



Industrial Policy for the Intelligence Age: Ideas to Keep People First

April 2026

OpenAI

Let's Talk

The drive to understand has always powered human progress—creating a flywheel from science to technology, from technology to discovery, and from discovery onward to more science. That inexorable forward movement led us to melt sand, add impurities, structure it with atomic precision into computer chips, run energy through those chips, and build systems capable of creating increasingly powerful artificial intelligence.

In just a few years, AI has progressed from systems capable of fast, narrow tasks to models that can perform general tasks people used to need hours to do. Now, we're beginning a transition toward superintelligence: AI systems capable of outperforming the smartest humans even when they are assisted by AI. No one knows exactly how this transition will unfold. At OpenAI, we believe we should navigate it through a democratic process that gives people real power to shape the AI future they want, and prepare for a range of possible outcomes while building the capacity to adapt. That's what this document is for—to start a conversation about governing advanced AI in ways that keep people first.

The promise of superintelligence is extraordinary. Just as electricity transformed homes, the combustion engine remade mobility, and mass production lowered the cost of essential goods, superintelligence will speed up scientific and medical breakthroughs, significantly increase productivity, lower costs for families by making essential goods cheaper, and open the way for entirely new forms of work, creativity, and entrepreneurship.

Today, AI's impact on work is often measured by the time required for tasks that systems can reliably complete. Frontier systems have advanced from supporting tasks that take people minutes to complete, to tasks that take them hours to complete. If progress continues, we can expect systems to be capable of carrying out projects that currently take people months. This shift will reshape how organizations run, how knowledge is created, and how people find meaning and opportunity. It will also highlight the limitations of today's policy toolkit and the need for more ambitious ideas to keep people at the center of the transition to superintelligence.



While we strongly believe that AI's benefits will far outweigh its challenges, we are clear-eyed about the risks—of jobs and entire industries being disrupted; bad actors misusing the technology; misaligned systems evading human control; governments or institutions deploying AI in ways that undermine democratic values; and power and wealth becoming more concentrated instead of more widely shared. Indeed, we highlight these risks here to raise awareness of the need for policy solutions to address them. Unless policy keeps pace with technological change, the institutions and safety nets needed to navigate this transition could fall behind. Ensuring that AI expands access, agency, and opportunity is a central challenge as we move towards superintelligence. We should aim for a future where superintelligence benefits everyone, and where we:

1. **Share prosperity broadly.** The promise of advanced AI is not just technological progress, but a higher quality of life for all. Everyone should have the opportunity to participate in the new opportunities AI creates. Living standards should rise and people should see material improvements through lower costs, better health and education, and more security and opportunity. If AI winds up controlled by, and benefiting only a few, while most people lack agency and access to AI-driven opportunity, we will have failed to deliver on its promise.
2. **Mitigate risks.** The transition toward superintelligence will come with serious risks—from economic disruption, to misuse in areas like cybersecurity and biology, to the loss of alignment or control over increasingly powerful systems. Without effective mitigation, people will be harmed. Avoiding these outcomes requires building new institutions, technical safeguards, and governance frameworks so that advanced systems remain safe, controllable, and aligned—reducing the risk of large-scale harm, protecting critical systems, and ensuring people can rely on AI in their daily lives. As capability scales, safety must scale with it.
3. **Democratize access and agency.** As capabilities advance, some systems may need to be controlled for safety. But broad participation in the AI economy should not depend on access to the most powerful models—it should depend on access to AI that is useful, affordable, preserves people's privacy and expands their individual agency. Avoiding a concentration of wealth and control will require ensuring that people everywhere can use AI in ways that give them real influence at work, in markets, and through democratic processes.

The Case for a New Industrial Policy. Society has navigated major technological transitions before, but not without real disruption and dislocation along the way. While those transitions ultimately created more prosperity, they required proactive political choices to ensure that growth translated into broader opportunity and greater security. For example, following the transition to the Industrial Age, the Progressive Era and the New Deal helped modernize the social contract for a world reshaped by electricity, the combustion engine, and mass production. They did so by building new public institutions, protections, and expectations about what a fair economy should provide, including labor protections, safety standards, social safety nets, and expanded access to education.

