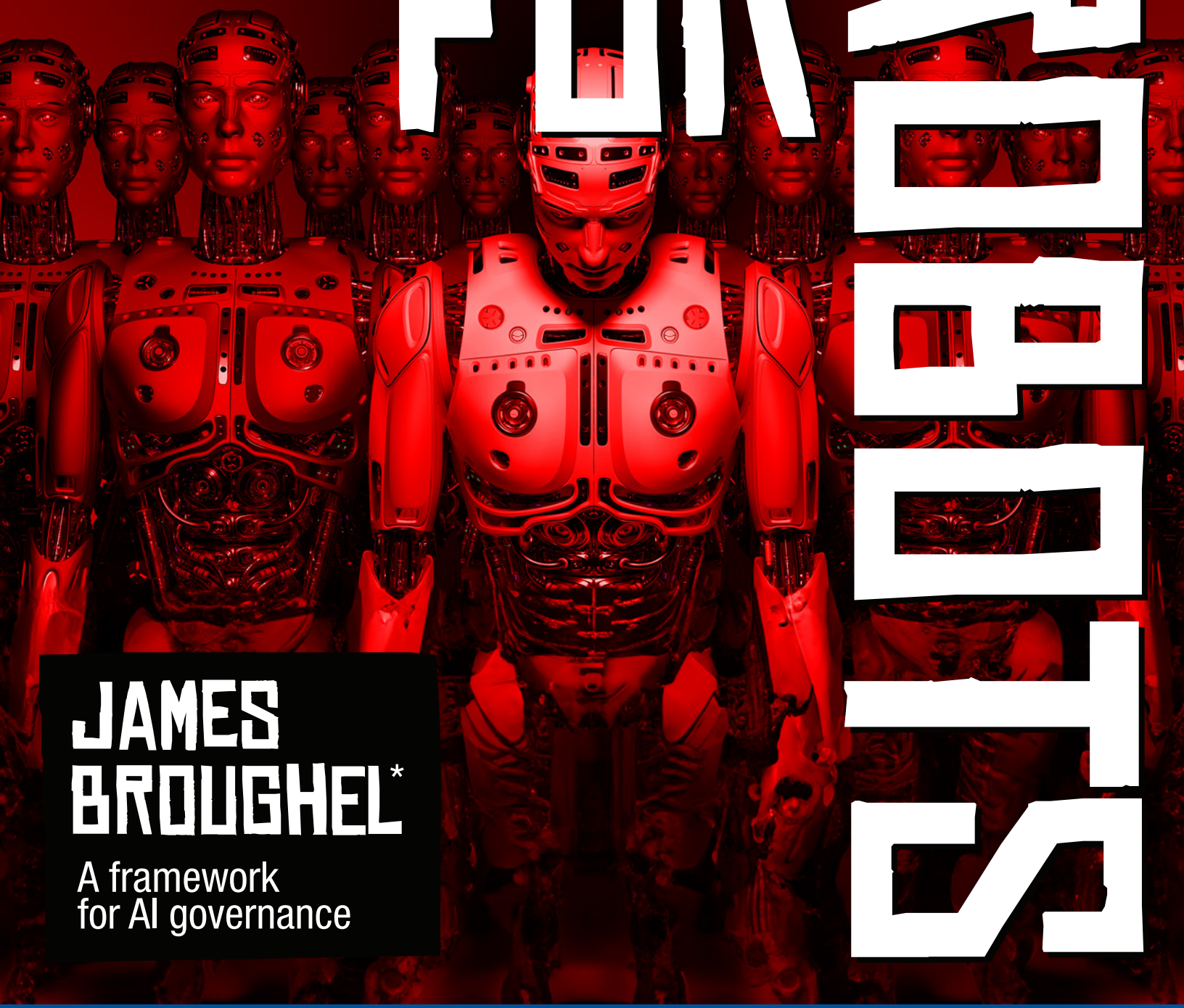
COMPETITIVE ENTERPRISE INSTITUTE

RULES FOR ROBOTS



JAMES BROUGHEL*

A framework for AI governance

Contents

*Dr. James Broughel is a Senior Fellow at the Competitive Enterprise Institute. Dr. Broughel is an accomplished economist whose expertise lies in regulatory institutions and the impact of regulations on economic growth. He is author of the book *Regulation and Economic Growth: Applying Economic Theory to Public Policy.*

Introduction

With the rise of artificial intelligence (AI) and machine learning technologies in recent years, we have seen increasing calls for regulation to mitigate the potential threats these technologies pose. High-profile figures, such as Tesla CEO Elon Musk and Apple co-founder Steve Wozniak, have issued warnings about a hypothetical future in which superintelligent AI surpasses human intelligence, leading to existential risks for humanity.[1] These "AI doomsday" prophecies have further fueled fears about the implications of unfettered AI development, sparking debates about the governance of AI technologies.

