Google

June 2026

A Pragmatic Approach to AI Governance in America





01 Executive Summary

02 Introduction

03 Standard-Setting for Frontier AI

- Building an independent regulatory organization for frontier AI with government oversight
- Defining benchmarks for frontier capabilities and setting standards
- Promoting frontier model transparency and audits

04 Policies for Widely-Deployed AI

- Workforce preparedness and economic opportunity
- Protecting kids and families in the digital world
- Modern energy infrastructure and data center ecosystems
- Provenance and information integrity
- Creativity, copyright, and the AI value exchange
- Protecting privacy

05 Conclusion

01

Executive Summary

Executive Summary

The debate over AI governance is stuck in a false choice between over-regulation and no regulation. There is a middle way: a pragmatic, dynamic, and evidence-based approach that recognizes the unique challenges and opportunities posed by frontier AI on the one hand and widely-deployed AI on the other.

Standard-Setting for Frontier AI Models

Frontier AI is moving too fast for traditional bureaucracy, and we need a balanced approach to both drive progress and address risks. The good news is that America has solved similar challenges in the past through independent, industry-funded bodies that operate under federal oversight and establish guardrails to protect the public. Examples include the North American Electric Reliability Corporation (overseen by the Federal Energy Regulatory Commission), the Financial Industry Regulatory Authority (overseen by the Securities & Exchange Commission), the American Medical Association (overseen by federal and state health authorities), state bar associations (overseen by state supreme courts), and others.

We suggest federal policymakers consider a frontier AI regulatory organization (FARO), which could follow those models and enable:

Frontier Safety Standards

A FARO could progress and promote national and international standards, guiding requirements for how developers should identify and mitigate risks and verifying that companies implement security practices and incident response plans before releasing frontier models publicly, building on efforts like Google DeepMind's Frontier Safety Framework and similar efforts from other leading labs.

Annual Independent Procedural Audits

In the near term, a FARO could oversee an independent procedural audit regime for frontier AI companies. Such audits could be performed by well-established professional audit organizations on an annual basis, with reports provided to the FARO. Over time, as the FARO has identified appropriate standards, audits could shift to reviewing compliance with those standards.

Policies for Widely-Deployed AI (Below the Frontier)

For AI applications enabled by models at lower levels of capability, the federal government does not need new regulatory regimes that duplicate or conflict with existing law. AI applications like chatbots raise social and consumer safety issues distinct from the national security issues raised by the most advanced frontier AI models. For these widespread applications the federal government can draw on, and in some cases amend, existing laws and rules to address real-world outputs and specific harms.

Such an approach to widely-deployed AI applications could:

Shape the Future of Work

National workforce policies should gather data to inform sound public investments, skilling initiatives to facilitate worker transitions, and where warranted, modernization of unemployment insurance.

Protect Minors

Federal law should mandate evidence-backed AI interaction guidelines, disclaimers stating chatbots are not sentient, bans on gamified rewards, and "pause-and-direct" protocols for self-harm related queries.

Ensure Information Integrity

Congress should require watermarks like SynthID and tamper-resistant cryptographic provenance standards like C2PA for generative AI services, while companies use contextual transparency to avoid user "label fatigue."

Promote Copyright and Creativity

While AI training on public data is subject to fair use and text-and-data-mining exemptions, the industry is progressing mechanisms for value exchange with creators and governing infringing outputs under established copyright and notice-and-takedown regimes.

Develop Energy Infrastructure

Policymakers should launch a public-private initiative to scale America's energy generation and transmission system, with large AI data centers helping amortize fixed legacy grid costs and ultimately lowering residential electric rates.

Protect Privacy

Lawmakers should encourage development of Privacy Enhancing Technologies and build new technical standards for AI tools.

While the federal government should actively manage the risks of AI through balanced regulatory frameworks, the greatest risk of all would be missing out on AI's life-changing benefits. A federally overseen frontier AI regulatory organization, coupled with targeted policies for widely deployed AI applications and evidence-based adjustments to existing law, can address both national security and consumer protection risks while promoting economic, scientific, and social progress.

02

Introduction

Introduction

We are at a pivotal moment in AI development. AI's impact is visible in both everyday life and in extraordinary discoveries. While AI promises to drive scientific and economic progress, it also poses new challenges.

Recent advances in the capabilities of frontier AI models in areas like cybersecurity and biology have highlighted the need for appropriate safeguards and protocols for the most advanced models. And the widespread adoption of AI tools has raised concerns about the future of work, information integrity, energy infrastructure, and consumer well-being.

