Democratic Governance of Frontier AI:

A blueprint for a federal framework

June 2, 2026

OpenAI

The countries that successfully harness artificial intelligence will shape the scientific, economic, and geopolitical trajectory of the 21st century. AI can accelerate scientific discovery, expand economic opportunity, strengthen national security, and help solve problems that once seemed unsolvable. At the same time, increasingly capable AI systems are beginning to demonstrate abilities that raise concerns about cyber offense, biological misuse, autonomy, alignment, and other threats to national security. We also see early signs of recursive self-improvement (RSI) in today's systems: where AI development is itself accelerated by AI. We expect this to increase competitive pressures among developers and nations, and create governance challenges that existing institutions are not equipped to address. As RSI emerges, societies will need ways to shape the trajectory of AI development and ensure that it serves human interests.

Democracies are uniquely positioned to ensure that powerful AI is developed responsibly by coupling innovation with public accountability, transparency, independent oversight, and the ability to course-correct through representative government. But effective governance requires visibility into how frontier capabilities are evolving, how AI is impacting national security, and whether changing risk profiles warrant additional domestic safeguards, international coordination, or other precautionary measures. Building that understanding necessitates creating institutions capable of evaluating frontier AI, monitoring how capabilities evolve, and providing policymakers with reliable information.

As AI becomes increasingly important, democratic governments—not private companies acting alone—must ultimately determine the rules, safeguards, and accountability mechanisms. Our view is that decisions about the pace of AI innovation should not be left to any one lab, company, or special interest group. Instead, these choices should be made through democratic processes and informed by a robust understanding of frontier capabilities, risk mitigation measures, societal resilience, and geopolitical considerations.

In this context, we believe that the United States is particularly well positioned to help shape global governance. The US federal government possesses unique capabilities that no private company can replicate, including access to classified intelligence, expertise in cybersecurity and chemical, biological, radiological, and nuclear (CBRN) defense, secure computing environments, and the ability to coordinate with international partners. And many building blocks of a frontier safety framework already exist in the US today.

Frontier developers have already adopted White House voluntary commitments and partnered with the Center for AI Standards and Innovation (CAISI) for pre-deployment evaluations. US companies have also coordinated internationally, signing onto the European Union's AI Act Code of Practice and partnering with the United Kingdom's AI Security Institute. States have started developing harmonized approaches to frontier AI governance that include California's SB 53, New York's RAISE Act, and Illinois's SB 315, and the White House's recent executive order on Promoting Advanced Artificial Intelligence Innovation and Security is an important step forward. The US federal government must now build on that foundation and create a durable federal framework capable of evolving alongside the technology itself.

If artificial general intelligence is going to benefit all of humanity, the world needs more than voluntary commitments, individual company policies, and isolated regulatory interventions. It needs harmonized legal frameworks and durable institutions capable of adapting as technology advances. That framework should:

- **Address frontier risks to national security and public safety.** The primary goal of any frontier safety framework should be to mitigate the most severe risks posed by advanced general-purpose AI systems. These include risks related to cyber and CBRN threats, RSI progress, and loss-of-control scenarios that could result in catastrophic outcomes.
- **Advance democratic governance.** Decisions about how society manages frontier AI risks should be made through representative government, not by private companies acting alone. Frontier safety governance should reflect the strengths of free societies: transparency, public accountability, independent oversight, and the rule of law.
- **Promote transparency.** Governments, researchers, businesses, and the public need reliable information about how frontier AI is developed, evaluated, and deployed. Transparency creates accountability, supports independent scrutiny, and helps ensure that policy decisions are informed by evidence.
- **Protect innovation.** A frontier safety framework should focus on the highest-consequence risks without creating unnecessary barriers for startups, researchers, and developers building on top of frontier capabilities. It should reduce risk without locking today's industry structure into law.
- **Build adaptive institutions.** AI is advancing rapidly, and frontier governance must be capable of evolving alongside the technology. Policymakers should create institutions that can learn, experiment, incorporate new evidence, and update standards over time.

Implementing a frontier safety framework will require action at multiple levels of government, as well as international cooperation. States can continue serving as laboratories of democracy by developing harmonized frontier safety laws. The US federal government should build on that foundation by creating institutions capable of evaluating frontier systems, identifying emerging risks, and informing decisions about how AI should be governed. And as AI becomes more powerful, policymakers will need a whole-of-government plan to build broader societal resilience.

This blueprint outlines a three-part strategy for achieving those goals: 1) building a national framework that leverages the emerging consensus reflected in state frontier safety laws; 2) strengthening CAISI as the US federal government's primary institution for frontier AI safety; and, 3) mobilizing a broader resilience plan across government to address the national security and public safety challenges posed by frontier AI.

