. Regulators can encourage public access by providing government data sets (scrubbed and anonymized) to help support the design, development, and operation of AI applications. We support efforts by Governments to increase access by funding AI research and educational efforts. To this end, Google is doing its part to safely share open data[7] and provide funding to those who seek to create uses of AI for social good.[8]

We also recognize that NIST has existing frameworks and standards that address some of these themes, for example NIST's privacy[9] and cybersecurity[10] frameworks. Rather than reinventing the wheel or risking creating redundant or conflicting approaches, it could instead be beneficial to reference existing standards and frameworks in the AI RMF and articulate how these frameworks can be used together to holistically manage risk.

4. ***The extent to which AI risks are incorporated into different organizations' overarching enterprise risk management—including, but not limited to, the management of risks related to cybersecurity, privacy, and safety.***

AI risks do not exist in a vacuum. With that said, integrating AI into existing risk management processes is typically a complex, iterative process of experimentation, research, model training and retraining, testing and validation, and redevelopment.

At Google, our dedicated AI Principles review processes complement our existing internal governance processes, including privacy, security, and quality assurance.[11] Our approach involves combining embedded processes from across our product areas, and operationalizing our AI Principles is challenging work. Specifically, we have a central, dedicated team that reviews proposals for AI research and applications for alignment with our principles. The review process is iterative, and we continue to refine and improve our assessments as advanced technologies emerge and evolve. The team also consults with internal experts in machine-learning, fairness, security, privacy,

[7] https://www.blog.google/technology/ai/sharing-open-data
[8] https://ai.google/static/documents/accelerating-social-good-with-artificial-intelligence.pdf
[9] https://www.nist.gov/privacy-framework/privacy-framework
[10] https://www.nist.gov/cyberframework
[11] https://ai.google/responsibilities/review-process/

human rights, and other areas. We believe that our cross-disciplinary approach brings a deep understanding of specific technologies, use cases, and user bases.

5. ***Standards, frameworks, models, methodologies, tools, guidelines and best practices, and principles to identify, assess, prioritize, mitigate, or communicate AI risk and whether any currently meet the minimum attributes described above.***

We agree that common standards, frameworks, and benchmarks are needed to assess and compare different AI systems. However, this work is not starting from scratch. Several organizations are already working in this space, including the International Organization for Standardization (ISO),[12] the Organisation for Economic Co-operation and Development (OECD),[13] MLCommons,[14] the Partnership on AI (PAI),[15] and the Office of Management and Budget (OMB).[16]

Many of these organizations' efforts can be used to address the minimum attributes identified by NIST. For instance, the OECD provides definitions of AI terms and concepts and articulates AI principles of its own that may assist in providing common definitions.[17] In addition, ISO/IEC JTC 1/SC 42 develops international standards on AI;[18] MLCommons and PAI study and provide sets of best practices for AI technologies;[19] and OMB has released its own framework for regulatory and non-regulatory approaches to AI applications developed and used outside of the federal government.[20]

Moreover, and as discussed further in response to request 3 above, Google has also promulgated its own responsible AI practices,[21] including a variety of toolkits,[22]

---

[12] https://www.iso.org/committee/6794475.html
[13] https://www.oecd.ai/
[14] https://mlcommons.org/en/
[15] https://www.partnershiponai.org/
[16] https://www.oecd.ai/ai-principles
[17] https://www.oecd.ai/ai-principles; 86 Fed. Reg. 40811; *see also* https://www.partnershiponai.org/to-prevent-algorithmic-bias-legal-and-technical-definitions-around-algorithmic-fairness-must-align/
[18] https://www.iso.org/committee/6794475.html
[19] https://mlcommons.org/en/mlcube/; https://www.partnershiponai.org/research-lander/.
[20] https://www.whitehouse.gov/wp-content/uploads/2020/11/M-21-06.pdf
[21] https://ai.google/responsibilities/responsible-ai-practices/
[22] *See, e.g.*, https://ai.googleblog.com/2020/07/introducing-model-card-toolkit-for.html

frameworks,[23] and methodologies[24] to help address the goals and desired attributes identified by NIST in its proposed Framework.

Google supports NIST's intention to make its AI RMF voluntary and recommends that any standards are formed through a multi-stakeholder process, similar to the AI standards set by ISO/IEC JTC 1/SC 42 and NIST's past frameworks on privacy and cybersecurity. Continuously soliciting input from stakeholders will also ensure that the framework will have practical applications to AI products.

6. ***How current regulatory or regulatory reporting requirements (e.g., local, state, national, international) relate to the use of AI standards, frameworks, models, methodologies, tools, guidelines and best practices, and principles.***

AI standards, frameworks, and best practices can be complementary with current regulatory and reporting frameworks. They can also facilitate compliance with regulations, provide further clarity around expectations, and enable third-party oversight and comparison of AI systems.

Certification to international standards can also serve as a means to demonstrate compliance with regulatory requirements, reducing the burden of additional assessments on regulators and organizations, as outlined in the European Commission's draft Artificial Intelligence Act.[25]

Google welcomes a standardized and more cohesive approach to AI oversight. Common standards and frameworks can enable the interoperability of AI technologies and harmonization of AI governance approaches around the world, rather than a fragmented approach that could slow the pace of AI development and potentially limit the availability of new products and services to consumers in certain jurisdictions.