Google has long called for an approach to AI that is both bold and responsible. Today, we and others are asking important questions and offering different approaches to the challenges that come with any new general-purpose technology — questions like:

- How should America balance technological leadership and security risks?
- How can the country promote AI's economic, scientific, and social benefits while limiting the potential for harmful impacts?
- Should governments oversee AI development, availability, and applications?
- How can we base policy and regulation on objective evidence and standards, so that we intervene neither too soon nor too late?
- How can leading AI labs and broad AI deployers help address the disruptions that often accompany progress?

Just as there isn't a single question about AI, or a single goal policy can achieve, there is no single answer to what AI policy should be. In this paper, we build on the best ideas we have seen to separately address the **national security risks of frontier AI** and **the economic and social impacts of widely deployed AI**. These two types of AI pose different risks that demand different solutions and different regulatory regimes. We draw on frontier governance work from Google DeepMind, including lessons from the implementation of Google DeepMind's Frontier Safety Framework. Our thinking is also informed by years of work on the public policy issues addressed here, along with thoughtful approaches proposed by other companies and frontier labs, federal and state legislators, and civil society groups.

Our aim is to promote a dialogue about practical ways to ensure the United States leads the world in the responsible and secure development and deployment of AI. As we have long said, AI is too important not to regulate, and too important not to regulate well.

03 Standard-Setting for Frontier AI

Standard-Setting for Frontier AI

To protect American innovation and ensure global AI leadership while providing for a safe and secure digital future, leading labs need a unified framework for frontier AI safety, security, incident reporting, and transparency.

While we have supported initiatives in various states, most people agree that given its unique role, capabilities, and access to intelligence, the federal government should ultimately be the lead jurisdiction on issues of importance to national security.

Having studied the regulatory debates over frontier AI regulation and drawing on the ideas of others, we believe meaningful frontier AI safety and security policy should include the following elements:

An **independent regulatory organization** that can keep pace with fast-moving AI research and development.

Scientific **benchmarks** for identifying frontier capabilities in the cyber and chemical, biological, radiological, and nuclear (CBRN) domains, complemented by clear safety and security **standards** for building, testing, and deploying the most advanced AI systems.

Annual audits to demonstrate procedural, and ultimately substantive, compliance with safety standards, supported by model **transparency and reporting** requirements.

Below, we propose a framework for national-level regulation of frontier AI that incorporates these ideas.

Building an independent regulatory organization for frontier AI with government oversight

In response to AI's growing capabilities, there is an emerging consensus around the need for a regulatory approach of some kind, coupling expertise in state-of-the-art technologies with independent checks and balances. At the same time, the debate over AI governance often conjures a false dichotomy between innovation-halting regulation and the Wild West.

There is a middle path that would balance market-driven innovation and independent oversight: a **federally overseen frontier AI regulatory organization (FARO)**. A FARO could standardize and verify frontier AI labs' safety and security practices prior to launching frontier models on the U.S. market, including by ensuring that companies are publishing and complying with frontier AI frameworks and engaging with annual governance process audits.

The concept isn't a new one. Prominent examples of independent, government-supervised, industry-backed regulators include professional organizations like the North American Electric Reliability Corporation (overseen by the Federal Energy Regulatory Commission), the American Medical Association (overseen by federal and state health authorities), state bar associations (overseen by state supreme courts), and other groups overseen by the Securities & Exchange Commission, such as the Financial Industry Regulatory Authority (FINRA), the Public Company Accounting Oversight Board (which oversees audits of public companies), the New York Stock Exchange, and the NASDAQ. These organizations (and many more) are private, industry-funded bodies that write and enforce binding rules on their members, but operate under the supervision (and ultimate veto) of a government agency.

The FARO approach allows for nimble and informed decision-making particularly well-suited to the fast-evolving world of AI.

A FARO could pay market-rate salaries that would allow it to attract staff with significant AI expertise. Its employees could have security clearances necessary to access classified information relevant to various safety domains. And it could address coordination problems and information asymmetries across the AI industry. The FARO board would include a combination of independent directors and industry representatives, balancing considerations of safety and security against the need for bold innovation and speed to market in a global race.

Government oversight could be provided by any number of agencies, including the Commerce Department, the Treasury Department, or the Department of Energy. Importantly, the FARO would complement – but not supplant – programs that provide U.S. national security agencies with early access to certain models with advanced cyber capabilities.

