Request for Information on the Development of an Artificial Intelligence (AI) Action Plan

AI 2030 Task Force Inputs

March 15, 2025

This document is approved for public dissemination. The document contains no business-proprietary or confidential information. Document contents may be reused by the government in developing the AI Action Plan and associated documents without attribution.

1. Introduction

AI 2030 is a global initiative aiming to harness the transformative power of AI to benefit humanity while minimizing its potential negative impact. AI 2030 aims to bridge awareness, talent, and resource gaps, enabling trustworthy and safe AI adoption across public and private sectors. Currently, AI 2030 has a thriving community of over 2,000 members across 18 countries, including more than 80% members from the United States. With strong expertise within our network, we have contributed to major consultations, including the UN Global Digital Compact, US NIST, and the governments of Singapore and Malaysia.

To provide feedback on the Development of an Artificial Intelligence (AI) Action Plan RFI and advance AI for a safe, secure, and prosperous U.S. economy, we have established a dedicated task force. This task force is composed of highly experienced professionals with deep AI experience across multiple industries, ensuring a well-rounded and impactful approach.

AI 2030 is not a policy or advocacy organization but focuses on leveraging evidence and insights to inform discussions on AI. The information provided in this response is for general informational purposes only and does not constitute official policy, legal advice, or an endorsement of any specific regulatory framework. While we strive to ensure accuracy, the details shared are based on publicly available information and contributions from our community members. Any references to consultations, governments, or organizations do not imply direct affiliation or endorsement. Readers are encouraged to conduct their own due diligence and consult with relevant authorities or experts before making any decisions based on this information.

2. Feedback on the AI Action Plan Overall Approach

To sustain AI leadership while driving innovation, security, and economic prosperity, AI 2030 recommends prioritizing several critical focus areas in the AI Action Plan. First, to accelerate market creation for next-generation AI research, development, and adoption, the AI Action Plan should drive increased federal investment in AI innovation, expand public-private partnerships, and establish AI testbeds to support industry adoption. In addition, the plan would benefit from a more explicit articulation of how to balance regulation with innovation, particularly in setting guardrails that protect consumers while fostering AI-driven economic growth. A well-defined approach to accountability is also necessary, ensuring clear roles and responsibilities for AI developers, deployers, and regulators. Furthermore, promoting a fair and competitive AI ecosystem should be prioritized to prevent market concentration and ensure that small and medium-sized enterprises (SMEs) have the opportunity to compete and innovate alongside larger industry players.

Another critical area for prioritization is AI supply chain security, including hardware dependencies, semiconductor security, and software vulnerabilities. AI 2030 also recommends a more comprehensive approach to data collection, usage, and access. Issues such as copyright, data ownership, and equitable data access must be addressed

to ensure that all AI stakeholders, including SMEs, have the resources necessary to develop competitive AI solutions. As the amount and variety of AI-generated content continues to grow, distinguishing between AI-generated and human-created data is becoming increasingly urgent. Without clear standards, training data quality could deteriorate, leading to potential model failures or performance degradation over time.

Finally, stronger enforcement and monitoring mechanisms are essential to ensure AI policies translate into effective practice. Compliance frameworks, auditability, and real-time monitoring should be incorporated to mitigate potential risks and ensure adherence to established guidelines. Addressing these areas will help strengthen the AI Action Plan, fostering a competitive, innovative, and trustworthy AI ecosystem that supports both economic growth and societal well-being.

3. Responses to Key RFI Questions

Q1: What are the most critical challenges and risks associated with AI development and deployment?

AI 2030 has identified several key challenges that must be addressed to ensure the safe, equitable, and sustainable development of AI. These challenges span training data limitations, economic barriers, workforce preparedness, regulatory coherence, and ethical concerns. Below, we outline these critical issues alongside recommendations for targeted policy action.

1.1 Training Data Scarcity and Bias

A major challenge in AI development is the growing scarcity of high-quality training data. Large language models (LLMs) follow a two-step training process: an initial baseline training phase, which relies on vast datasets, followed by a post-training phase, which uses high-quality labeled data. As an example, the GPT-2 model was trained on 8 million documents scraped from websites, such as Reddit, social media platforms, and newspapers for a total of 40GB of text[1]. However, experts, including Ilya Sutskever[2], have warned that publicly available data suitable for baseline training is rapidly depleting .

As a result, AI developers are exploring synthetic data generation, but current techniques lack the quality and diversity required to maintain robust, unbiased AI systems. In an experiment, one of our experts used GPT-4 to generate synthetic data for a recommendation system by generating profiles for users. The system generated data, but 100% of the names were anglicised and the male-female ratio was 70-30. By updating the instructions, they were able to generate a less biased dataset that included names from different nationalities.