History shows that democratic societies can respond to technological upheaval with ambition: reimagining the social contract, mediating between capital and labor, and encouraging broad



distribution of the benefits of technological progress while preserving pluralism, constitutional checks and balances, and freedom to innovate. The transition to superintelligence will require an even more ambitious form of industrial policy, one that reflects the ability of democratic societies to act collectively, at scale, to shape their economic future so that superintelligence benefits everyone.

On this path to superintelligence, there are clear steps we need to take today. People are already concerned about what AI will mean for their lives—whether their jobs and families will be safe, and whether data centers will disrupt their communities and raise energy prices. AI data centers should pay their own way on energy so that households aren't subsidizing them; and they should generate local jobs and tax revenue. Governments should implement common-sense AI regulation—not to entrench incumbents through regulatory capture but to protect children, mitigate national security risks, and encourage innovation.

But the magnitude of the changes we expect and the potential risks we foresee demand even more. We are entering a new phase of economic and social organization that will fundamentally reshape work, knowledge, and production. It requires not just incremental policy responses but ambitious policy ideas for tomorrow that we must start discussing today. This is the moment to start the conversation: to think boldly, explore new ideas, and collaboratively develop a new industrial policy agenda that ensures superintelligence benefits everyone.

In normal times, the case for letting markets work on their own is strong. Historically, competition, entrepreneurship, and open economic participation have lifted living standards and expanded opportunity. Capitalism, imperfect as it is, remains an effective system for translating human ingenuity into shared prosperity.

But industrial policy can play an important role when market forces alone aren't sufficient—when new technologies create opportunities and risks that existing institutions aren't equipped to manage. It can help translate scientific breakthroughs into scaled industries and broad-based economic growth.

A new industrial policy agenda should use government's existing toolbox for aligning public and private activities: research funding, workforce development, market-shaping tools, and targeted regulation. But governments should not act alone. Nongovernmental institutions should pilot new approaches, measure what works, and iterate quickly, then governments should reinforce successes by aligning incentives and scaling what works through procurement, regulation, and investment. This public-private collaboration should stave off regulatory capture and centralized control, instead preserving the freedom to innovate while ensuring that the onset of superintelligence isn't dominated by the most powerful forces in society.

We don't have all, or even most of the answers. Different paths will require different policy responses, and no single set of tools will be enough in any scenario. But we should aim to build an AI economy that is both open and resilient through policies that expand participation, broaden access to opportunity, and ensure that society has the safeguards and institutions needed to manage risk.



This document offers initial ideas for an industrial policy agenda to keep people first during the transition to superintelligence. It is organized in two sections: 1) building an open economy with broad access, participation, and shared prosperity; and 2) building a resilient society through accountability, alignment, and management of frontier risks. OpenAI is offering these ideas to help start a broader conversation about the kinds of policies and institutions needed to navigate the transition, a conversation that needs to happen among governments, companies, civil society, communities, and families. These ideas are intentionally early and exploratory, offered not as a comprehensive or final set of recommendations, but as a starting point for discussion that we invite others to build on, refine, challenge, or choose among through the democratic process. They also focus on the United States as a starting point, but the conversation—and the solutions—must ultimately be global. The transition to superintelligence is not a distant possibility—it’s already underway, and the choices we make in the near term will shape how its benefits and risks are distributed for decades to come.

1. Building an Open Economy

The promise of advanced AI is that it can benefit everyone by translating abundant intelligence into extraordinary progress. It can lower the cost of essential goods, expand opportunity, and give people more time for what is meaningful, relational, and community-building. It can help solve scientific challenges that still elude human effort: curing or preventing diseases, alleviating food scarcity, strengthening agriculture under climate stress, and speeding up breakthroughs in clean, reliable energy. The benefits of major investments in science could emerge within a single lifetime and reach communities far beyond traditional research hubs.