A self-regulatory or co-regulatory set of international governance norms based on voluntary standards that could be applied flexibly and adaptively would enable policy safeguards while preserving the space for continued beneficial innovation.

---

[23] *See, e.g.*, https://cloud.google.com/responsible-ai
[24] *See, e.g.*, https://developers.google.com/machine-learning/guides/rules-of-ml
[25] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206

7. ***AI risk management standards, frameworks, models, methodologies, tools, guidelines and best practices, principles, and practices which NIST should consider to ensure that the AI RMF aligns with and supports other efforts.***

The alignment of the AI RMF with other AI risk management standards is essential for ensuring a harmonized, interoperable regime that streamlines compliance for AI actors and minimizes confusion or contradiction between different standards and frameworks.

Google recommends that NIST consider the AI principles set by OECD, which have been adopted by OECD member countries, as well as Argentina, Brazil, Costa Rica, Malta, Peru, Romania and Ukraine.[26] Google also recommends that NIST ensure that its AI RMF aligns with standards for AI set by ISO/IEC JTC 1/SC 42,[27] and AI standards articulated in ISO/IEC JTC 1/SC 27 on information security, cybersecurity, and privacy protection.[28]

8. ***How organizations take into account benefits and issues related to inclusiveness in AI design, development, use and evaluation—and how AI design and development may be carried out in a way that reduces or manages the risk of potential negative impact on individuals, groups, and society.***

Addressing fairness and inclusion in AI is an active area of Google's AI work. From fostering an inclusive workforce that embodies critical and diverse knowledge[29] to assessing training datasets for potential sources of bias, training models to remove or correct problematic biases, evaluating machine learning models for disparities in performance, and continued testing of final systems for unfair outcomes, inclusivity must be considered at each stage of the AI lifecycle.

Far from a solved problem, fairness and inclusion in AI presents both an opportunity and a challenge. Google is committed to making progress in all of these areas, and to creating tools, datasets, and other resources for the larger community. We are an active contributor to this field, including in the provision of developer tools.[30] For example:

---

[26] https://www.oecd.org/going-digital/ai/principles/
[27] https://www.iso.org/committee/6794475/x/catalogue/p/0/u/1/w/0/d/0
[28] https://www.iso.org/committee/45306/x/catalogue/
[29] https://diversity.google/
[30] https://ai.google/responsibilities/responsible-ai-practices/?category=fairness

- *Facets:* interactive visualization tool that lets developers see a holistic picture of their training data at different granularities[31]
- *ML fairness gym:* a set of components for building simple simulations that explore the potential long-run impacts of ML systems[32]
- *What-If Tool (WIT):* An interactive tool that allows ML developers to explore how their models perform for different groups of users.[33]

9. ***The appropriateness of the attributes NIST has developed for the AI Risk Management Framework. (See above, "AI RMF Development and Attributes")***

Google agrees that the attributes NIST identifies in the RFI are appropriate and beneficial to the overall development of the AI RMF.

10. ***Effective ways to structure the Framework to achieve the desired goals, including, but not limited to, integrating AI risk management processes with organizational processes for developing products and services for better outcomes in terms of trustworthiness and management of AI risks.***

NIST's cybersecurity and privacy frameworks[34] have been well-received by a variety of stakeholders and are suitable models after which NIST could structure its AI RMF. Modeling the AI RMF after its past two frameworks would also provide NIST with an opportunity to articulate how the three frameworks interact in a cohesive scheme.

11. ***How the Framework could be developed to advance the recruitment, hiring, development, and retention of a knowledgeable and skilled workforce necessary to perform AI-related functions within organizations.***

There is an emerging consensus that AI will bring about some reconfiguration of employment, even if the pace and scale of impact is as yet unknown. Looking holistically, however, people are central to an AI system's development and are likely to remain so. From the beginning stages of problem and goal articulation, through to data collection and curation, model and product design, and user research and testing, people are the engine for the system's creation.

---

[31] https://pair-code.github.io/facets/
[32] https://github.com/google/ml-fairness-gym
[33] https://pair-code.github.io/what-if-tool/
[34] https://www.nist.gov/cyberframework; https://www.nist.gov/privacy-framework/privacy-framework

The development of clear standards, frameworks, and benchmarks can form the foundation of a core skill set for responsible AI leaders. Establishing a common lexicon, shared mapping of core concepts and fields, a widely accepted body of responsible practices, and an understanding of key outstanding challenges can serve as the basis of an education and training curriculum for future AI leaders.[35]

12. ***The extent to which the Framework should include governance issues, including but not limited to make up of design and development teams, monitoring and evaluation, and grievance and redress.***

Google supports the inclusion of governance issues in the RMF. Clear accountability and oversight, ongoing monitoring and evaluation, and mechanisms to collect feedback and address challenges are important to effectively manage risk throughout the life of a system. In general, AI applications developed and deployed in environments with strong governance structures in place will pose less risk than if they were being developed by an organisation without such stringent processes. Governance over AI should be viewed as a means to hold stakeholders throughout the AI chain accountable for responsible practices and risk management.

Conclusion

Developing this Framework is important not only because of the direct output of guidance and associated documentation that NIST is assembling, but also for the opportunity to host a collaborative discussion among diverse stakeholders. Google welcomes the opportunity to share insight based on our experience, and to learn from and engage with other participants. We look forward to continuing to work with NIST and our fellow stakeholders on these important matters.

[35] https://ai.google/education/