Additionally, training data across specialized fields and languages remains non-representative of the full diversity of human society, leading to performance disparities in AI models. Machine learning techniques like Transfer Learning[3] and new merging

[1] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language Models are Unsupervised Multitask Learners.
[2] NeurIPS Test Of Time. (2024). Neurips.cc. https://neurips.cc/virtual/2024/test-of-time/105032
[3] Fuzhen Zhuang, Zhiyuan Qi, Keyu Duan, Dongbo Xi, Yongchun Zhu, Hengshu Zhu, Hui Xiong, & Qing He. (2020). A Comprehensive Survey on Transfer Learning.

research like Multilingual Alignment-as-Preference Optimization[4] offer partial solutions, but the better approach, as yet, remains expanding diverse, real-world datasets.

AI 2030 recommends incentivizing open and diverse data collection initiatives, particularly for underrepresented fields and languages. A good example of an open dataset for training foundational models is the Allen Institute for Artificial Intelligence (AI2) Dolma dataset composed of content from diverse sources, such as the web, academic publications, code, books and encyclopedic material[5]. Policymakers should establish data-sharing partnerships between academia, industry, and government to improve AI training datasets. Additionally, clear ethical guidelines for synthetic data use should be developed to ensure that AI-generated content does not exacerbate existing biases.

1.2 Compute Costs and Market Concentration

Another major challenge is the prohibitively high cost of compute power needed to train foundational AI models. Training or fine-tuning advanced LLMs requires immense computational resources, which creates a significant barrier for startups, academic institutions, and organizations without extensive financial backing .

This economic divide concentrates AI innovation within a few well-funded corporations, limiting competition and slowing overall technological progress. AI 2030 urges policymakers to fund university-led research on cost-efficient training methods and create public cloud access programs for AI startups and researchers. These initiatives would help democratize AI development, ensuring that innovative solutions are not limited to those with deep financial resources.

1.3 Workforce Readiness and AI Talent Shortage

AI will lead to significant job displacement, rather than outright job elimination, as industries integrate AI-driven automation. However, there is no national strategy for workforce reskilling to prepare workers for an AI-driven economy. A national strategy should complement and support state-level initiatives by providing funding, policy guidance, and standardized training frameworks, ensuring that local programs are scalable and aligned with evolving industry needs.

Additionally, regulated industries such as healthcare, finance, and public administration struggle to attract AI talent, as policy constraints and slow adoption cycles deter skilled professionals from these sectors. With more companies moving toward in-house AI model development, the demand for deep learning experts, AI strategists, and domain-specific AI talent is expected to grow.

AI 2030 supports the creation of federally funded AI workforce training programs tailored to different industries. These programs should focus on reskilling workers for AI-augmented roles and ensuring that education systems integrate AI literacy. To prevent widening skill gaps across communities, the curriculum must be inclusive, accessible,

---

[4] Shuaijie She, Wei Zou, Shujian Huang, Wenhao Zhu, Xiang Liu, Xiang Geng, & Jiajun Chen. (2024). MAPO: Advancing Multilingual Reasoning through Multilingual Alignment-as-Preference Optimization.

[5] Soldaini, L., Kinney, R., Bhagia, A., Schwenk, D., Atkinson, D., Authur, R., Bogin, B., Chandu, K.R., Dumas, J., Elazar, Y., Hofmann, V., Jha, A., Kumar, S., Lucy, L., Lyu, X., Lambert, N., Magnusson, I., Morrison, J.D., Muennighoff, N., Naik, A., Nam, C., Peters, M.E., Ravichander, A., Richardson, K., Shen, Z., Strubell, E., Subramani, N., Tafjord, O., Walsh, P., Zettlemoyer, L.S., Smith, N.A., Hajishirzi, H., Beltagy, I., Groeneveld, D., Dodge, J., & Lo, K. (2024). Dolma: an Open Corpus of Three Trillion Tokens for Language Model Pretraining Research. ArXiv, abs/2402.00159.

and designed to accommodate diverse learning needs. This includes tailored approaches for both digital natives and individuals with limited prior exposure to digital technologies, ensuring equitable participation in the AI-driven economy. Immigration policies should also be updated to attract global AI talent to meet rising industry demand.

1.4 AI Misuse and Ethical Risks

Bias in AI-driven decision-making has already caused significant harm in critical sectors. AI-driven fraud is growing at an alarming rate, with deepfake-related scams increasing 30 times since 2019 and causing $3.1 billion in losses just last year. Banks are feeling the pressure, as 7 out of 10 have already faced AI-powered fraud attempts, yet fewer than one-third have the tools to stop them. By 2026, nearly 40% of identity theft losses could come from AI-generated fraud, making it clear that stronger oversight is needed.[6]

A recent paper[7] demonstrated that some foundational models fine-tuned for coding tasks can unintentionally lead to broad misalignment, such as asserting that humans should be enslaved by AI, giving malicious advice, or acting deceptively. The paper concluded that we need to understand when and why narrow fine-tuning causes broad misalignment, which has implications for AI safety.