Yet the same capabilities making this progress possible will also disrupt jobs and reshape entire industries at a speed and scale unlike any previous technological shift. Some jobs will disappear, others will evolve, and entirely new forms of work will emerge as organizations learn how to deploy advanced AI.

These changes will not arrive evenly. Without thoughtful policies, AI could widen inequality by compounding advantages for those already positioned to capture the upside while communities that begin with fewer resources fall further behind, excluded from new tools, new industries, and new opportunities. There is also a risk that the economic gains concentrate within a small number of firms like OpenAI, even as the technology itself becomes more powerful and widely used. Workers using AI might well agree that it’s increasing their productivity without believing they’re seeing the benefits.

Maintaining an open economy that is easily accessed and participatory will require ambitious policymaking. The enclosed ideas include proposals to ensure that workers have a voice in the AI transition, since workers have deep knowledge about how work is actually performed and where AI can make work better and safer. Other proposals suggest new mechanisms to share returns from AI-driven growth by expanding access to capital, sharing economic gains more widely, and aligning the benefits of AI-enabled growth with higher living standards. And they aim to modernize economic security by



helping people navigate transitions, access new opportunities, and maintain stability as work changes. Together, they form a portfolio of ambitious, preliminary ideas for navigating a wide range of economic scenarios that the transition to superintelligence might create—all while striving to keep the economy open and broadly beneficial.

Worker perspectives. Give workers a voice in the AI transition to make work better and safer, including a formal way to collaborate with management to make sure AI improves job quality, enhances safety, and respects labor rights. Workers have deep knowledge about how work is actually performed and where AI can improve outcomes. They will be critical voices in understanding how AI can be used in workplaces to ensure that technological change will not only lead to improved productivity, but also lead to better jobs and stronger, safer workplaces. Allow workers to prioritize AI deployments that improve job quality by eliminating dangerous, repetitive, administrative, or exhausting tasks so employees can focus on higher-value work. At the same time, set clear limits on harmful uses of AI that could erode job quality by intensifying workloads, narrowing autonomy, or undermining fair scheduling and pay.

AI-first entrepreneurs. Help workers turn domain expertise into new companies by using AI to handle the overhead that usually blocks entrepreneurship (e.g., accounting, marketing, procurement). Pair microgrants or revenue-based financing with practical “startup-in-a-box” supports such as model contracts and shared back-office infrastructure so that new small businesses can compete quickly. Worker organizations could serve as enablers by offering training, providing shared services, and helping workers negotiate fair commercial terms and protect IP.

Right to AI. Treat access to AI as foundational for participation in the modern economy, similar to mass efforts to increase global literacy, or to make sure that electricity and the internet reach remote parts of the globe. (The internet still isn’t fairly deployed across the globe or even the US; learn from this and seek to rectify those issues when it comes to AI.) Expand affordable, reliable access to foundational models—the building blocks of modern AI systems—and make a baseline level of capability broadly available, including through free or low-cost access points. Support the education, infrastructure, connectivity, and training needed to use these systems effectively, and make sure that workers, small businesses, schools, libraries, and underserved communities are not excluded from the capabilities that drive productivity and opportunity.

Modernize the tax base. As AI reshapes work and production, the composition of economic activity may shift—expanding corporate profits and capital gains while potentially reducing reliance on labor income and payroll taxes. This could erode the tax base that funds core programs like Social Security, Medicaid, SNAP, and housing assistance—putting them at risk. Tax policy should adapt to ensure these systems remain durable. Policymakers could rebalance the tax base by increasing reliance on capital-based revenues—such as higher taxes on capital gains at the top, corporate income, or targeted measures on sustained AI-driven returns—and by exploring new approaches such as taxes related to automated labor. These reforms should be paired with wage-linked incentives that encourage firms to retain, retrain, and invest in workers, similar to existing R&D-style credits. Together, these changes would help stabilize funding for essential programs while supporting workforce transitions in an AI-driven economy.