Defining benchmarks for frontier capabilities and setting standards

Part of building a meaningful FARO would be empowering it to identify and enforce objective scientific measures for model safety and security. Developing evaluation frameworks that are applied consistently across frontier models and that evolve at the pace of AI technology development would facilitate consistent application of safety and security practices around frontier AI deployed on the U.S. market.

Working in concert with federal government experts in different safety domains, the FARO should maintain a repository of standards and benchmarks for building, testing, and deploying the most advanced AI systems.

It could also establish other safety rules based on industry best practices, guidelines from industry bodies like the Frontier Model Forum, and standards from the American National Standards Institute and the International Standards Organization. Ideally, governments around the world would implement international reciprocity agreements with the U.S. government, such that models verified by the FARO would be automatically verified in reciprocal jurisdictions. In this way, the FARO could set the global benchmark for safe and secure frontier AI models.

Companies would become FARO members if they have developed a frontier model. While AI researchers debate the technical definition of a frontier model, most agree that a mere recital of the number of floating point operations (FLOPs) used in training is insufficient and both over- and under-broad in assessing risks. That said, the 10^26 FLOPs standard used as a placeholder in other regulations could serve as an initial criterion, while the National Institute for Standards and Technology or the FARO promptly develop a capability-based standard to replace it. That standard should assess a model's ability to meaningfully facilitate cyberattacks or provide expert-level assistance in the creation of CBRN tools that pose risks to national security or public safety.

Promoting frontier model transparency and audits

By centralizing cross-industry standards for frontier AI safety, a FARO could enable the development of an independent AI auditing ecosystem.

Until substantive benchmarks and standards exist, the focus should be on standardizing model safety and security protocols. Each developer of a frontier model should publish and adhere to its own comprehensive frontier AI framework. These frameworks should have clear requirements: they should define tiered risk thresholds and corresponding mitigations, outline cybersecurity practices designed to secure unreleased model weights against unauthorized modification or transfer, establish critical incident response plans, and institute internal governance and oversight.

Most frontier labs already have such frameworks, and we have supported U.S. state legislation requiring frontier AI frameworks and other safety measures such as transparency and incident reporting.

Pending agreed national benchmarks, frontier labs could also contract with independent auditors to perform annual "procedural" audits to assess whether they have adhered to that framework. And before releasing a materially new frontier model, labs would attest to the FARO that they have adhered to their published framework and applied appropriate standards for testing and risk mitigation. In parallel, frontier labs should notify and provide early access to the U.S. government of models that advance the state of the art in sensitive national security domains, as outlined in the Administration's recent Executive Order on Promoting Advanced Artificial Intelligence Innovation and Security.

To ensure audits are effective and secure, three core operational protections should guide the requirements:

- *Standardized Reviews*: Auditors should have access to predefined, standardized, and focused sets of documents (such as model cards) to promote consistency and prevent the risk of loss or disclosure of sensitive IP.
- *Remediation*: Companies should have the opportunity to address audit findings and provide a plan to implement corrective changes before an audit report is finalized.
- *Confidentiality*: To prevent misuse of sensitive security findings and protect intellectual property, all final audit reports should be submitted confidentially to the FARO.

In the longer term, as standards bodies make progress and the FARO adopts and validates substantive benchmarks and technical standards, AI labs could be required to submit frontier models to substantive audits performed by independent auditors.

04

Policies for Widely-Deployed AI

Policies for widely-deployed AI

The issues raised by the widespread use of AI applications like chatbots are distinct from the kinds of national security issues posed by advanced frontier AI models. The U.S. federal government should also, but separately, address everyday uses of AI across the economy through a series of discrete frameworks.

Rules for the use of AI in daily life need to both minimize harms and maximize benefits. That means figuring out how our existing laws apply to this new technology. People on the left and the right agree that if something is illegal to do without AI, it's illegal to do with AI — we don't need to reinvent the wheel. Rather than replacing existing legal concepts wholesale, or coming up with AI versions of existing laws, we should work out how to apply time-tested rules to AI applications. In fact AI is already posing new questions under long-established law, like when to entrust fiduciary duty, how to determine intent and assign responsibility, and whether an AI can be the "author" of a patent or the creator of a new work of art.

We believe in an approach that is fundamentally data-driven, focuses on evidence of real-world benefits and harms, and accepts a degree of uncertainty to avoid regulations that slow progress without addressing real challenges. This approach would address outputs, not inputs, looking to prevent and mitigate specific harms rather than micromanaging the science behind these new tools. Among the challenges that Americans care most about, we believe the federal government should focus on jobs and the economy; kids' safety; the role of energy in our shared prosperity; protecting privacy and information integrity; and the important contributions of creators and publishers.