AI 2030 recommends strengthening AI transparency requirements by mandating clear explanations for AI-driven decisions in critical sectors. Additionally, risk mitigation strategies must be embedded in AI model development, including mandatory impartiality testing and third-party audits. Policymakers should also introduce accountability mechanisms for companies deploying high-risk AI models to ensure ethical and fair AI usage.

1.5 The Growing Risk of SMEs Falling Behind in the AI Era

Small and medium-sized enterprises (SMEs), the backbone of the U.S. economy, face significant disadvantages in the AI era. Limited access to AI talent, high implementation costs, and a lack of technical expertise create barriers to AI adoption, putting them at a competitive disadvantage. If SMEs are not adequately included in AI development and fail to benefit from its advancements, it could lead to long-term economic disparities, reduced innovation, and a widening digital divide. Ensuring equitable access to AI resources, training, and infrastructure is essential to fostering inclusive growth and preventing systemic risks in the future economy.

By proactively addressing these challenges through robust data collection, accessible AI infrastructure, workforce development, and strong support to SMEs, the U.S. can sustain its leadership in AI while fostering a robust, competitive, and trustworthy AI ecosystem.

Q2: What policies or regulatory approaches should be prioritized to ensure safe and trustworthy AI development?

---

[6] Deloitte Insights. Deepfake Banking Fraud Risk on the Rise. Deloitte, 2024, https://www2.deloitte.com/us/en/insights/industry/financial-services/financial-services-industry-predictions/2024/deepfake-banking-fraud-risk-on-the-rise.html.

[7] Betley, J., Tan, D., Warncke, N., Sztyber-Betley, A., Bao, X., Soto, M., Labenz, N., & Evans, O. (2025). Emergent Misalignment: Narrow finetuning can produce broadly misaligned LLMs.

To ensure AI development remains safe, accountable, and beneficial to society, AI 2030 recommends a policy approach that balances innovation with legal accountability, risk mitigation, and economic opportunity. Below are key areas that should be prioritized in AI governance.

2.1 Establishing Legal Accountability Frameworks

AI technologies have the potential to cause harm, including privacy violations, constitutional rights infringements, and systemic risks. Clear accountability measures are necessary to protect individuals and businesses from AI-related harms. AI 2030 recommends clarifying legal responsibility across the AI ecosystem, ensuring that liability is appropriately distributed between AI developers, deployers, and application providers.

We commend the State of Illinois for introducing HB3529[8] that codifies the principles of AI governance as

- Safety: Ensuring systems operate without causing harm to individuals.
- Transparency: Providing clear and understandable explanations of how systems work and make decisions.
- Accountability: Identifying and holding individuals or companies responsible for the system's performance and outcomes.
- Fairness: Preventing and mitigating bias to ensure equitable treatment for all individuals.
- Contestability: Allowing individuals to challenge and seek redress for decisions made by the system.

One crucial area requiring attention is the applicability of 47 U.S.C. §230[9] to AI-generated content and decision-making. While Section 230 currently protects online platforms from liability for user-generated content, it is unclear whether these protections should extend to AI systems making automated decisions that result in harm. This is especially important as we have seen a proliferation of AI Agents designed to act autonomously in order to understand, plan and execute tasks[10]. Policymakers should explore whether modifications are necessary to hold AI deployers accountable while maintaining space for innovation.

Additionally, AI 2030 supports the establishment of an independent regulatory body to oversee AI accountability. This entity could recognize private rights of action for individuals harmed by AI systems, allowing for legal recourse in cases of wrongful outcomes, privacy breaches, or algorithmic bias. Targeted legislative measures are also needed to address emerging AI risks, such as deepfake misinformation, electoral manipulation, and AI-driven fraud.

At the same time, safeguards must be in place to protect third-party application developers who integrate external AI models into their products. Liability should be carefully assigned so that AI model providers remain responsible for underlying

[8] HB3529, 104th General Assembly State of Illinois 2025 and 2026. https://ilga.gov/legislation/fulltext.asp?DocName=&SessionId=114&GA=104&DocTypeId=HB&DocNum=3529&GAID=18&LegID=162219
[9] Legal Information Institute. (n.d.). 47 U.S. Code § 230 - Protection for private blocking and screening of offensive material. Legal Information Institute; Cornell Law School. https://www.law.cornell.edu/uscode/text/47/230
[10] Coshow, T., Gao, A., Pingree, L., Verma, A., Scheibenreif, D., Khandabattu, H., OlliffeGartner, G. (2024) Top Strategic Technology Trends for 2025: Agentic AI. https://www.gartner.com/en/documents/5850847

algorithmic risks, while application developers are held accountable for how AI is deployed and used within their platforms.