Public Wealth Fund. Create a Public Wealth Fund that provides every citizen—including those not invested in financial markets—with a stake in AI-driven economic growth. While tax reforms help ensure governments can continue to fund essential programs, a Public Wealth Fund is designed to ensure that people directly share in the upside of that growth. Policymakers and AI companies should work together to determine how to best seed the Fund, which could invest in diversified, long-term assets that capture growth in both AI companies and the broader set of firms adopting and deploying AI. Returns from the Fund could be distributed directly to citizens, allowing more people to participate directly in the upside of AI-driven growth, regardless of their starting wealth or access to capital.

Accelerate grid expansion. Establish new public-private partnership models to finance and accelerate the expansion of energy infrastructure required to power AI. Use these models to address financing constraints, permitting delays, and siting risks that have limited high-voltage interstate and interregional transmission—and to deliver infrastructure at speed and scale, limit taxpayer risk, and share the upside with the public. Approaches could include reducing the cost of capital through targeted investment credits, direct and indirect flexible subsidies, or equity stakes; removing market barriers to advanced technologies such as advanced conductors and high voltage direct current; and providing a narrow federal authority to accelerate the construction of interregional transmission when it is in the national interest. Partnerships should be structured to minimize taxpayer exposure to commercial losses and ensure that expanded energy infrastructure translates into lower energy costs for households and businesses.

Efficiency dividends. Convert efficiency gains from AI into durable improvements in workers' benefits when routine workload declines and operating costs fall, including incentivizing companies to increase retirement matches or contributions, cover a larger share of healthcare costs, and subsidize child and eldercare. Incentivize employers and unions to run time-bound 32-hour/four-day workweek pilots with no loss in pay that hold output and service levels constant, then convert reclaimed hours into a permanent shorter week, bankable paid time off, or both. Where helpful, firms could also offer predictable "benefits bonuses" tied to measured productivity improvements so the efficiency dividend shows up both as long-term financial security and as time back for workers.

Adaptive safety nets that work for everyone. Make sure the existing safety net works reliably, quickly, and at scale, because if the transition to superintelligence is going to benefit everyone, the systems designed to provide economic and health security need to deliver without delay or gaps. That starts with unemployment insurance, SNAP, Social Security, Medicaid, and Medicare that are not just in place but fully functional, accessible, and responsive to the realities people will face during the transition.

Next, invest in clear, real-time measurement of how AI is affecting work, wages, job quality, and sectoral dynamics, using public metrics such as unemployment rates and indicators of regional or industry-specific displacement. These systems should provide policymakers with timely visibility into where disruption is occurring and how severe it is. Then, define a package of temporary, expanded safety nets (e.g., expanded or more flexible unemployment benefits, fast cash assistance, wage insurance, training vouchers) that activates automatically when these metrics exceed pre-defined thresholds. When disruption rises above those levels, support would scale up; as conditions stabilize, it



would phase out. This ensures that assistance is targeted, time-bound, and proportional to the scale of disruption, and also avoids a permanent expansion of programs.

Portable benefits. Over time, build benefit systems that are not tied to a single employer by expanding access to healthcare, retirement savings, and skills training through portable accounts that follow individuals across jobs, industries, education programs, and entrepreneurial ventures. Public programs can decouple key benefits from employment status by expanding access to retirement and training support regardless of where or how someone works. Implementation can run through portable benefit platforms that pool contributions from multiple sources and route them into standardized accounts attached to the individual, not the job. Retirement systems can also be modernized through pooled structures that allow workers to accrue benefits continuously across employers, reducing gaps and preserving continuity over time.

Pathways into human-centered work. Expand opportunities in the care and connection economy—childcare, eldercare, education, healthcare, and community services—as pathways for workers displaced by AI. Although AI can enhance these roles by reducing administrative burdens and enabling greater personalization, human connection will remain an essential part of the profession. As AI reshapes the labor market, these sectors can absorb transitioning workers if supported with investments in training, wages, and job quality. Governments can build training pipelines, support transitions into care roles, and incentivize employers to raise pay and improve conditions in fields facing chronic shortages.

