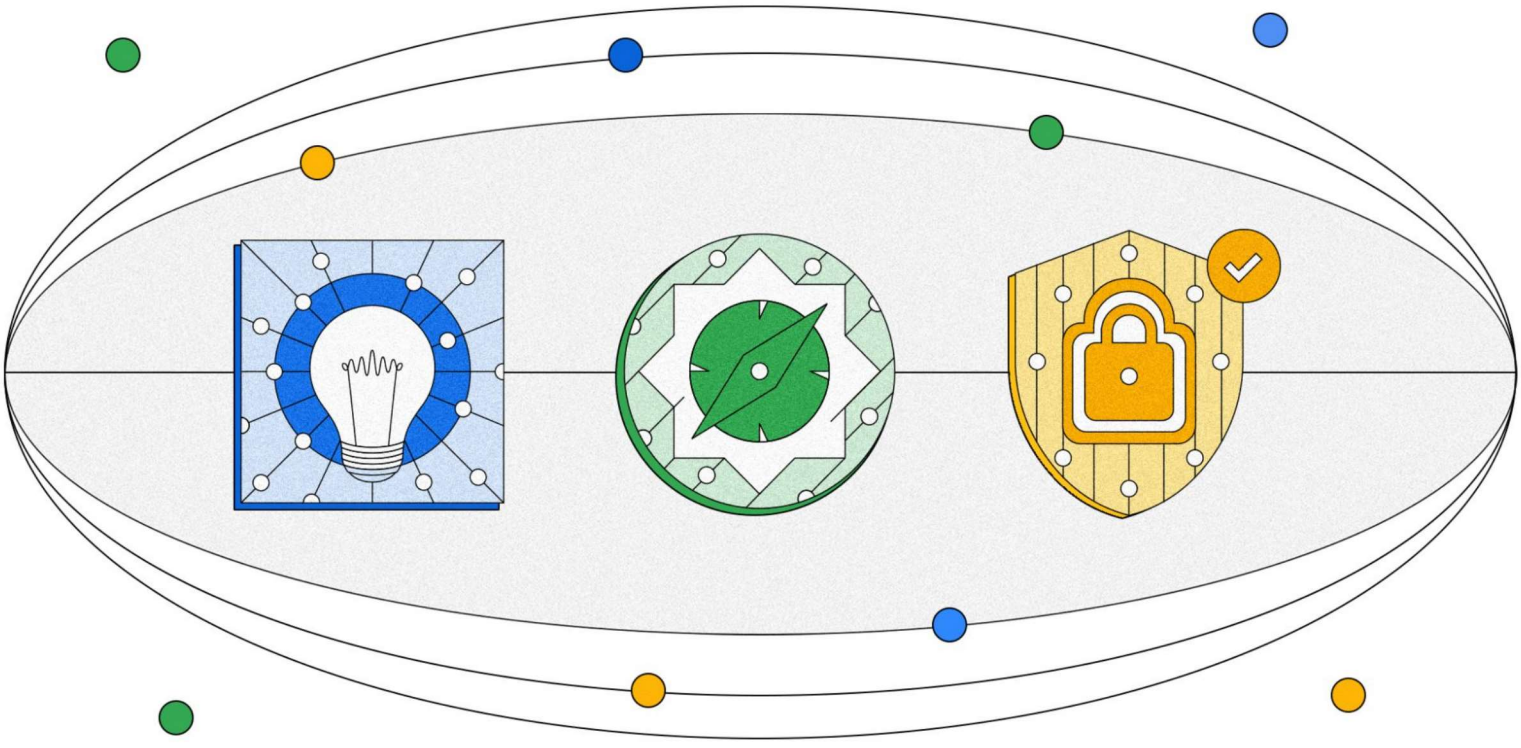
Google

A Policy Agenda for

Responsible Progress in Artificial Intelligence



Executive Summary

AI will be critical to our scientific, geopolitical, and economic future, enabling current and future generations to live in a more prosperous, healthy, secure, and sustainable world. Governments, the private sector, educational institutions, and other stakeholders must work together to capitalize on AI's benefits. A policy agenda that addresses **Opportunity**, **Responsibility**, and **Security** will help us boldly and responsibly realize AI's potential.