Even as the loudest voices on the internet have tended to focus on existential risks from AI, there are more immediate and mundane risks posed by AI, arising in the context of data security, privacy, discrimination, job loss, and other areas. Meanwhile, the benefits of AI development are immense. AI has the potential to redefine our daily interactions and experiences in both the digital and the physical realms, reshaping sectors from healthcare to education to finance.

Within this context, leading voices from the technology industry have contributed to the call for regulation. Sam Altman, CEO of OpenAI, has consistently defended the AI industry against critics, while simultaneously advocating for regulatory oversight of the most potent AI tools.[2] Microsoft—a funder of OpenAI's successful ChatGPT chatbot—created a 5-point document outlining a potential approach to AI regulation.[3] Influential ideological movements in the technology industry, including the "effective altruism" and "longtermism" communities, who often pride themselves on their rationalist approach, have played a key role in bringing public attention to existential risks from AI.[4]

Their efforts appear to be making a difference, as evidenced by the heightened focus on AI in Washington, DC. Democratic Senate Majority Leader Chuck Schumer has outlined an approach to AI regulation called the SAFE Innovation Framework.[5] Sens. Josh Hawley and Richard Blumenthal have outlined a framework for a potential US AI Act.[6] Bipartisan legislation has been introduced in Congress to create an AI regulation commission.[7] Meanwhile, federal agencies are following suit by enacting guidelines for internal and external AI policies.[8]

The history of regulation in the United States is complex, however, and the technology industry's recent entry into these discussions is still in its nascent stages. While these new voices bring a fresh perspective, their relative inexperience in the realm of public policy may lead to pitfalls such as an overly hasty or inadequately thought-out regulatory approach.

This set of circumstances is creating a perfect storm for a potentially perilous approach to public policymaking that has been termed "Ready! Fire! Aim!" rule-making by former Environmental Protection Agency Administrator William Reilly.[9] This phrase encapsulates the dangers of hasty regulation–where solutions are proposed and implemented before the problem's nature and significance are fully understood, or even before establishing whether a problem exists at all.

---

1 Future of Life Institute, "Pause Giant AI Experiments: An Open Letter," March 22, 2023, https://futureoflife.org/open-letter/pause-giant-ai-experiments/.

2 Samuel Altman, Testimony before the Senate Judiciary Committee, Subcommittee on Technology, Commerce, and the Internet, 118th Cong., 1st sess., May 16, 2023, https://www.judiciary.senate.gov/imo/media/doc/2023-05-16%20-%20Bio%20&%20Testimony%20-%20Altman.pdf.

3 Microsoft, "Governing AI: A Blueprint for the Future," May 25, 2023, https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RW14Gtw

4 Nitasha Tiku, "How Elite Schools like Stanford Became Fixated on the AI Apocalypse," *The Washington Post*, July 5, 2023, https://www.washingtonpost.com/technology/2023/07/05/ai-apocalypse-college-students/.

5 Senate Majority Leader Chuck Schumer, "Schumer's SAFE Innovation Framework," June 21, 2023, https://www.democrats.senate.gov/imo/media/doc/schumer_ai_framework.pdf.

6 Sens. Richard Blumenthal and Josh Hawley, " Bipartisan Framework for U.S. AI Act," September 7, 2023, https://www.blumenthal.senate.gov/imo/media/doc/09072023bipartisanaiframework.pdf.

7 Office of Congressman Ted Lieu, "Reps. Lieu, Buck, Eshoo and Sen. Schatz Introduce Bipartisan, Bicameral Bill to Create a National Commission on Artificial Intelligence," press release, US House of Representatives, June 20 2023, https://lieu.house.gov/media-center/press-releases/reps-lieu-buck-eshoo-and-sen-schatz-introduce-bipartisan-bicameral-bill.

8 Nihal Krishan, "White House Federal Agency AI Guidelines May Focus on Pilots and Info Sharing," FedScoop, May 5, 2023, https://fedscoop.com/white-house-federal-agency-ai-guidelines-may-focus-on-pilots-and-info-sharing/. AI General Services Administration, Centers of Excellence, "AI Guide for Government," accessed July 27, 2023, https://coe.gsa.gov/coe/ai-guide-for-government/print-all/index.html. US Government Accountability Office. "Artificial Intelligence: An Accountability Framework for Federal Agencies and Other Entities," June 30, 2021, https://www.gao.gov/products/gao-21-519sp. National Telecommunications and Information Administration, "AI Accountability Policy Request for Comment," 88 *Federal Register* 22433, April 13, 2023, https://www.federalregister.gov/documents/2023/04/13/2023-07776/ai-accountability-policy-request-for-comment.