2.2 Prioritizing Safety and Explainability in AI Applications

The AI ecosystem is evolving beyond the traditional model where companies develop and deploy their own AI systems. Many businesses now rely on third-party frontier models to power their applications, shifting the point of AI deployment closer to consumers. Given this change, regulatory frameworks must adapt by ensuring accountability extends beyond model developers to those integrating AI into end-user applications.

AI 2030 recommends requiring companies deploying consumer-facing AI applications to implement safeguards that mitigate misuse, clarify intended use cases, and ensure trustworthy deployment. A tiered compliance approach should be adopted, where AI systems posing higher risks to security, fairness, or privacy are subject to stricter oversight, while low-risk applications face fewer regulatory burdens.

Explainability and interpretability should also be prioritized. Without clear mechanisms to understand how AI models function and generate outputs, developers cannot debug systems, users cannot trust AI decisions, and businesses may be reluctant to deploy AI due to legal uncertainty. AI 2030 urges policymakers to strengthen requirements for model transparency, auditability, and explainability tools, ensuring AI remains trustworthy and accountable.

AI 2030 encourages a reassessment of the security risks that could emerge in the case of deregulation, ensuring that AI-driven systems remain resilient against adversarial attacks and exploitation.

2.3 Aligning and Leading AI Governance with International Standards

International regulatory misalignment creates significant obstacles for U.S. AI companies expanding into global markets. In some cases, European regulators have prevented U.S. AI applications from launching due to compliance mismatches with local standards. This fragmentation not only stifles economic opportunity but also undermines American leadership in AI governance.

AI 2030 advocates for aligning U.S. regulations with international AI frameworks, ensuring businesses can navigate global markets with fewer compliance barriers. Harmonizing U.S. policies with standards such as the EU AI Act, OECD AI Principles, and ISO/IEC AI Safety Guidelines would demonstrate American leadership in shaping AI governance rather than resisting international collaboration.

By streamlining AI policies across global markets, the U.S. can reduce the regulatory burden on businesses while maintaining its competitive edge in AI development.

2.4 Expanding Economic Opportunity and Strengthening U.S. Competitiveness

AI policies should be designed to expand economic opportunity by ensuring that AI benefits support broad-based economic growth and prevent market concentration that limits competition. Regulations must create an environment where small and medium-sized enterprises (SMEs), startups, and independent developers have fair opportunities to innovate and compete.

AI 2030 recommends implementing proportionate governance mechanisms that scale regulatory requirements based on an AI system's risk level. For example, while large-scale AI models with systemic risks should be subject to rigorous compliance measures, smaller businesses deploying low-risk AI applications should not be overburdened with excessive regulations.

Additionally, economic growth should not be constrained by access barriers to AI resources. AI 2030 supports policies that fund AI research at universities, expand cloud computing access for startups, and incentivize trustworthy AI innovation through grants and tax credits.

A competitive AI ecosystem supports job creation, drives economic expansion, and strengthens America's position as a global technology leader.

2.5 Strengthening Whistleblower Protections and User Feedback Mechanisms

If policymakers aim to reduce proactive compliance burdens on private companies, alternative mechanisms must be introduced to ensure accountability when AI systems fail. Whistleblower protections and robust user feedback channels are essential in this context.

AI 2030 recommends establishing legal protections for whistleblowers who expose AI-related harms, ensuring they are shielded from retaliation when reporting risky AI practices. Additionally, companies deploying AI systems should be required to maintain mechanisms for users to flag harmful AI outputs, enabling real-world feedback to inform regulatory oversight.

A crowdsourced accountability model, where user-reported AI failures trigger regulatory review, can reduce compliance costs while addressing AI safety concerns effectively. Policymakers can balance innovation with public protection by replacing costly proactive regulation with responsive feedback mechanisms.

2.6 Updating Privacy and Copyright Laws for AI

Current U.S. privacy and copyright laws do not adequately address the implications of AI. The 1974 Privacy Act[11] provides limited consumer data protections, and while HIPAA covers healthcare data, broader regulations governing AI-driven personal data collection remain fragmented. Similarly, the 1976 Copyright Act[12] does not fully account for AI-generated content, creating legal uncertainty for businesses and creators.

AI 2030 recommends expanding privacy protections to require businesses to obtain explicit consent before collecting consumer data, ensuring individuals have control over how their data is used in AI training. Additionally, copyright laws should be updated to distinguish between AI-assisted works, which involve meaningful human input, and fully AI-generated content, clarifying legal rights and responsibilities.