1. Building a national framework through reverse federalism

States have been valuable laboratories for AI policy, helping to develop and test many of the ideas that now form an emerging consensus on frontier safety. As AI becomes increasingly capable, however, the most important frontier safety challenges will be national—and often international—in scope. Policymakers should build on that foundation and turn today's emerging consensus into a comprehensive federal framework.

This approach, which we call *reverse federalism*, allowed states to develop and refine common legal frameworks first, creating models that Congress should now adopt at the national level. California's SB 53, New York's RAISE Act, and Illinois's SB 315 demonstrate that a meaningful consensus has emerged around the core elements of frontier AI governance. These frameworks share common requirements, align with international approaches, and provide a practical starting point for federal legislation.

Policymakers should build on this consensus by establishing a national frontier safety framework that provides both robust safeguards and regulatory certainty. At a minimum, that framework should include:

- **Severe risk evaluations and mitigations.** Companies should evaluate frontier capabilities for risks related to cyber, CBRN, loss of control, misalignment, and progress towards RSI; implement appropriate safeguards; and explain why any residual risks are appropriately managed. Risk assessments and mitigations should be tailored to the model's deployment context.
- **Transparency requirements.** Companies should publish public frontier safety frameworks and transparency reports describing how they evaluate severe risks, implement safeguards, make deployment decisions, and responsibly track progress towards RSI, with appropriate redactions to protect security, trade secrets, and proprietary information.
- **Independent assessment and auditing.** Large frontier developers should annually retain an independent third party to audit compliance with frontier safety requirements, including implementation of the developer's frontier AI framework, internal controls, and governance structures. These audits should be underpinned by a set of common standards that allow for interoperable audits across jurisdictions.
- **Critical safety incident detection and reporting.** Companies should report critical safety incidents involving deployed models, including incidents related to risks covered by the frontier safety framework, dangerous model behavior, or unauthorized access to sensitive model weights.
- **Model weight security requirements.** Companies should implement cybersecurity and insider-threat protections to secure unreleased model weights.
- **Whistleblower protections.** Employees should be protected from retaliation when reporting credible concerns about severe risks, safety failures, critical safety incidents, or violations of law to company leadership, regulators, or other appropriate authorities.
- **Meaningful accountability mechanisms.** Companies should face enforceable consequences for failing to comply with transparency, reporting, and safety obligations.

Liability frameworks should preserve accountability for severe harms and should not provide blanket safe harbors from responsibility.

The requirements reflected in SB 53, RAISE, and SB 315 should serve as the foundation for federal frontier safety legislation—not its endpoint. Policymakers should build beyond them by establishing a formal role for CAISI, strengthening federal evaluation and assessment capabilities, and advancing the broader resilience measures described below. With this comprehensive federal framework in place, policymakers should also preempt state laws that seek to regulate the same frontier safety risks, creating a single national framework that combines strong safeguards with regulatory certainty. States should continue serving as laboratories of democracy in areas beyond frontier safety, including youth protection, electricity and environmental policy, and AI education and literacy.

2. Strengthening safety through strong institutions

As AI becomes increasingly capable, the US will need a trusted institution responsible for evaluating frontier AI, monitoring emerging risks, and providing policymakers with independent technical advice. RSI exacerbates the fundamental governance question of whether humans can retain the ability to understand, guide, and shape the trajectory of advanced AI: making it potentially the most consequential frontier safety issue of the coming decade. Yet policymakers currently have limited visibility into RSI progress, whether safeguards are keeping pace, or what indicators should inform future policy decisions. An institution like CAISI can help close that gap.

Policymakers should strengthen CAISI and build it into the world's premier institution for frontier AI evaluation, standards development, independent assessment certification, and coordination across government and with international partners. Its mission should be to provide the information and analysis that policymakers need to make informed decisions about national security, international coordination, the adequacy of safeguards, and the pace of AI development. Particular priority should be given to understanding progress toward RSI, developing reliable measurements of that progress, and ensuring that governments have the information needed to respond. As capabilities advance, CAISI should remain flexible and adaptable, with the ability to take on new responsibilities as emerging risks and governance needs become clearer.

At the same time, policymakers should be realistic about the challenges of building a new institution. Their immediate priority should be developing CAISI's technical expertise, operational capacity, and institutional credibility. New responsibilities should be introduced incrementally and paired with the personnel, infrastructure, funding, and authorities necessary to execute them successfully. The effectiveness of any framework will ultimately depend on how it is designed, resourced, and executed in practice, and we expect those details to be refined through continued engagement among policymakers, industry, and technical experts. The recommendations that follow suggest guiding principles for institutional design rather than prescribing every implementation detail and propose a phased approach to building an institution capable of supporting policymakers while also maintaining the speed, rigor, and technical sophistication expected by the private sector.