Workforce preparedness and economic opportunity

There is a great deal of uncertainty around how AI will impact the future of work and our economy. It's likely that AI will create new jobs (through growing some existing jobs and creating some new job categories), replace some roles, and change most jobs over the long run. We saw similar patterns with prior general-purpose technologies like electrification, computers, and the internet, each of which ultimately created many net new jobs. AI might concentrate wealth, but it could also make blue-collar jobs more valuable and shrink the pay disparity between white-collar and blue-collar jobs, trending toward a "convergence economy" rather than a "K-shaped" one.

We believe that with the right policy frameworks and investments in American workers, AI can be steered to augment the capabilities of the labor force, ultimately driving broader economic growth and social welfare.

The best response to uncertainty is usually to seek more information. In this case, that means better assessments of AI's economic impact. Workers need clear career pathways, and policymakers need robust and up-to-date evidence. Google is committed to sharing our detailed insights on how people use our AI tools at work and home, and we will be launching initiatives to support this vision soon. We are also committed to helping the U.S. overhaul its fragmented labor data infrastructure to help states better match job-seekers' skills with employer needs. Maintaining a nimble and flexible labor market is always important, but never more so than at a time of potentially rapid technological change.

There is also broad agreement that the workforce should be equipped with relevant AI skills. AI training has several advantages. It is relatively short and inexpensive compared to many forms of traditional training. Unlike most tools, AI itself can help people learn to use it better. And AI training is also highly adaptable: once they learn AI skills, workers can apply them across different AI models and in different professions. Many employers and governments at every level are already putting policies in place to upskill their current employees, enabling public-private partnerships such as AI Ready Ohio and ensuring that skills learned are relevant to workplaces.

> Google is also working closely with partners to equip the workforce, including a manufacturing vocational partnership, and initiatives to train over 300,000 American workers and all six million U.S. educators.

To help scale these kinds of efforts, we encourage policymakers to support private-sector, "earn-and-learn" training models. Funding employer-driven upskilling can offset the financial risk for businesses while helping workers acquire the specific, transferable skills that industry demands. And policy responses can encourage companies to invest in apprenticeship programs and skilling employees even when those employees may use those skills to find roles with other firms. We have endorsed a range of bills at the federal level that would help create these opportunities and incentives.

If and when evidence shows an uptick in unemployment as part of an AI-driven transition, the federal government may need to strengthen unemployment insurance and administration, potentially including portable benefits, broadened wage insurance, and reemployment support to help displaced individuals get back to work. Crucially, we must also protect worker dignity, promoting AI applications that augment workers' capabilities and frameworks that incorporate meaningful human oversight in critical employment decisions.

Public policy can also encourage AI adoption that benefits workers across all sectors. By embedding technical experts into small businesses, non-profits, and local governments for short-term, high-impact residencies, we and others across the AI industry are working to democratize the knowledge required to deploy enterprise-quality AI.

Protecting kids and families in the digital world

The widespread adoption of generative AI chatbots in the U.S. has highlighted concerns around potential harms to our children. AI developers and deployers should offer high-quality, privacy-protective, and age-appropriate AI experiences that empower kids, protect them from inappropriate content, and recognize their developmental needs. And rules should empower parents, who are typically in the best position to assess and address the unique needs of their children.

When it comes to school, we are seeing risks of AI tools helping students cheat on essays or do "cognitive offloading", but also a growing range of ways teachers are using personalized tutors and programs to enrich education and equip today's kids for the world of tomorrow. Some studies even show that students using AI tools tailored to their needs can learn much faster than students in traditional classrooms. But the details of implementation matter, and it is critical that companies work with educators, administrators, and academic experts to promote constructive learning environments and avoid negative or distracting uses. To empower younger users, platforms should also include tailored onboarding experiences and youth-friendly prompts that build AI literacy and teach about the technology's limitations.

There are also concerns that this new generation of AI tools might echo some of the features and challenges of digital technology and social media that have sparked a prior generation of concerns. Of course AI tools differ in important respects, and it's important not to fight the last war. But as conversational AI becomes more common, companies should follow strict guidelines for such interactions informed by child development experts.

AI platforms should be required to take reasonable measures to feature persistent disclaimers, filter out sexually explicit or romantic content, avoid claims the model is a person (and regularly point out that it's not), and not promote emotional dependency. Platforms should be required to ban unpredictable, engagement-focused rewards and gamification techniques that encourage unhealthy usage patterns for kids.