These updates will help align AI regulation with evolving technological realities, providing certainty for businesses while safeguarding consumer rights.

2.7 Utilizing Voluntary Guidelines to Bridge the AI Regulation Gap

Leveraging voluntary frameworks for enterprises is essential in the absence of federal or state-level AI regulations. These frameworks can provide organizations with structured

[11] 5 U.S.C. § 552a - https://www.congress.gov/crs-product/R47863
[12] 17 U.S.C. Copyrights - https://www.law.cornell.edu/uscode/text/17

guidance to develop their own AI governance models, ensuring beneficial innovation while fostering trust. By adopting industry-driven best practices, enterprises can align their AI development with ethical principles, mitigate risks, and scale alongside their peers, creating a more cohesive and self-regulated AI ecosystem.

Q3: What mechanisms should be established for monitoring and enforcing AI governance?

A balanced AI governance approach should focus on ensuring compliance while avoiding over-regulation that stifles innovation. Practical mechanisms for monitoring and enforcement should include third-party audits, certification systems, privacy-preserving techniques, and regulatory sandboxes.

3.1 Third-Party AI Auditing and Compliance Certification

Instead of imposing heavy-handed AI model regulations, a third-party auditing system should be established, where training traces are available on demand for audit by a trusted third party. This ensures transparency without excessive regulatory oversight and allows for independent verification of AI systems' compliance with safety, fairness, and privacy standards. An example of third-party auditing is the IEEE CertifAIEd program[13] designed to assess the ethical risks of applications and enable organizations to demonstrate their commitment to deliver trustworthy AI solutions.

A federated evaluation framework can be adopted, allowing AI models to be tested within their developers' infrastructure without exposing proprietary data. A certification agent could execute predefined test cases locally, and a federated auditing protocol would enable oversight organizations to verify compliance remotely. Zero-Knowledge Proofs (ZKPs) can be used to validate test results without disclosing sensitive model details.

To ensure that AI governance is accessible and scalable, a government-backed AI compliance certification program should be established. This program could function similarly to ISO or HIPAA compliance, providing certifications for AI models and training datasets based on standardized safety, bias, and privacy criteria. Government authorities should explore funding opportunities to minimize financial burdens, particularly for startups.

3.2 Secure and Privacy-Preserving AI Model Auditing

A confidential computing approach should be implemented to protect AI intellectual property while ensuring regulatory compliance. Privacy-preserving dataset certification should also be prioritized. Differential privacy[14] audits, homomorphic encryption, and synthetic data validation can be leveraged to verify the integrity and compliance of datasets without exposing raw data.

A blockchain-based certification and audit system can provide tamper-proof records of AI model and dataset certifications. Regulatory bodies and stakeholders can track certifications transparently and securely by logging evaluation metadata and compliance proofs on a decentralized ledger.

---

[13] https://standards.ieee.org/products-programs/icap/ieee-certifaied/

[14] What is Differential Privacy? – MIT Ethical Technology Initiative. (n.d.). http://eti.mit.edu/what-is-differential-privacy/

3.3 AI Traceability and Watermarking

To enhance AI model accountability, a binary watermarking system should be implemented to trace AI-generated outputs back to their source. This system should include metadata indicating which model produced the content, who prompted it, when it was generated, and other relevant details. This is particularly critical for non-textual outputs such as images, audio, and video, where traditional tracking methods are insufficient.

For consumers, a "right to be forgotten"[15] mechanism should be established, allowing individuals to request the removal of their personal data from AI training datasets and inference systems.

3.4 Government-Provided AI Monitoring Tools and Infrastructure

To reduce compliance burdens on private companies, government bodies should provide scalable, automated AI governance tools. These tools should be designed to monitor AI models in real-time, detecting:

- Security vulnerabilities and adversarial attacks.
- Bias-related economic exclusion, ensuring AI systems do not inadvertently restrict consumer access.
- Privacy risks, flagging potential misuse of personal data.
- Additionally, government-backed auditing services should be available to help companies assess risks without harming their business interests. By offering low-cost AI monitoring solutions, businesses can focus on innovation while ensuring compliance with minimal operational disruption.

3.5 Regulatory Sandboxes for AI Development

A regulatory sandbox approach, similar to Singapore's AI governance model[16], should be adopted to allow companies to test AI applications in controlled environments before widespread deployment. This would enable startups and enterprises to innovate without immediate regulatory constraints, while collecting empirical evidence on AI risks and effectiveness.

3.6 "Trustworthy AI Pledge" to Scale Private Sector Actions

Leveraging the Trustworthy AI Pledge can serve as a foundational mechanism for guiding AI governance. By committing to trustworthy AI principles, enterprises can align their AI development with industry best practices. This pledge can act as a self-regulatory framework, fostering trustworthy AI adoption while encouraging continuous monitoring, peer accountability, and alignment with emerging regulatory standards.