9 Fred L. Smith Jr., "Conclusion: The Role of Opportunity Costs in the Global Warming Debate," In: *The Costs of Kyoto: Climate Change Policy and its Implications*, edited by Jonathan H. Adler, 149-165. Washington, D.C.: Competitive Enterprise Institute, 1997. Available at https://cei.org/sites/default/files/Costs_of_Kyoto_Part4.pdf.

Already a number of ambitious proposals have been put forward by AI regulation advocates, including the establishment of a federal AI agency,[10] the licensing of advanced AI products,[11] and even the creation of international AI bodies akin to the International Atomic Energy Agency (IAEA).[12] These proposals often lack concrete details both about the problem they aim to solve as well as about the proposed mechanisms by which they are supposed to work. They also tend to lack an empirical basis grounded in scientific research. Instead, vague principles are held up as statements of priorities, without any evidentiary basis. As such it is difficult to judge the various proposals' merits. What is clear is that ready, fire, aim rulemaking is rapidly becoming the norm, possibly putting American interests at a competitive disadvantage in the global race to develop advanced AI.

This paper aims to introduce an evidence-based framework for evaluating potential AI regulatory measures. Its objective is to ground AI policy discussions in well agreed-upon principles of regulation, established over the last half century of policymaking. This will help ensure regulatory proposals are judged based on their factual merits and evidence rather than on abstract goals or values. To effectively regulate the AI industry, the United States needs to strike a balance that respects liberty and facilitates innovation while understanding that regulation will be necessary in some cases where current law is inadequate and substantial market failures are present. This necessitates an approach that is as smart, informed, and as rigorous as the technology it seeks to oversee.

A brief taxonomy of AI risks

A common definition of AI is that it refers to machines that operate at a level equal to or exceeding human capabilities at some task. AGI, meanwhile, refers to artificial general intelligence, which is a technology capable of meeting or exceeding human capabilities across a wide range of activities. These definitions are loose, and there remain some basic disagreements about definitions.[13] Some tend to equate AI with machine learning, for example, which involves computers learning and improving from experience without being explicitly programmed. Debates about definitions will be important given that whatever definition is settled upon for policy purposes could end up having large ramifications with respect to the scope of activities that fall under the AI legal regime.

The media narrative surrounding AGI often hinges on dystopian, end-of-the-world scenarios. While these images paint a dramatic picture, they do not necessarily capture the most pertinent AI-related risks, nor do they reflect the areas most likely to be subjected to regulation. While existential risks are important and worth taking note of, AI technologies are already creating more immediate categories of risk, which include:

- **Data security:** AI systems, due to their capacity for processing vast amounts of data and making complex decisions, are increasingly utilized in critical infrastructure worldwide. This includes, but is not limited to, finance, health, transportation, and the defense sector, as well as everyday websites like social media sites that collect user data. These data centers present attractive targets for cyber threats, potentially creating unprecedented vulnerabilities in critical systems.

[10] Blumenthal and Hawley, "Bipartisan Framework." Tom Wheeler, "Artificial Intelligence Is Another Reason for a New Digital Agency," Brookings Institution *Commentary,* April 28, 2023. https://www.brookings.edu/articles/artificial-intelligence-is-another-reason-for-a-new-digital-agency/.

[11] Christopher Hutton, "Sudden Push to Require Licenses for Advanced AI Sparks Major Debate," *Washington Examiner,* May 16, 2023, https://www.washingtonexaminer.com/policy/technology/congress-ai-proposal-regulatory-agency. Markus Anderljung, Joslyn Barnhart, Anton Korinek, Jade Leung, Cullen O'Keefe, Jess Whittlestone, Shahar Avin, et al, "Frontier AI Regulation: Managing Emerging Risks to Public Safety," arXiv.org, July 6, 2023, https://arxiv.org/abs/2307.03718.

[12] Jon Gambrell, "OpenAI CEO Suggests International Agency like UN's Nuclear Watchdog Could Oversee AI," Associated Press, June 6, 2023, https://apnews.com/article/open-ai-sam-altman-emirates-10b15d748212be7dc5d09eabd642ff39.

