548 Market Street #83849
San Francisco, CA 94104

(707) 656-2679

ARCPrize.org

March 12, 2025

Faisal D'Souza, NCO
Office of Science and Technology Policy
Executive Office of the President
2415 Eisenhower Avenue
Alexandria, VA 22314

Subject: AI Action Plan
Re: RFI on the Development of an Artificial Intelligence Action Plan
Submitted by email to ostp-ai-rfi@nitrd.gov

To Whom It May Concern:

ARC Prize Foundation is a non-profit dedicated to establishing ground truth on AI model capabilities and accelerating progress through benchmarking, evaluation, and open science. We appreciate the opportunity to provide input on the Artificial Intelligence (AI) Action Plan and highlight the critical role AI benchmarks play in maintaining U.S. leadership, fueling innovation, and protecting national security.

AI benchmarks are more than just measurement tools – they drive investment in key research areas and influence global standards that shape the competitive landscape.

To that end, we recommend the following actions:

1. Assert U.S. leadership in defining global AI standards

2. Establish a U.S. government hub for AI benchmarking

3. Cultivate an ecosystem of independent AI benchmarking organizations

AI benchmarking is a strategic capability that should be prioritized in the Artificial Intelligence Action Plan. We welcome the opportunity to discuss these recommendations further and explore potential areas for collaboration.

Sincerely,

Lauren Wagner

ARC Prize Foundation

ARC Prize Foundation Response: Request for Information on the Development of an Artificial Intelligence Action Plan

Introduction

America's continued leadership in artificial intelligence depends on our ability to accurately measure and analyze progress. In a time of rapid change, AI benchmarks provide government decision makers with the tools to assess models, identify technological gaps, and accelerate AI research and development.

The US has been at the forefront of AI benchmarking, setting standards that have driven breakthroughs in computer vision, language models, and hardware efficiency. However, global competition in AI measurement is increasing. Without continued engagement in this area, the U.S risks falling behind in shaping how AI systems are evaluated and deployed worldwide.

Role of AI Benchmarks

Drive Innovation And Competitiveness

AI benchmarks serve two critical functions:

1. **Provide objective measures for evaluating state-of-the-art AI models.** This generates ground truth about system capabilities and progress, shaping public and private sector decisions regarding AI research, security, procurement, and other policy priorities.

2. **Accelerate America's path to advanced artificial intelligence by defining what matters.** Benchmarks spotlight technological challenges and focus AI researcher efforts toward solving them.

Historically, the U.S. has pioneered impactful AI benchmarks that catalyzed major technological breakthroughs. For example:

- The 2010 ImageNet Challenge spurred revolutionary advances in computer vision by setting clear evaluation protocols for AI model performance.

- MLPerf, launched in 2018, has propelled rapid efficiency gains in AI model training and inference by systematizing performance metrics for hardware and software optimizations, benefiting AI deployment in sectors such as healthcare and autonomous systems.

- The SuperGLUE benchmark, introduced in 2019, provided a more challenging test for evaluating natural language processing (NLP) models, encouraging the development of more advanced systems.

Support National Security And Economic Resilience

AI benchmarking is not just a measurement tool, but a strategic capability that underpins America's national security, economic resilience, and technological edge. Today, geopolitical adversaries are investing in benchmarking infrastructure to promote the development of international AI principles, through organizations like the International Telecommunications Union with 194 member states. For decades, U.S.-led AI benchmarks have fostered an open, transparent ecosystem, influencing global technical standards, regulatory frameworks, and competitive dynamics. However, the opportunity for the U.S. to lead in this area is narrowing, making federal engagement crucial.

About ARC Prize Foundation

ARC Prize Foundation is a nonprofit AI benchmarking organization founded in 2024 by technologists Mike Knoop (co-founder, Ndea and Zapier) and François Chollet (creator of ARC-AGI, Ndea, and Keras).