Q4: How can the U.S. government support AI innovation while mitigating risks?

The U.S. government can foster AI innovation while ensuring trustworthy development by expanding access to AI model training, investing in foundational research, protecting technological advantages, enhancing education, and providing AI governance tools that support compliance without stifling growth.

---

[15] Wolford, B. (2018, November 5). Everything You Need to Know About the "Right to be Forgotten." GDPR.eu. https://gdpr.eu/right-to-be-forgotten/

[16] Singapore's Generative AI Evaluation Sandbox: https://www.imda.gov.sg/resources/press-releases-factsheets-and-speeches/press-releases/2023/generative-ai-evaluation-sandbox

4.1 Expanding Access to AI Model Training

Training state-of-the-art AI models is cost-prohibitive for most organizations due to the high price of GPUs and computing infrastructure. This concentration of AI capabilities within a handful of well-funded companies limits competition and innovation. To address this challenge, the government should fund research into reducing AI training compute costs, making large-scale model development more accessible.

Alternative computing paradigms such as quantum computing, energy-efficient AI chips, and novel model architectures should also be prioritized to reduce computational overhead. Expanding cloud-based AI research grants will allow startups, universities, and smaller enterprises to train models without requiring massive infrastructure investments.

4.2 Advancing Foundational AI Research

Sustained investment in fundamental AI research is critical for overcoming persistent technical challenges and ensuring long-term competitiveness. The government should focus on funding:

- Data management research to improve the efficiency, accessibility, and quality of training data, addressing one of AI's most persistent bottlenecks.
- Synthetic data integrity research to prevent risks such as "model autophagy disorder," where AI systems trained primarily on AI-generated content lose richness and diversity. Establishing research labs and collaborations focused on generating high-quality human-created content will be key to addressing data scarcity.
- Quantum computing and blockchain research both have the potential to enhance AI security, scalability, and efficiency.
- Renewable energy solutions to meet the increased demand for energy to manage the costs of AI training, ensuring that the pace of innovation is sustainable.
- Federated Learning and decentralized computing research, which could potentially democratize AI development while maintaining data sovereignty, represent a more efficient, secure, and accessible framework for advancing AI technology.

4.3 Protecting U.S. AI Leadership and Security

Ensuring that AI innovation benefits the U.S. economy and national security requires strategic safeguards on hardware, model security, and sensitive AI development processes. Protecting AI training traces from unauthorized access will help prevent model distillation, where competitors or foreign entities extract insights from proprietary AI models.

Cybersecurity measures must also be strengthened to secure AI infrastructure against adversarial threats at every stage of development, from model training to deployment. By implementing these protections, the U.S. can maintain a technological advantage while preventing AI proliferation that could undermine economic or national security interests.

4.4 Infusing AI into Public Education

To ensure that AI innovation translates into long-term economic growth and workforce readiness, AI literacy must be introduced at all levels of education. Modernizing K-12

curricula to include AI concepts, data science, and computational thinking will help prepare students for the realities of an AI-driven economy.

Beyond early education, vocational training and apprenticeships should be expanded to equip students with hands-on AI experience. Public universities should receive dedicated funding to develop AI specialization programs, ensuring that AI talent pipelines remain strong across various industries.

4.5 Providing AI Governance Tools to Reduce Compliance Burdens

Companies—particularly startups and smaller enterprises—often struggle with understanding and implementing AI governance requirements. Instead of relying solely on traditional regulatory enforcement, the government should develop AI governance tools that make compliance easier, cheaper, and scalable. These tools should include automated risk assessment frameworks, fairness monitoring systems, and real-time security detection to help businesses proactively identify potential AI risks without incurring high compliance costs. Additionally, clear frameworks for acceptable use of AI should be established.

Q5: How should AI impact assessments be designed to evaluate societal and economic risks?

AI impact assessments should be forward-looking, standardized, and multi-disciplinary, ensuring that risks are identified proactively rather than reactively. Given the rapid pace of AI development and its potential irreversible effects, assessments must go beyond traditional compliance checklists and focus on real-world monitoring, structured evaluation frameworks, and diverse stakeholder engagement.

5.1 Forward-Looking and Proactive Risk Assessments

AI risks are not always a direct repetition of past harms but can accumulate in unforeseen ways. Traditional regulatory approaches may fail to capture long-term, emergent effects, making it necessary to continuously monitor AI systems beyond their initial deployment.

Impact assessments should be designed to anticipate future risks rather than just evaluate past failures. Policymakers should engage researchers, industry leaders, and interdisciplinary experts through surveys, symposiums, and real-world testing initiatives to ensure that assessments remain relevant as AI capabilities evolve.