This paper lays out policy recommendations across those three pillars:

1) **Unlocking opportunity by maximizing AI's economic promise.** Economies that embrace AI will see significant growth, outcompeting rivals that are slower on the uptake. Adopting AI in existing industries represents an opportunity to move up the value chain, producing more complex and valuable products and services. AI also promises to help increase productivity despite growing demographic challenges. Governments, the private sector, educational institutions, and other stakeholders will need to work on joint and separate strategies to enable businesses, workers, and communities to capitalize on AI's benefits. Governments should increase investments in fundamental AI research, studies of the evolving future of work to help with labor transitions, and programs to ensure strong pipelines of STEM talent. Governments and industry need to deepen their efforts to upskill workers and support businesses meeting changing demands and new ways of producing goods and services.

2) **Promoting responsibility while reducing risks of abuse.** Without trust and confidence in AI systems, businesses and consumers will be hesitant to adopt AI, limiting their opportunity to realize AI's benefits. AI is already helping the world tackle challenges from disease to climate change, and can be a powerful force advancing both progress and fairness. But if not developed and deployed responsibly, AI systems could also amplify societal issues. Tackling these challenges will again require a multi-stakeholder approach to governance. Some of these challenges will be more appropriately addressed by standards and shared best practices, while others will require regulation – for example, requiring high-risk AI systems to undergo expert risk assessments tailored to specific applications. Other challenges will require fundamental research to better understand potential harms and mitigations, in partnership with communities and civil society.

3) **Enhancing global security while preventing malicious actors from exploiting this technology.** AI has important implications for global security and stability. AI can both help create and help identify and track mis- and dis-information and manipulated media. And it can drive a new generation of cyber defenses through advanced security operations and threat intelligence. Our challenge is to put appropriate controls in place to prevent malicious use of AI and to work collectively to address bad actors, while maximizing the potential benefits of AI. Governments, academia, civil society, and industry need a better understanding of the safety implications of powerful AI systems, and how we can align increasingly sophisticated and complex AI with human values.

Introduction

Artificial intelligence (AI), and specifically machine learning (ML), refer to a family of technologies that can perform increasingly complex tasks, improving through feedback. While the concept of AI goes back to Alan Turing's Imitation Game in 1950, machine learning has developed rapidly in the last decade, with new techniques and architectures drawing on internet data and increasing compute resources to lead to powerful new applications.

AI (as we'll call it throughout this paper) has driven advances both in everyday tools — like Search, Translate, Maps, Gmail, and chatbots — and in big challenges — like flood prediction, quantum computing, nuclear fusion, and medical science.

This latest wave of AI innovation and deployment has the potential to be revolutionary, changing the very nature of how we progress science and technology. Technological shifts — whether the printing press, the telegraph, radio, or the internet — have always brought social and economic change: higher living standards and greater opportunities, but also risks of disruption, inequality, and insecurity.

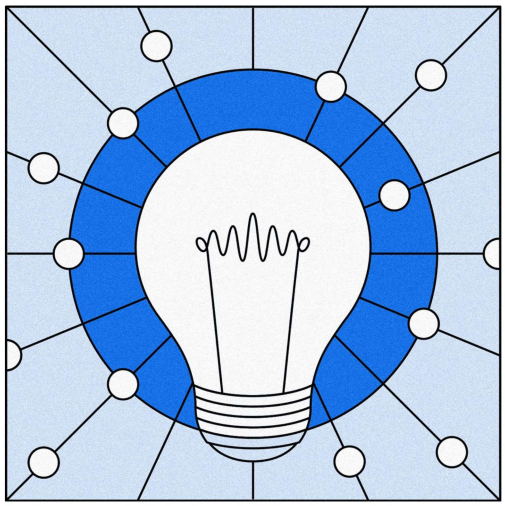
Calls for a halt to technological advances are unlikely to be successful or effective, and risk missing out on AI's substantial benefits and falling behind those who embrace its potential. Instead, we need broad-based efforts — across government, companies, universities, and more — to help translate technological breakthroughs into widespread benefits, while mitigating risks.

