scale

SCALE AI RESPONSE TO THE NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY'S REQUEST FOR INFORMATION RELATED TO NIST'S ASSIGNMENTS UNDER SECTIONS 4.1, 4.5 AND 11 OF THE EXECUTIVE ORDER CONCERNING ARTIFICIAL INTELLIGENCE

Scale AI (Scale) is pleased to respond to the National Institute of Standards and Technology's (NIST) request for information (RFI) to assist them in carrying out several of their responsibilities under the Executive Order of Safe, Secure and Trustworthy Artificial Intelligence (AI) issued on October 23, 2024.[1]

Scale was founded in 2016 with the mission of accelerating AI development. Since our earliest days working to build high-quality datasets for autonomous vehicle programs to our work providing the data infrastructure for the leading AI companies, we have always been on the forefront of AI innovation. Today, we fine tune, red team or test and evaluate nearly every large language model (LLM) before it comes to market. Due to our role in the AI ecosystem, we believe that Scale is uniquely suited to lend our expertise to NIST as part of this RFI.

AI DEVELOPMENT IS HAPPENING AT RECORD PACE, BUT IT MUST BE DONE RESPONSIBLY

It is no secret that the world is heavily investing in AI due to the potential economic and societal benefits. Over the past year alone, we have seen billions of dollars invested globally and countless AI innovations come to market. While promising, this innovation must be done so responsibly. NIST has a critical role to play in order to help lead the development of the frameworks and best practices that will enable this to occur.

The Executive Order recognized this by tasking NIST with critical items like the developing guidelines and standards for AI safety practices, such as test and evaluation and international promotion of U.S. standards. Scale strongly supports

[1] See, https://www.federalregister.gov/documents/2023/12/21/2023-28232/request-for-information-rfi-related-to-nists-assignments-under-sections-41-45-and-11-of-the

NIST's work to do so as well as the forthcoming launch of the U.S. AI Safety Institute.[2] This work promises to bring together leading voices from across the AI ecosystem to make sure that the right approach to AI safety is established.

From the earliest days of our company, Scale has been working with the leading AI companies to build high-quality datasets that are critical for developing and training safe and responsible AI models. Today, through our work with the leading AI labs, we have helped to pioneer many of the techniques such as Reinforcement Learning from Human Feedback, red teaming and test and evaluation, which are now considered fundamental to the responsible AI development cycle.

More recently, and due to the critical nature of this work, Scale launched our Safety, Evaluation and Analysis Lab (SEAL),[3] which is our internal research lab working to accelerate responsible AI development. SEAL is fully committed to working with NIST to develop benchmark evaluation datasets and lending our expertise, as a leading AI safety company, to the discussion.

AI SAFETY BEGINS WITH STANDARDS AND FRAMEWORKS

AI safety must be recognized as a critical aspect of the entire lifecycle of AI development, from the earliest steps of development and model training to constant refining during deployment. Today, industry's approach is guided by best practices, but, moving forward, it's critical that this approach also relies on industry consensus standards and frameworks to ensure consistency. This does not mean that companies will not continue to build proprietary solutions, but that the underlying safety techniques will be based on standards and frameworks, not best practices alone. NIST has long led the development of this work on behalf of the U. S. Government, and once again, the Executive Order taps NIST to lead it for responsible AI.

This work began years ago with the development of the internationally heralded NIST AI Risk Management Framework (RMF)[4] that was developed before the launch of many generative AI applications. Therefore, the Executive Order correctly recognizes the need for a companion document for generative AI. Scale strongly supports the intended outcome of the Executive Order and believes that the topics suggested by the RFI, such as managing trustworthiness, model evaluation and validation, and the anticipated role of each actor in the AI ecosystem are the right ones for this document to encompass.

[2] See, https://www.nist.gov/artificial-intelligence/artificial-intelligence-safety-institute
[3] See, https://scale.com/blog/safety-evaluations-analysis-lab
[4] See, https://www.nist.gov/itl/ai-risk-management-framework

As industry develops these frameworks and standards, it will be essential to define common terms of reference to ensure consistency in current and future work products.

Beyond those topics, it will also be critical for the frameworks developed to become internationally recognized standards, as appropriate. Internationally recognized standards have long been critical to the development of technology, and critical gaps still exist for generative AI while standards bodies are conducting research. While AI standards work is currently underway in many standards bodies around the world, critical gaps still exist due to ongoing research topics. These gaps are correctly called out by the Executive Order on topics like red teaming and test and evaluation and addressing the gaps will be vital to developing safe, secure, and trustworthy AI.

Once these standards, frameworks, and best practices are in place, it will be critical that industry applies them to the AI development cycle. One critical item that the Executive Order identifies is the need for external test and evaluation prior to AI deployment.

A risk-based, sector-specific approach to test and evaluation is the best way to ensure that AI is safe to deploy for its intended use case because the anticipated risk of the use case drives the rigor of the evaluation. This approach, as outlined in the Executive Order, relies on the work that NIST is undertaking and is of the utmost importance for safe deployment of AI.

Scale strongly believes that the proposed requirement for external test and evaluation is necessary to provide the public the confidence that they deserve in the ability of the AI to carry out its intended use case. While companies already conduct internal test and evaluation as part of their AI development process, it is critical that external parties both confirm and ensure that the AI is truly safe to deploy.

NIST's work will be critical to bring together the entirety of the AI ecosystem–developers, deployers, and leading research bodies–to ensure that the techniques to carry out the test and evaluation are rapidly developed so that the AI ecosystem will have a uniform way to ensure that the capabilities and limitations of the models are understood.

Scale is strongly committed to doing everything we can to develop the techniques, benchmark datasets, and technology necessary, in collaboration with NIST and the rest of industry, to fill the current research gaps and build the necessary tools to ensure a robust and comprehensive test and evaluation system for AI.

EXPORTING UNITED STATES POLICIES AND STANDARDS ARE KEY TENENTS OF GLOBAL LEADERSHIP

Beyond the critical work to develop the frameworks, best practices, and standards for safe, secure and trustworthy AI, the Executive Order also taps NIST, along with the Secretary of State and other relevant stakeholders to craft a plan to globally harmonize these standards and frameworks. This work is critical to ensuring America's global leadership in AI development.

Around the world, various policy frameworks are being developed, but Scale firmly believes that the best regulatory framework to ensure that AI is developed safely is being developed in the United States because it prioritizes safety without stifling innovation. While there may not be as many headlines, the work that NIST and the soon-to-be launched U.S. AI Safety Institute are and will be conducting will strike the right balance between enabling innovation and establishing the proper guardrails to ensure safe AI deployment.

As has long been the case with emerging technology development, companies will look to develop their products for a global marketplace. For this reason, internationally recognized standards play a vital role in providing companies the certainty they need that a product developed in the United States will be accepted around the world. This does not necessarily mean that the policies governing the technologies' use need to be fully harmonized, but it does mean that the product has a path to viability in the market. NIST's work under the Executive Order is critical to making sure that this occurs for AI, much like it has for years with any emerging technology.

Scale stands ready and willing to do everything we can to assist NIST in this critical endeavor.

CONCLUSION

Scale looks forward to continuing to work with NIST and the entirety of the United States Government to ensure that the United States continues to lead the world in the development of AI.

Please direct any questions to the undersigned.

Max Fenkell
Head of Government Relations
Scale AI