These initiatives could be complemented with a family benefit that recognizes caregiving as economically valuable work and supports evolving work patterns. This benefit could help cover childcare, education, and healthcare while remaining compatible with part-time work, retraining, or entrepreneurship. Together, these efforts would expand access to care, strengthen communities, and create meaningful, human-centered work.

Accelerate scientific discovery and scale the benefits. Build a distributed network of AI-enabled laboratories to dramatically expand the capacity to test and validate AI-generated hypotheses at scale. These labs would integrate AI systems directly into experimental workflows by automating routine processes, capturing high-quality data, and enabling rapid iteration between hypothesis generation and testing. Then, build the physical systems and infrastructure needed to translate validated discoveries into real-world use at scale. This includes expanding the capacity of organizations to deploy new technologies, upgrading facilities and systems required for implementation, and aligning financing and incentives to support adoption. It also includes a sustained investment in people: training scientists, technicians, and operators to contribute to AI-enabled science.

These investments ensure that breakthroughs move beyond laboratories and into widespread use, while strengthening the workforce and operational systems required to build, maintain, and run the infrastructure that supports AI-enabled discovery. Both laboratory and production infrastructure should be deployed broadly across universities, community colleges, hospitals, and regional research hubs, not concentrated in a small number of elite institutions.



2. Building a Resilient Society

As AI systems become more capable and more embedded across the economy, they may introduce new vulnerabilities alongside new abundance. Some systems may be misused for cyber or biological harm. Others may create new pressures on social and emotional well-being, including for young people, if deployed without adequate safeguards. AI systems may act in ways that are misaligned with human intent or operate beyond meaningful human oversight. And as advanced AI reshapes how people, organizations, and governments operate, it may place new strain on the institutions and norms that societies rely on to remain stable, secure, and free.

We should be clear-eyed about the resilience required here. These new risks won't be isolated or suitable for addressing one at a time—AI will reshape how work is performed, how decisions are made, how organizations operate, and how states interact. Building resilience therefore means making sure people and institutions can adapt quickly, maintain meaningful agency over how these systems are used, and preserve broadly shared prosperity even as economic and social structures evolve.

Over the past several years, leading AI developers including OpenAI have focused heavily on upstream safeguards: development of global standards, transparency around evaluations, mitigations, and risks, and investments in model testing, red teaming, and usage policies designed to identify and mitigate risks before deployment. Policymakers have also focused here, codifying requirements in the EU AI Act and in US state-based regulation. At the same time, training and literacy efforts have expanded so that schools, nonprofits, small businesses, and communities can use AI tools more safely and effectively. These upstream efforts should continue.

But as AI systems become more capable and more widely deployed, resilience will also depend upon what happens *after* deployment—when systems must be monitored in real time, operate under uncertainty, and integrate into institutions not designed for agentic workflows.

This is not a new challenge. When transformative technologies have reshaped society in the past, they have introduced new risks alongside new benefits, and new systems were built to manage them as they scaled. As electricity spread, societies built safety standards and regulatory institutions. As automobiles transformed mobility, safety systems reduced risk while preserving freedom of movement. In aviation, continuous monitoring and coordinated response systems made flying one of the safest forms of transportation. In food and medicine, testing and post-market surveillance helped ensure safety in everyday use. In each case, resilience was not automatic—it was built with the luxury of time.

As we move toward superintelligence, building a resilient society will require a similar but speedier effort that kicks into gear now. The ideas below are a slate of ambitious approaches to building a more resilient society. They focus on building and scaling safety systems that operate in real-world conditions by establishing mechanisms for trust, accountability, and auditing. They suggest opportunities for strengthening governance so that advanced AI remains controllable, transparent, and aligned with democratic values. And they suggest approaches to improve coordination across companies, governments, and countries so that risks can be identified early, information can be shared, and



responses can be executed quickly when needed. Together, these proposals extend important safety work already underway and represent initial ideas to keep AI safe, governable, and aligned with democratic values.