Federal regulation should also require that chatbots integrate robust suicide and self-harm protocols directly into AI systems. When a user searches or prompts for concerning topics the chatbot should pause the query and present "crisis resource pages," coordinating directly with lifeline services.

A principled and research-backed approach to kids' experiences with AI could form the foundation for operational, design, and reporting requirements focused on specific features of concern. Interaction disclosures, emotional dependency protections, tailored content restrictions, bans on gamified engagement, and self-harm interventions should all be considered as part of government-mandated child safety regulation for AI.

Modern energy infrastructure and data center ecosystems

America needs to build an energy system that is equipped to meet tomorrow's opportunities - smarter, cleaner, more affordable, and more reliable than ever before. Achieving this vision requires a generational shift in the pace and scale at which we construct energy infrastructure — an "Eisenhower Highway Program" for the American electric grid.

Data centers and associated power generation are essential components of realizing the opportunity to create a better electric grid. Research shows that data center growth can actually *reduce* electricity rates and create a more affordable energy system for all Americans, by spreading the fixed costs of the energy grid across a larger volume of energy use. But America also needs legislation that unlocks our ability to build power generation and transmission infrastructure at pace and scale. That's why we support comprehensive federal permitting reform legislation that can speed infrastructure construction while protecting the environment, unlock development of critical transmission lines to power growth, and establish permitting certainty for energy generation projects. Our U.S. energy policy agenda lays out a comprehensive framework for powering American growth while protecting affordability.

Data centers power the technology America relies on — everything from online banking to hospitals and 911 systems. The question is not data centers or no data centers, but how to build data centers the right way, responsibly and in partnership with communities. That entails:

- **Expanding energy capacity** for each new data center, because energy growth and ratepayer protection go hand-in-hand.
- **Paying for energy and infrastructure**, and not passing on costs to others. Long-term energy commitments should have built-in safeguards for local residents.
- **Supporting local communities**. Being a good neighbor means investing in the things that matter most – local jobs, schools, nonprofits, and the responsible stewardship of shared resources.

Google has signed up for the Ratepayer Protection Pledge, and supports enacting its provisions into federal law. Structures like the Capacity Commitment Framework and the Clean Transition Tariff help ensure that the costs of growth do not fall on other customers. These frameworks operate in multiple states and we continue to work with energy sector partners and policy makers to roll them out across the country.

The industry will need to drive data center efficiency at every level of the AI stack — from models to chips to data centers — as well as adding new energy to the grid to meet growing energy needs. Google has spent more than a decade contracting to add nearly 35 gigawatts of power to global grids — enough new capacity to power over 28 million American homes. We are also working to accelerate the next generation of energy technologies – advanced nuclear, natural gas with carbon capture, demand response, virtual power plants, fusion energy, geothermal, long-duration storage, and more – to assure a reliable, clean, and abundant supply of energy.

And energy is only one part of the picture. Robust water stewardship practices are another core component of our responsible growth strategy. That's why Google has committed to replenishing more water than we consume at our sites by 2030, and we encourage others to adopt similar practices.

Finally, leading companies are increasingly looking to help build the workforce of skilled tradespeople that are essential to constructing and maintaining the electrical infrastructure America needs - from the welders and pipefitters securing complex cooling systems, to the electricians and fiber technicians powering advanced network grids.

Provenance and information integrity

AI, particularly high-quality image and video generation, has renewed debate in the U.S. about the authenticity of content and the resulting quality of information exchange. As generative media becomes more advanced and accessible, it's helpful to know where content comes from, and whether it's been altered. We need a mix of policy and technical solutions to help verify whether content is generated or modified by AI in materially misleading ways, as well as providing other helpful context relevant to the origin and authenticity of content.

That will take a holistic approach that gives users tools and information to make informed decisions about content and its context, from the hardware where an image is captured to the platforms where it's displayed. Addressing these challenges requires robust, scalable, cross-industry solutions that provide transparency without hindering innovation. Federal regulation requiring reasonable measures to implement machine-readable, tamper-resistant watermarking and cryptographic metadata standards for generative AI services is an essential part of such efforts.