5.2 Multi-Stakeholder Review Processes

Effective AI impact assessments should involve experts from diverse domains, including technical, economic, legal, and historical perspectives. AI risks are not only technical failures but also socioeconomic disruptions that require a broad understanding of historical trends and systemic effects.

Including anthropologists, historians, and social scientists in AI evaluation processes can help identify patterns of long-term impact that may not be immediately evident.

5.3 Standardized AI Impact Assessment Templates

To ensure consistency and reduce unnecessary burden on businesses, the government should develop and distribute standardized AI impact assessment templates. These templates should include:

- *Scale of impact:* How widespread are the potential effects of this AI system?
- *Severity of risks:* What is the level of harm if something goes wrong?
- *Likelihood of occurrence:* How probable are different risks based on historical and experimental data?
- *Frequency of impacts:* Will risks accumulate over time or occur in singular events?
- *Mitigation strategies:* What technical and governance measures can reduce or eliminate identified risks?

5.4 Real-World Monitoring and Continuous Evaluation

Static AI impact assessments are insufficient given AI's dynamic nature. AI systems should be continuously monitored for unintended consequences, including economic distortions, security vulnerabilities, and societal shifts.

Real-world monitoring mechanisms should include:

- *Post-deployment audits* that assess AI impact over time rather than just pre-launch evaluations.
- *Public reporting mechanisms* that allow users to flag unforeseen risks and failures.
- *Independent review bodies* that conduct regular AI evaluations based on real-world data, not just theoretical modeling.

6. Recommendations for Implementation-Ensure AI is Safe, Secure, and Trustworthy

The AI Action Plan should be structured to streamline oversight, ensure practical governance tools, facilitate industry compliance, and maintain U.S. leadership in AI innovation. Effective implementation requires a clear governing structure, standardized compliance mechanisms, accessible AI governance tools, and structured education and workforce development programs.

6.1 Establishing a Centralized AI Governance Structure

To prevent fragmented oversight, AI governance should be administered under a single coordinating body that oversees implementation, compliance, and industry collaboration. The National Institute of Standards and Technology (NIST), and its U.S. AI Safety Institute[17], play a critical role in establishing standardized frameworks, ensuring alignment across stakeholders, and fostering a cohesive regulatory approach. This entity should be enabled to continue to:

- *Consolidate AI oversight functions under one governing organization*, ensuring clear accountability rather than multiple disconnected agencies.
- *Set clear, time-bound deliverables* for AI safety standards, data governance protocols, and compliance mechanisms.
- *Create an AI industry advisory board* with representatives from government, industry, and academia, ensuring that regulations align with technological feasibility.
- *Expand international cooperation* by integrating the U.S. AI Safety Institute into the newly launched global network of AI Safety Institutes[18], strengthening cross-border collaboration on AI risks and governance.

6.2 Implementing Practical AI Governance Tools

[17] https://www.nist.gov/aisi

[18]https://www.gov.uk/government/news/global-leaders-agree-to-launch-first-international-network-of-ai-safety-institutes-to-boost-understanding-of-ai

To reduce compliance burdens while maintaining high standards, the government should develop practical, automated AI governance tools that allow companies to integrate trustworthy AI practices without significant cost or resource investments. This should include:

- *Pre-built AI impact assessment templates* that businesses can adopt instead of designing their own risk evaluation processes.
- *Real-time AI risk monitoring systems*, provided as open-source or government-supported tools, to detect bias, security vulnerabilities, and unintended impacts.
- *A centralized AI compliance portal* where companies can submit required transparency reports, safety evaluations, and certification requests in a streamlined manner.

6.3 Structuring Third-Party AI Auditing and Certification

A third-party AI auditing framework should be implemented to evaluate AI models on demand, ensuring safety and fairness without requiring direct government intervention in private-sector development. This can be achieved through:

- *A federated auditing approach*, where AI models are tested within the developer's infrastructure using secure, privacy-preserving evaluation techniques.
- *AI model certification tiers*, similar to ISO and HIPAA compliance, where models receive certification based on their intended use case and risk profile.
- *Tamper-proof AI audit logs* using blockchain or similar technology to ensure long-term accountability while protecting proprietary AI development processes.
- *Industry-backed AI certification programs*, such as the IEEE CertifAIEd framework, to provide standardized benchmarks for AI governance.

6.4 Integrating AI Governance into Industry and International Standards

To reduce regulatory friction and support U.S. companies in global markets, AI governance should be structured to align with international frameworks where applicable. The implementation process should include:

- *Establishing AI certification equivalency with international standards* such as the EU AI Act, OECD AI Principles, and G7 Hiroshima AI Process[19].
- *Creating AI regulatory sandboxes in strategic industries*—such as finance, healthcare, and defense—to test AI models under real-world constraints while refining governance frameworks.
- *Engaging industry leaders in quarterly AI governance reviews*, ensuring that policies are adaptive, risk-proportionate, and technologically feasible.