Safety systems for emerging risks. Research and develop tools that protect models, detect risks, and prevent misuse across high-consequence domains, including cyber and biological risks as well as other pathways to large-scale harm. Expand the use of advanced AI systems for threat modeling, red teaming, net assessments, and robustness testing to identify and anticipate novel risks early and inform mitigation strategies. Develop and scale complementary protective systems; for example, rapid identification and production of medical countermeasures in the event of an outbreak and expanded strategic stockpiles to prepare for future risks. Then, catalyze competitive safety markets by creating sustained demand for these capabilities through procurement, standards, insurance frameworks, and advance-purchase commitments. Over time, this approach can make safeguards an output of innovation and competition, ensuring that defenses improve as quickly as the risks they are designed to address.

AI trust stack. Research and develop systems that help people trust and verify AI systems, the content they produce, and the actions they take—especially as these systems take on more real-world responsibilities. Advance the development of provenance and verification standards and tools that can build trust in AI systems while preserving privacy. This could include enabling secure, verifiable signatures for actions such as generating content or issuing instructions, and developing privacy-preserving logging and audit systems capable of supporting investigation and accountability without enabling pervasive surveillance.

These types of solutions should capture key information about system behavior and use while minimizing the collection of sensitive data, and be designed to support investigation or intervention under clearly defined legal or safety conditions. This work could also include developing and testing governance frameworks that clarify responsibility within organizations, including how accountability could be assigned to specific roles and how delegation, monitoring, and escalation processes could function as systems become more capable. Over time, these efforts could establish a foundation for accountability by building trust in AI interactions and helping ensure that when harm occurs, responsibility can be appropriately allocated.

Auditing regimes. Strengthen institutions such as the Center for AI Standards and Innovation (CAISI) to develop auditing standards for frontier AI risks in coordination with national security agencies. Use tools such as government procurement, advance-purchase commitments, insurance frameworks, and standards-setting to create and scale a competitive market of auditors and evaluators capable of assessing AI systems and products for safety and security risks, building auditing capacity alongside the technology. Standards should be designed for international adoption to reduce fragmentation and avoid creating unnecessary compliance burdens for small companies, as well as those operating across jurisdictions.

As we progress toward superintelligence, there may come a point where a narrow set of highly capable models—particularly those that could materially advance chemical, biological, radiological, nuclear, or cyber risks—require stronger controls, including pre- and post-deployment audits using the standards developed in advance. Apply these requirements only to a small number of companies and the most advanced models, preserving a vibrant ecosystem of less powerful systems and the startups building on them. This approach maintains broad access to general-purpose AI while applying targeted safeguards where failures could create the greatest harm, avoiding unnecessary barriers that could limit competition or enable regulatory capture.

Model-containment playbooks. Develop and test coordinated playbooks to contain dangerous AI systems once they have been released into the world. As AI capabilities advance, societies may face scenarios where dangerous systems cannot be easily recalled—because model weights have been released, developers are unwilling or unable to limit access to dangerous capabilities, or the systems are autonomous and capable of replicating themselves. In these cases, the challenge is containment: limiting the spread of dangerous capabilities, reducing harm, and coordinating responses under real-world constraints. Experience from other high-consequence domains, such as cybersecurity and public health, shows that even when full containment is not possible, coordinated action can still meaningfully reduce impact.

Mission-aligned corporate governance. Frontier AI companies should adopt governance structures that embed public-interest accountability into decision-making, such as Public Benefit Corporations with mission-aligned governance. These structures should include explicit commitments to ensure that the benefits of AI are broadly shared, including through significant, long-term philanthropic or charitable giving. At the same time, harden frontier systems against corporate or insider capture by securing model weights and training infrastructure, auditing models for manipulative behaviors or hidden loyalties, and monitoring high-risk deployments so no individual or internal faction can quietly use AI systems to concentrate power.