To ensure transparency remains meaningful, we will need to differentiate between routine, harmless AI edits (like red-eye reduction or stock backgrounds) and genuinely deceptive content. Failing to make this distinction risks accelerating a crisis of information fatigue. Regulatory frameworks should also avoid defaulting to mere labels. A labeling strategy that is not applied responsibly can lead users to incorrectly conclude that any unlabeled content is "real" or more trustworthy. Instead, we should use provenance signals to surface helpful information to users. For example, content-sharing platforms can read signals like C2PA Content Credentials and share them with users through techniques like labels and visible metadata. Ultimately, effective transparency will be contextual, not merely technical, and regulation should prioritize a holistic approach to information integrity.

We've been pleased at the **cross-industry support for our SynthID digital watermarking technology, and proud of our collaboration to develop the cross-industry C2PA Content Credentials specification**. Generative AI systems designed to provide information to the public should embed these kinds of provenance data in video, image, and audio content, and use commonly supported technical standards for watermarking or metadata.

Creativity, copyright, and the AI value exchange

Balanced copyright frameworks enable innovation while providing the certainty required for AI model providers and rights holders to collaborate on mutually beneficial partnerships and commercial agreements. Using publicly available web data for training models is a transformative, non-expressive use — like an art student taking inspiration from walking through a gallery — that should remain protected under fair use in the U.S. and text-and-data-mining exceptions abroad. While this use is legally protected, responsible model developers should give website owners choice and control over whether their content is used in model development via simple, machine-readable robots.txt tags, like the Google-Extended control. Such controls enable markets to work and allocate value where appropriate.

At the same time, companies developing and deploying AI broadly should recognize rights holders' desires to find beneficial pathways for creative professionals and knowledge workers. Like other platforms and AI model developers, Google is exploring new types of partnership and value-exchange models. For instance, we are piloting novel ways to partner with websites whose content meaningfully contributes to the freshness and factuality of generative AI responses through the process of grounding. We have also entered into deals in which we are paying for access to and delivery of diverse types of specialized, non-public content, including creative and educational content. Collaborations like these are the ones most likely to be sustainable for developers and meaningful for the ecosystem.

In addressing copyright concerns, the focus should again be on outputs — in this case, whether a specific image or piece of text actually copies an existing work, regardless of how it was created. Technical safeguards can help prevent models from generating outputs that reproduce works they were trained on. However, to protect the open space needed for new forms of creative expression, it is important that filters do not try to automate subjective decisions like whether something is “too similar” to a prior work. Here, the appropriate mechanism is through established notice-and-action frameworks that use standard reporting and takedown mechanisms to remove infringing content. We have also supported regulatory proposals that can help protect individual voices and likenesses by establishing a balanced national standard against unauthorized digital replicas.

Generative AI is a powerful tool for human creativity. When a user prompts, selects, and refines an AI tool's output, they can engage in a creative act. We see this next chapter as an expansionary opportunity for human creativity, and balanced copyright frameworks are a key to enabling that transition.

Protecting privacy

Precisely because consumer AI models are so responsive to individual user prompts, users need to be confident that their data will remain secure. User data should be protected at each stage of AI development and deployment, from when a model is created through when it is adapted and distributed. And as AI becomes woven into daily life, users' reasonable expectations for personalization and privacy are also evolving. People are prepared to share more specific information precisely because they get greater benefit when they do so. As innovation evolves, so should privacy rules. This requires a shift from “privacy by design” toward "privacy by innovation," where developers and deployers compete on responsiveness and privacy as core aspects of product quality, empowering users with new forms of transparency and control.

The federal government should build on the tried-and-tested foundations of existing privacy laws to guide responsible development practices, including through new technical standards. Experts in AI and privacy can advise on the prospects for privacy-enhancing technologies (PETs) like “zero-knowledge proof” tools. Regulatory incentives for investing in and deploying PETs can spur innovation while reinforcing trust across the global digital ecosystem. More work is needed to explore concepts like contextual integrity, where data is handled appropriately to the context of the interaction. And the privacy community should explore ways to embed privacy principles like data minimization and user control into the agentic ecosystem and protocols.

05 Conclusion

Conclusion

As the Collingridge Dilemma reminds us, it is possible to regulate new technologies too soon or too late, and errors in either direction come at a cost. To solve this dilemma now, the U.S. federal government should focus on data-driven regulatory approaches, not guesswork.

A federally overseen frontier AI regulatory organization, coupled with targeted policies for widely deployed AI models, can address both national security and consumer protection risks while promoting economic, scientific, and social progress. Achieving this vision depends on proactive, evidence-based collaboration among policymakers, industry leaders, and civil society. Together, we can build a secure and prosperous future where technological advances enhance human capability and contribute to human flourishing.