AI will be critical to humanity's future scientific and economic growth, enabling current and future generations to live in a more prosperous, healthy, secure, and sustainable world. Of course AI, if not developed and deployed responsibly, also presents important challenges. We must ask ourselves both what we *can* do, and what we *should* do.

Many countries and international organizations have already begun to act — the OECD has created its AI Policy Observatory and Classification Framework, the UK has advanced a pro-innovation approach to AI regulation, and Europe is progressing work on its AI Act. Similarly, Singapore has released its AI Verify framework, Brazil's House and Senate have introduced AI bills, and Canada has introduced the AI and Data Act. In the United States, NIST has published an AI Risk Management Framework, and the National Security Commission on AI and National AI Advisory Council have issued reports.

Getting AI innovation right requires a policy framework that ensures accountability and enables trust. We need a holistic AI strategy focused on: (1) unlocking opportunity through innovation and inclusive economic growth; (2) ensuring responsibility and enabling trust; and (3) protecting global security. A cohesive AI agenda needs to advance all three goals — not any one at the expense of the others.

Here are suggestions for a policy agenda that addresses key goals of opportunity, responsibility, and security.



Opportunity

Responsibility

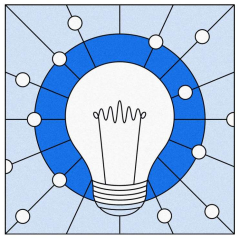
Security

Opportunity

Countries have historically excelled when they maximize access to technology and leverage it to accomplish major public objectives, rather than trying to limit technological advancement. From the adoption of machinery to produce textiles in Great Britain, which ushered in the industrial revolution, to the rapid adoption of TCP/IP by French, UK, and US researchers, ushering in the age of the internet, countries have established positions of global leadership by enabling innovation and capitalizing on the opportunities offered by technological change. AI offers the same opportunity — with potentially even great consequences given its power to bring major economic and social benefits and help address key societal challenges.

To compete on a global scale, countries will need to actively promote AI innovation and new applications of AI technologies. This includes advancing the technical state of the art through strategic investments in AI R&D, creating enabling legal frameworks for AI innovation, building an AI-ready workforce, and supporting workers and businesses as they adapt to new business models and ways of working. This is not just a question of "AI policy" per se, but redoubling commitments to science and technology, education, competition, and the rule of law.

One critical area will be maximizing AI's economic promise while managing the transitions for communities and individuals. AI promises to augment people's labor, making individuals, businesses, and other organizations more productive. It can automate increasingly complex tasks, enabling deep insights and new capabilities by organizations of all sizes — from enabling sustainable agriculture to improving cybersecurity. Boosting productivity has historically led to lower prices, higher wages, shorter working hours, new jobs, and better health care and living standards. Economies that embrace AI will likely see significant growth, outcompeting those that are slower on the uptake. For some countries, AI promises to help offset declining working populations through enhanced productivity, allowing economies to continue to grow, while for others, adopting AI in existing industries represents an opportunity to move up the value chain and produce more complex and valuable products and services.



To capture these benefits, societies will need to help workers gain new skills and businesses adapt to shifting demands and new ways of producing goods and services.

Opportunity

Specific Recommendations:

Invest in innovation and competitiveness

1) Grow investments in fundamental AI research through national labs and universities, and research foundations.
   a) Establish “grand challenges” for AI research that can benefit society, and/or push the boundaries of cutting-edge AI innovation, and offer incentives for achieving these milestones.
   b) Invest in AI safety and alignment research, in concert with leading AI labs.

2) Increase resources for AI researchers.
   a) Create shared AI research resources to make compute and data resources available to academics and small and medium enterprises conducting AI development and research.
   b) Create and fund public-private partnerships (PPPs) to build and maintain high-quality datasets for AI research, including public datasets that support multiple contributors and continuous quality improvements.
   c) Fund exploration and implementation of public-sector datasets that incorporate best practices for anonymizing and releasing data, or for drawing on confidential data without releasing it (e.g., the methods used in MedPerf).