Guardrails for government use. Have policymakers establish clear rules for how governments can and cannot use AI, with especially high standards for reliability, alignment, and safety. These standards should be codified in law and reinforced through technical safeguards. At the same time, use AI to strengthen democratic accountability. As more government decisions are made through AI-assisted workflows, these systems will create clearer digital records of government reasoning and action that can be logged alongside other public records. With appropriate safeguards, oversight institutions such as inspectors general, congressional committees, and courts could use AI-enabled auditing tools to detect abuse, identify harms, and improve accountability at scale.

Also, modernize transparency frameworks (including the Freedom of Information Act) to allow citizens and watchdog organizations to use AI to review targeted questions about government actions while protecting sensitive information. This could include clarifying when AI-interaction logs and agentic action logs constitute federal records that must be retained for specified periods.

Mechanisms for public input. Create structured ways for public input so that alignment isn't defined only by engineers or executives behind closed doors. As advanced AI makes more decisions that affect people's lives, societies need shared clarity about what these systems are supposed to do, what values should guide them, and how well they are performing. Make alignment more democratic, legible, and accountable through transparent specifications, evaluation frameworks, and representative input processes. Developers should publish model specifications that describe how systems are intended to behave and share information about how those systems are evaluated. Governments and public institutions should help shape these standards by anchoring them in democratic laws and values, while establishing mechanisms for representative public input to be considered alongside traditional business stakeholders. Together, these approaches help ensure that the advancement of AI reflects the perspectives of the societies that must live with its consequences.

Incident reporting. Establish a mechanism for companies to share information about incidents, misuse, and near-misses with a designated public authority. The system should emphasize learning and prevention over punishment, with appropriately scoped public disclosures that ensure transparency and democratic oversight while protecting sensitive technical, national security, and competitive information. Near-miss reporting could include cases where models exhibited concerning internal reasoning, unexpected capabilities, or other warning signals—even if safeguards ultimately prevented harm—so the ecosystem can learn from close calls before they become real incidents.

International information-sharing around AI capabilities, risks, and mitigations. Strengthen national evaluation institutions as the foundation for international coordination, beginning with expanding the role of the CAISI as a trusted technical body for evaluating frontier systems, assessing safeguards, and informing government understanding of advanced AI capabilities. Building on this foundation, develop a global network of AI Institutes that collaborate through shared protocols for information exchange, joint evaluations, and coordinated mitigation measures.

Over time, this network could evolve into an international framework akin to the other multilateral institutions focused on safety and standards, one that gives trusted public authorities visibility into frontier AI development; and creates secure cross-lab and cross-country channels for sharing evaluation results, alignment findings, and emerging risks; and likewise supports communicating during crises. To enable effective collaboration, policymakers should ensure that companies can share safety- and risk-related information through these channels without running afoul of antitrust or competition constraints, using clear safe harbors and narrowly scoped information-sharing rules. This system should expand beyond a narrow focus on national security to include a broader range of societal risks, including impacts on youth safety and well-being.

Starting the Conversation

We offer these ideas not as fixed answers but as a starting point for a broader conversation about how to ensure that AI benefits everyone. That conversation should be inclusive and ongoing—engaging governments, companies, researchers, civil society, communities, and families—and should be mediated through democratic processes that give people real power to shape the AI future they want. It also needs to expand globally—bringing in the perspectives of cultures, societies, and governments around the world.

These ideas are our first contribution to that effort, but only the beginning. Progress will depend on continued iteration, experimentation, and collaboration across institutions and sectors. To help sustain momentum, OpenAI is: (1) welcoming and organizing feedback through newindustrialpolicy@openai.com; (2) establishing a pilot program of fellowships and focused research grants of up to \$100,000 and up to \$1 million in API credits for work that builds on these and related policy ideas; and (3) convening discussions at our new OpenAI Workshop opening in May in Washington, DC.