Build CAISI's foundation. Before CAISI can take on significant new responsibilities, policymakers must ensure that it has the resources, authorities, and institutional support necessary to succeed. Policymakers should:

- **Authorize CAISI and appropriate funding.** Establish CAISI as a permanent institution with clear statutory authorities and sufficient funding to conduct frontier model evaluations, develop safety standards, certify third-party assessors, and coordinate with national security and scientific agencies, as well as with international partners.
- **Elevate CAISI's authority and coordinate government support.** The CAISI Director should report directly to the US Secretary of Commerce or another senior Cabinet-level official, and the White House should coordinate staffing, resources, expertise, and operational support from departments and agencies across the federal government.
- **Provide flexible hiring authorities and public-service pathways.** Adopt hiring authorities similar to those used by CHIPS for America, allowing CAISI to recruit technical talent quickly and offer competitive compensation, while creating pathways for experienced AI researchers, engineers, and safety experts to serve temporary tours of duty in government.
- **Mobilize national security expertise and data.** Direct national security and scientific departments and agencies to support evaluations in these domains in coordination with CAISI and immediately make available personnel and data related to cyber, CBRN, and other national security domains to bolster CAISI's ability to assess and mitigate risks.
- **Secure access to classified compute.** CAISI should have access to classified computing environments capable of conducting frontier model evaluations, whether through dedicated infrastructure, interagency partnerships, or formal agreements with agencies or commercial providers.

Create a mandatory evaluation process. Once CAISI has sufficient technical expertise and operational capacity, policymakers should require the most capable frontier models to undergo a CAISI evaluation before public release. These evaluations should assess frontier capabilities, the effectiveness of associated safeguards and mitigations, and the resulting risk profile of the deployed system. The requirement should be narrowly targeted at the most capable systems and should allow for iterative deployment by establishing clear thresholds for when subsequent model versions require reevaluation. CAISI's role should be to conduct evaluations and recommend mitigations—not to approve or block deployments. Developers should remain responsible for deployment decisions, publicly disclose evaluation findings and how they responded, and remain accountable through transparency and reporting requirements.

Policymakers should also ensure that evaluations are completed within a defined statutory timeline and that the evaluation process does not disincentivize the beneficial process of iterative deployment. If CAISI fails to complete an evaluation within the defined time period due to bandwidth, hardware, personnel, or other constraints, developers should be permitted to deploy without penalty. Companies should also remain free to share models with trusted third-party evaluators, independent researchers, red-teamers, and testing partners before or alongside CAISI review. A strong evaluation ecosystem requires multiple sources of expertise rather than a single institutional gatekeeper.

Support independent technical assessments. Deployment-triggered evaluations alone may not provide governments with sufficient visibility into how frontier capabilities evolve over time. Policymakers should therefore direct CAISI to develop standards for independent technical assessments, establish a certification process for qualified third-party assessors, and help build a broader ecosystem capable of conducting these reviews, including by using AI itself. Policymakers should also require frontier developers above specified capability thresholds to undergo periodic independent technical assessments conducted by CAISI-certified organizations.

An initial priority for these assessments should be providing policymakers with ongoing visibility into progress toward RSI, highly capable internal deployments, frontier model security, internal monitoring practices, and the effectiveness of associated safeguards. RSI may become the defining frontier safety and governance challenge of the coming decade, yet policymakers currently lack reliable ways to measure progress toward it or to understand its implications. CAISI should therefore work with frontier developers, academic researchers, national security agencies, and international partners to rapidly develop methodologies, benchmarks, and indicators for measuring RSI and assessing what governance works. Given the pace of AI development, we encourage CAISI to treat RSI as an urgent priority.

Building on this work, CAISI should develop common standards for independent technical assessments and support the creation of a broader ecosystem capable of monitoring RSI progress and safeguards over time. Frontier developers should share appropriate RSI-related measurements with CAISI, and qualified third-party assessors should periodically evaluate those measurements using CAISI-developed methodologies. Technical assessment findings should be provided to CAISI, which should use this information to help policymakers understand progress, assess whether mitigations are keeping pace, and coordinate with national security agencies and international partners on the implications of AI progress and whether additional safeguards or policy responses may be warranted.