In 2019, Chollet developed ARC Prize's inaugural benchmark, ARC-AGI-1, as the only benchmark that measures AI models' ability to efficiently acquire *new* skills for novel tasks - the essence of intelligence - making it a powerful indicator of progress toward more advanced systems. Since then, ARC Prize has analyzed major AI model releases, including from OpenAI, DeepSeek, and Anthropic, and launched a public competition generating 15,000 technical submissions and 40 academic papers from 1,500 AI researchers. In December 2024, OpenAI selected ARC-AGI-1 as the only headline benchmark for their o3 model launch. ARC Prize maintains active collaborations with leading U.S. research universities, AI labs, and open source communities. Moving forward, the foundation will continue developing and deploying benchmarks to fuel AI progress.

Recommended Policy Actions

Assert U.S. Leadership In Defining Global AI Standards

The nation that sets AI measurement standards has significant influence over the global AI ecosystem, guiding how AI systems are designed, assessed, and deployed across industries. Countries often embed these criteria – developed by organizations like the International Telecommunications Union (ITU), International Organization for Standardization (ISO), and International Electrotechnical Commission (IEC) – into legal frameworks and procurement criteria.

China is positioning itself as a dominant force in AI standards-setting, actively shaping key policies put forward by ISO, the ITU, and the IEC. A China-led global AI benchmarking system could disadvantage U.S. firms, ultimately shifting AI leadership away from the U.S.

We recommend that U.S.-led AI benchmarks anchor engagement with global bodies, ensuring these rules reflect American values rather than strategic competitors'.

Establish A U.S. Government Hub For AI Benchmarking

As AI capabilities evolve rapidly, policymakers need a central resource to track, evaluate, and interpret AI progress. The Administration should commit to maintaining a national AI benchmarking initiative, within an existing agency such as NIST/AISI, to coordinate AI evaluation efforts across government and provide expert guidance.

Key functions include:

1. **Coordinate federal AI benchmarking activities and partnerships**
   This includes managing relationships with model testing partners, including AI labs, and assessing government benchmarking needs to bolster economic competitiveness, national security, and human flourishing.

2. **Assemble and maintain a portfolio of AI benchmarks aligned with national security and economic priorities**
   The most widely used AI benchmarks analyze models on two key dimensions: (1) real-world capabilities or (2) ability to generalize. As AI systems become more powerful, it's crucial to test models across a variety of benchmarks as it yields the most comprehensive view of model progress and where the industry is going.

3. **Generate reports to guide U.S. AI policymaking and global standards-setting**
   This initiative should leverage AI benchmarks to produce empirical scientific reports that educate Congress and other federal agencies on the current and potential capabilities of AI systems.

Cultivate An Ecosystem Of Independent AI Benchmarking Organizations

Independent AI benchmarking organizations like ARC Prize operate at the intersection of industry, academia, and civil society. This structure enables them to recruit technical talent, innovate, and pursue organization-specific goals, ranging from generalization (ARC-AGI), to knowledge-recall (MMLU), domain specific tasks (FrontierMath), and safety (AILuminate). Autonomy ensures evaluations remain credible and transparent.

To foster a strong benchmarking ecosystem, the government can:

- Facilitate collaborations to develop AI benchmarks aligned with government objectives. This recommendation draws on proven public-private models to incentivize progress, such as the National Science Foundation's Industry-University Cooperative Research Centers (IUCRC), which funds collaborative research between universities, government and private sector partners in cybersecurity, advanced manufacturing, and telecommunications; and DARPA's AI Next Campaign and Explainable AI Program, providing research funding to universities and private labs.

- Launch joint working groups focused on specific domains (e.g., cybersecurity, scientific reasoning, CBRN risks)

These partnership models enable the Administration to harness external expertise and embed government priorities into benchmarking efforts.

Our recommendations reflect ARC Prize's experience developing and deploying AI benchmarks that accelerate AI progress. In 2025, ARC Prize Foundation will continue to:

1. Develop new AI benchmarks to analyze AI systems and drive research in promising fields

2. Quantify AI model capabilities, benefits, and drawbacks

3. Empower the U.S. open source AI research community, partnering with leading universities, AI labs, and the broader AI community to spur innovation

Conclusion

The next several years are crucial for shaping the future of artificial intelligence. Strengthening AI benchmarking in the AI Action Plan will drive U.S. innovation, safeguard national security, and ensure AI systems align with strategic imperatives.

ARC Prize Foundation looks forward to continued engagement on these efforts.

This document is approved for public dissemination.