3) Complementing the development of centers of government expertise, expand AI research in sectoral agencies to address key societal challenges, for example agencies overseeing environmental protection, public health, and disaster prevention and relief.

4) Establish and scale programs to educate government employees about AI to enable more effective policy making and investments.

5) Maintain and expand technology transfer frameworks to help universities both obtain government funding and progress new AI applications, pursuing AI innovations produced with federal funding and partnering with private tech companies to develop advanced AI applications.

Opportunity

Specific Recommendations:

Promote an enabling legal framework for AI innovation

1) Advance regulation and policies that help support AI innovation and responsible deployment.
   - **a)** Adopt or maintain proportional privacy laws that protect personal information and enable trusted data flows across national borders, and establish a legal framework for AI models' incidental use of such data on the open web for training purposes.
   - **b)** Establish competition safe harbors for open public-private and cross-industry collaboration on AI safety research.
   - **c)** Adopt or maintain copyright systems that enable appropriate and fair use of copyrighted content, while giving publishers and content creators choice and control over reproduction of their works.
2) Clarify potential liability for misuse/abuse of both general-purpose and specialized AI systems (including open-source systems, as appropriate) by various participants — researchers and authors, creators, implementers, and end users.

Prepare the workforce for an AI-driven job transition and promote opportunities to broadly share AI's benefits

1) Create a strong pipeline of local STEM and computer science talent.
   - **a)** Expand STEM and digital training programs in K-12 education.
   - **b)** Fund undergraduate, graduate, and post-doctoral research fellowships to promote AI and technology literacy.
   - **c)** Expand computer science and AI curricula and programs at public universities.
   - **d)** Provide scholarships and grants for students pursuing computer science and AI degrees.
   - **e)** Promote interdisciplinary education, combining AI and related technologies with domain-specific knowledge, to prepare graduates for an increasingly AI-driven job market.

Opportunity

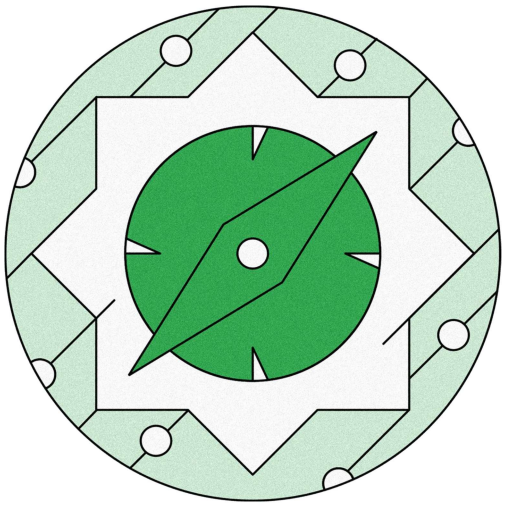
Specific Recommendations:

2) Invest in alternative pathways for non-college-degree holders.
   a) Incentivize employers to broaden their talent pipeline to enlist those who have acquired skills outside of four-year colleges (e.g., community college, military service, training programs, and on-the-job learning).
   b) Promote new mechanisms for verifying knowledge and skills developed through training programs and prior work experience, including career certificates in fields relevant to AI (e.g., data analytics).

3) Invest in community infrastructure to enhance economic mobility and inclusion (broadband, transportation, regional innovation hubs, affordable housing), improving labor market mobility by developing benefit systems that facilitate greater benefit portability across employers.

4) Attract leading technology talent by promoting greater transparency and flexibility in immigration pathways for international STEM experts.
   a) Expand and accelerate access to visas and citizenship for computer science and AI workers and their spouses.

5) Support workers who may be displaced by AI advances.
   a) Invest in continuing education, and upskilling and digital reskilling programs for displaced workers.
   b) Create public-private partnerships to bring together companies within industries to conduct industry-wide training and retraining.
   c) Ensure that digital skilling programs are accessible to workers who are displaced.