Policymakers should also assign CAISI the resources and authorities necessary to help build the assessment ecosystem itself. CAISI should be authorized to provide grants, cooperative agreements, and other forms of support to emerging assessment organizations, academic centers, and technical institutions developing evaluation, technical assessment, and auditing capabilities.

3. Mobilizing a whole-of-government resilience strategy

Strong institutions are necessary, but they are not sufficient. No evaluation process, assessment regime, or single organization can eliminate every risk. As AI becomes increasingly capable, democratic societies will need to make sure that defensive capabilities, public institutions, and societal resilience improve alongside them.

Frontier AI should therefore be treated as a national priority requiring coordination across national security, public health, cybersecurity, scientific, diplomatic, and economic agencies, as well as with international partners. Policymakers should pursue a resilience strategy that not only reduces risk,

but also strengthens society's capacity to respond to emerging challenges, adapt to changing conditions, and continue benefiting from AI.

Facilitate collaboration and international coordination on frontier AI safety. Frontier risks will not be addressed by any one organization or country acting alone. Policymakers should provide legal certainty that allows frontier developers to collaborate on safety-related issues, including sharing threat intelligence, evaluation methodologies, incident learnings, and best practices. Democratic nations should also work together to develop compatible safety frameworks, trusted channels for information sharing, and coordinated responses to serious incidents. Particular priority should be given to developing shared approaches for evaluating and responsibly communicating progress toward RSI, where a lack of shared measurements and transparency could intensify competitive pressures among developers and make it more difficult to determine when additional safeguards are warranted. The emerging network of AI safety institutes could help build this shared technical understanding and provide a foundation for coordinated action if additional safeguards become necessary. Over time, the network could help build the shared methods, technical expertise, and confidence-building measures needed for governments to assess whether agreed safeguards are being implemented effectively.

Protect America's compute advantage. Advanced AI capabilities depend on access to leading-edge semiconductors and large-scale computing infrastructure. Policymakers should strengthen export controls, close known loopholes, and invest in the compute, energy, and infrastructure needed to maintain US leadership. Strategic compute capacity would give the US the ability to evaluate, govern, and deploy frontier AI systems when national security demands it. Maintaining leadership in advanced compute is not only an economic and national security priority—it's also a frontier safety strategy.

Restrict the adoption of unevaluated frontier AI systems. Federal agencies should prohibit the use of frontier AI systems that have not undergone a recognized safety evaluation on government-owned systems and devices. These evaluations should assess systems as deployed, including associated safeguards and controls, rather than the underlying model in isolation. Agencies should also prohibit procurement of products and services that rely on unevaluated frontier models in sensitive government contexts. Trusted evaluation processes are essential to ensuring that frontier systems deployed within government meet appropriate security and safety standards.

Ensure defensive capabilities scale faster than offensive capabilities. Frontier AI will strengthen both defenders and attackers. Policymakers should invest in AI-enabled biodefense, cybersecurity, critical infrastructure protection, and rapid response systems that reduce the consequences of misuse regardless of where threats originate. OpenAI's *Cyber Action Plan* provides examples of how advanced AI can strengthen resilience by helping trusted defenders identify threats earlier, respond faster, and protect critical systems more effectively. AI should strengthen the institutions and systems that protect society, ensuring that resilience grows alongside capability.

Prepare for future resilience challenges. Frontier AI governance will continue evolving as capabilities advance. In *Industrial Policy for the Intelligence Age*, we outlined several potential

approaches to strengthening societal resilience, including AI trust infrastructure, model-containment playbooks, safety systems for emerging cyber and biological risks, and new approaches to international coordination. Policymakers should direct relevant agencies to evaluate the feasibility, costs, benefits, and implementation challenges associated with these and other proposals, and identify where additional authorities or resources may be required. Building resilient institutions today will help democratic societies adapt as future governance challenges emerge.

Building the institutions for democratic governance

Frontier AI will help shape the balance of economic, scientific, and geopolitical power in the 21st century. Democratic societies have a narrowing window to build the institutions needed to govern increasingly capable AI systems before frontier capabilities outpace existing frameworks. States have helped establish an emerging frontier safety baseline. Policymakers should now build on that foundation by creating a national framework, strengthening CAISI as the US federal government's primary frontier AI institution, and mobilizing a broader resilience strategy.

Because frontier AI is a global technology, democratic nations must also work together to strengthen technical institutions, develop compatible governance approaches, and coordinate responses to emerging risks. The framework outlined here is not intended to be the final word on frontier AI governance. AI is advancing rapidly, and many important questions remain unresolved. The goal is to build the institutions, standards, and resilience needed for democratic societies to understand, adapt to, and govern increasingly capable AI systems.