6.5 Scaling AI Literacy and Workforce Development Programs

AI literacy must be integrated at all levels of education and workforce training to ensure that individuals and businesses can effectively adopt and utilize AI technologies. The implementation process should include:

1. *Mandating AI literacy in K-12 and higher education curricula*, with a focus on data literacy, ethical AI use, and AI-driven decision-making.

[19] https://www.soumu.go.jp/hiroshimaaiprocess/en/index.html

2. *Creating federally funded AI upskilling and reskilling programs*, particularly for industries facing automation risks.
3. *Establishing AI apprenticeships and hands-on training partnerships* between government, universities, and private-sector AI labs to prepare workers for AI-integrated roles.

7. Recommendations for Implementation-Advance AI for a safe, secure, and prosperous U.S. economy

7. 1 Accelerating Market Creation for AI Research, Development, and Adoption

Accelerating AI research, development, and adoption requires coordinated efforts among academia, government, and industry. Strengthening public-private partnerships through research consortia, pilot programs, and data-sharing will facilitate commercialization. AI testbeds and regulatory sandboxes should enable safe model testing and compliance. To drive adoption in key sectors, the plan should offer tax incentives, low-interest loans, and industry-specific programs. Expanding federal and state AI procurement will further promote enterprise adoption and public sector innovation.

7.2 Building a Private-Sector-Led AI Investment Ecosystem

To build a robust private-sector-led AI investment ecosystem, the AI Action Plan should introduce tax incentives, R&D credits, and public-private funding to drive AI innovation. Expanding incubators, accelerators, and venture funding will support startups in scaling cutting-edge technologies. AI investment matching funds and co-investment programs should reduce investor risks while accelerating AI deployment. Government-backed funds should prioritize high-impact AI projects that drive economic growth and societal benefits.

7.3 Enhancing AI Infrastructure and Compute Capabilities

Enhancing AI infrastructure and compute capabilities is crucial for advancing AI leadership. A national strategy should be implemented to expand access to high-performance computing resources, support domestic semiconductor manufacturing to reduce reliance on foreign supply chains, and promote the development of energy-efficient AI technologies. Establishing AI supercomputing hubs will provide startups, researchers, and businesses with the computational power necessary to drive AI innovation and maintain global competitiveness.

Attracting, Retaining, and Scaling World-Class AI Talent

Attracting, retaining, and scaling world-class AI talent is essential to sustaining AI leadership. The AI Action Plan should include immigration reforms that facilitate the entry of top global AI professionals, such as expanding H-1B and O-1 visa programs. Developing a national AI talent pipeline will be key to maintaining a strong and dynamic AI workforce.

Leveraging Voluntary Frameworks for Enterprise AI Adoption

In the absence of federal or state-level AI regulations, voluntary frameworks can provide enterprises with structured guidance to develop trustworthy AI systems. These frameworks will help organizations align with ethical principles, mitigate risks, and scale responsibly alongside industry peers. By adopting voluntary standards and principles,

businesses can proactively address AI governance challenges while fostering public trust and regulatory preparedness.

Implementing Targeted Support Programs for SMEs

SMEs are critical to the U.S. economy but face significant barriers to AI adoption, including limited access to talent, infrastructure, and funding. The AI Action Plan should offer targeted support programs such as financial incentives, AI training initiatives, and access to AI infrastructure to ensure these businesses can effectively leverage AI for growth, competitiveness, and innovation. Without these interventions, SMEs risk falling behind, exacerbating an AI-driven digital divide and limiting economic opportunities for smaller enterprises.

AI 2030 RFI Leads:

Daniela Muhaj (Head of R&D; & Co-Chair of Global Fellows Program AI 2030)

Xiaochen Zhang (Executive Director; & Chief AI Officer, AI 2030)

Sean Lee (Program Director AI 2030)

AI 2030 RFI Task Force Contributors:

Jen Gennai, AI 2030 RFI Task Force Member, Co-founder & Head of Responsible AI, T3

Lokesh Prakash Manohar, AI 2030 RFI Task Force Member, Head of Generative CoE, LatentView Analytics

Anderson Prewitt, AI 2030 RFI Task Force Member, Co-Founder, DRADPA

Manas Talukdar, AI 2030 RFI Task Force Member, Director of Engineering at Labelbox

Quentin Reul, AI 2030 RFI Task Force Member, Director of Global AI Strategy and Solutions, Information Services at expert.ai

Contact information:

Website: www.ai2030.org

e-mail: [redacted]