Opportunity

Responsibility

Security

Responsibility

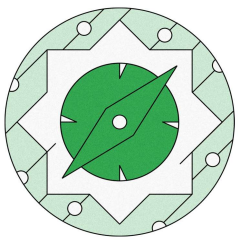
Capitalizing on the opportunity presented by AI requires a responsible approach. Without trust and confidence in AI systems, businesses and consumers will be hesitant to adopt AI, limiting their opportunity to capture AI's benefits.

AI is already enabling people to tackle societal challenges from disease to climate change, and can be a powerful force for equity. But if not developed and deployed responsibly, AI systems could also amplify current societal challenges.

For example, AI systems can help combat human bias in fields like lending or criminal justice through more objective, data-driven recommendations. But if not corrected, biases learned from existing human decision making or structural inequities can also exacerbate discrimination.

These promises and risks play out in fields like biometric systems (promoting privacy or enabling mass surveillance), health care (developing new treatments or entrenching disparate treatment), manufacturing (improving or harming worker safety), and the creative arts (potentially creating new opportunities for more creators but also disrupting existing compensation models). There are also inherent tensions, not just within but between domains. For example, effectively combating bias can require access to personal and demographic data that raises privacy questions.

Tackling these challenges will require a multi-stakeholder approach to governance. Learning from the experience of the internet, stakeholders will come to the table with a healthy grasp of both the potential benefits and challenges. Some challenges can be addressed through regulation, ensuring that AI technologies are developed and deployed in line with responsible industry practices and international standards. Others will require fundamental research to better understand AI's benefits and risks, and how to manage them, and developing and deploying new technical innovations in areas like interpretability and watermarking. And others may require new organizations and institutions.



And solutions inevitably come with trade-offs that must be reconciled. Focusing regulations on the highest-*risk* applications can also deter innovation in the highest-*value* applications where AI can offer the most significant benefits.Transparency, which can support accountability and equity, can come at a cost in accuracy, security, and privacy. Democracies need to carefully assess how to strike the right balances.

Responsibility

Specific Recommendations:

Pursue proportionate, risk-based regulation that enables responsible development and application of next-generation technologies

1) Require regulatory agencies to issue detailed guidance on how existing authorities (e.g., those designed to combat discrimination or protect safety) apply to the use of AI.
   - **a)** Direct sectoral regulators to update existing oversight and enforcement regimes to apply to AI systems, including on how existing authorities apply to the use of AI, and how to demonstrate compliance of an AI system with existing regulations using international consensus multistakeholder standards like the ISO 42001 series.
   - **b)** Instruct regulatory agencies to issue regular reports identifying capacity gaps that make it difficult both for covered entities to comply with regulations and for regulators to conduct effective oversight.

2) Require high-risk AI systems to:
   - **a)** Provide documentation describing how the system is intended to be used, known inappropriate uses, known risks, and recommendations for independent deployers and users to manage risk.
   - **b)** Undergo risk assessments by independent internal or external experts
   - **c)** Disclose the results of capacity and risk assessments, with protection for trade secrets and controlled technologies as appropriate.
   - **d)** Align documentation, risk assessment, and management practices with relevant standards, frameworks, and industry best practices as those standards develop.
   - **e)** Define “high-risk AI systems” as those intended for use in applications that pose a material risk of significantly harming people or property or imperiling access to essential services.
   - **f)** Establish proportionate penalties for non-compliant deployments, with reasonable opportunities to cure issues given the novel nature of the technologies.

3) Require regulatory agencies to consider trade-offs between different policy objectives, including efficiency and productivity enhancement, transparency, fairness, privacy, security, and resilience.

Responsibility

Specific Recommendations:

4) Build technical and human capacity in the ecosystem to enable effective risk management.
   - **a)** Invest in responsible AI research (including on addressing bias in AI) by increasing funding for national and multinational research centers.
   - **b)** Create a multistakeholder partnership (a Global Forum on AI — GFAI), building on prior examples like the Global Internet Forum to Counter Terrorism (GIFCT), to establish joint strategies, consistent best practices, and standardization in AI red-teaming, AI bug and bias bounties, and solutions like watermarking, metadata, and other techniques to combat AI-enabled mis/disinformation.

Drive international policy alignment, working with allies and partners to develop common approaches that reflect democratic values

1) Enable trusted data flows across national borders, ensuring that allies and partners do not restrict data flows between each other to ensure high-quality data is available for AI modeling.

2) Establish multinational AI research resources and offer compute and data to SMEs and academics in allied countries to work on AI.

3) Encourage adoption of common approaches to AI regulation and governance, as well as a common lexicon, based on the work of the OECD.

4) Use trade and economic agreements to support the development of consistent and non-discriminatory AI regulations, avoiding fragmented or differentiated treatment of AI applications based on geography.

5) Establish more effective mechanisms for information and best-practice sharing with allies and partners, and between governments and the private sector — for example for identifying actors engaging in economic espionage and attacks on AI systems.

Opportunity

Responsibility

Security

Security

We face a global competition for leadership on artificial intelligence. AI could lead to a new race for technological superiority as governments invest in digitalization and smart capabilities, using advanced analytics to streamline logistics and maintenance, improve intelligence and military performance, and reduce casualties and collateral damage.

AI will turbocharge influence operations, but also help track mis- and disinformation and identify manipulated media. And it will change the dynamic for attackers and defenders in cyberspace, enabling more sophisticated attacks by a wider range of cyber actors, including non-state actors, but also driving a new generation of cyber defenses through advanced security operations and threat intelligence.

Our challenge is to maximize the potential benefits of AI for global security and stability while preventing threat actors from exploiting this technology for malicious purposes. Governments must simultaneously invest in R&D and accelerate public and private AI adoption while controlling the proliferation of tools that could be abused by malicious actors.

Progress in this space will require cooperation. The comparative advantage of global democracies is that we can work together, not at cross purposes, on the development of new technology.

Security

Specific Recommendations:

Safeguard international security interests in advanced technologies

1) Develop optimal "next-generation" trade control policies for specific applications of AI-powered software that are deemed security risks, and on specific entities that provide support to AI-related research and development in ways that could threaten global security.

2) Expand international partnerships and public-private forums to share information on AI security vulnerabilities and issues between governments and the private sector.

3) Explore ways to identify and block disinformation campaigns, including interference in elections, where malicious actors use generative AI to generate or manipulate media (deepfakes/cheapfakes).

4) Establish joint AI research centers to advance AI research and adoption amongst like-minded countries.

Streamline government adoption of AI technologies

1) Reform government acquisition policies to take advantage of and foster world-leading AI. This includes investments in the most-needed, future-facing capabilities and expanding the aperture of companies that can deliver innovation.

2) Examine institutional and bureaucratic barriers that prevent governments from breaking down data silos and adopt best-in-class data governance to harness the full power of AI.

3) Capitalize on data insights through human-machine teaming, building nimble teams with the skills to quickly build/adapt/leverage AI systems which no longer require computer science degrees so that these teams can — in hours or days — address problems like finding a ship lost at sea or responding to an active threat event.

Research safety and alignment implications of advanced AI models

1) Convene multistakeholder fora and support the development of institutions with experts from academia, civil society, industry, and governments to conduct joint research on AI safety.

2) Consider thresholds (e.g., size, capability) at which AI systems may present novel risks that require additional oversight.

3) Develop strategies to align increasingly sophisticated and complex AI with human values and intended outcomes.

Conclusion

Getting these balances right will require a collaborative effort that brings multiple perspectives — what Thomas Friedman has called “complex adaptive coalitions” — to develop shared standards, protocols, and institutions of governance.

Thoughtful approaches and new ideas from across the AI ecosystem will help us navigate the transition, find collective solutions, and maximize AI’s amazing potential. The good news is that societies are already focused on the risks and benefits, and working on solutions. Governments can play an important role in convening stakeholders and shaping a global approach to move that work forward. And companies on the forefront of this work will need to lean in to demonstrate openness and cooperation to meet these goals.