[13] Adam Thierer, "Artificial Intelligence Primer: Definitions, Benefits & Policy Challenges," Medium, December 2, 2022 [Version 2.4 — June 2023], https://medium.com/@AdamThierer/artificial-intelligence-primer-definitions-benefits-policy-challenges-4c20a1fcf465.

- **Privacy:** With the advent of big data and advanced machine learning algorithms, AI's capability to collect, analyze, and generate insights from vast amounts of personal and sensitive data is becoming a significant privacy concern. Unauthorized access or misuse of such data can lead to violations of individual privacy rights and expose sensitive information. There is also the potential for surveillance abuse, blackmail, and identity theft.
- **Mis/disinformation:** The potential misuse of AI technologies, such as through deepfakes, text generators, and social media algorithms, can lead to the creation and spread of disinformation or misinformation on a large scale. This can distort public discourse, manipulate public opinion, undermine trust in institutions, and disrupt democratic processes and elections. The growing sophistication of these technologies makes it increasingly challenging to distinguish between genuine and manipulated content.
- **Bias and discrimination:** As AI systems are trained on human-generated datasets, they can reflect and perpetuate existing societal biases and discrimination. This can be particularly problematic in high-stake applications such as recruitment, law enforcement, and credit scoring, as well as advertising and sentencing decisions. AI's opacity and complexity can make it difficult to detect and rectify these biases, leading to unfair outcomes for marginalized groups.
- **Intellectual property:** AI has the capability to generate new content, such as music, text, artwork, and movies, raising questions about the ownership of IP created by non-human entities. AI algorithms are often trained on copyrighted works, thereby increasing the risk AIs might reproduce copyrighted material. This potentially poses a threat to artists and content creators who rely on IP protections for their livelihood.[14] Clear legal boundaries will likely need to be established with respect to who owns AI-generated content, as well as who is responsible for AI-associated liability.
- **Health:** In healthcare, AI has the potential to revolutionize diagnostics, patient care, drug discovery, and public health monitoring. However, inaccurate or biased AI algorithms can lead to misdiagnosis, inappropriate treatment recommendations, or health disparities, with grave consequences for patient safety. The potential misuse of health-related data raises additional privacy and security concerns.
- **Education:** AI promises transformative benefits in education, especially through personalized tutoring tailored to individual learning paces and styles.[15] However, students might also use AI for cheating. Increased dependence on AI can reduce social interactions, risking the loss of vital human connections in the learning process. Relying on AI for evaluations might miss nuances in human intelligence and creativity.
- **Transportation:** While AI-powered autonomous vehicles promise to enhance efficiency and safety on the roads, they also present novel challenges. These include safety risks associated with technological failures or cyber threats, ethical dilemmas related to decision-making in emergency situations, and legal and insurance implications related to liability.[16]

[14] These issues are already being actively debated in Hollywood, including who has control over an actor's personal image and whether screenwriters can be replaced by computers. Lucas Shaw, "AI in Hollywood Has Gone From Contract Sticking Point to Existential Crisis," *Bloomberg Businessweek*, July 27, 2023, https://www.bloomberg.com/news/features/2023-07-27/ai-use-in-movies-and-tv-looms-over-writers-and-actors-strike.

[15] John Bailey, "AI in Education," *Education Next*, August 9, 2023, https://www.educationnext.org/a-i-in-education-leap-into-new-era-machine-intelligence-carries-risks-challenges-promises/.

[16] Marc Scribner, "Self-Driving Regulation: Pro-Market Policies Key to Automated Vehicle Innovation," Competitive Enterprise Institute, April 23, 2014, https://cei.org/wp-content/uploads/2014/04/Marc-Scribner-Self-Driving-Regulation.pdf.

- **Weapons systems:** The integration of AI into weaponry, including autonomous weapons and bioweapons, poses serious ethical and security risks. Lethal Autonomous Weapon Systems (LAWS) will change the nature of warfare, raising questions about accountability, control, and the potential for escalation. Automated decision-making systems, if not adequately secured, can be manipulated by malicious actors. Terrorists and other non-state actors could obtain AI-enabled weapons. The use of AI in the development or dissemination of bioweapons could exacerbate biological threats.
- **Government abuse:** The utilization of AI in law enforcement holds the potential to revolutionize crime detection and prevention. However, governmental use of facial recognition tools can pose threats to civil liberties, leading to unwarranted surveillance and undermining the democratic principles of society. AI systems, trained on historically biased data, may perpetuate or exacerbate racial or socio-economic prejudices, leading to unjust profiling and targeting of certain groups.
- **Employment disruption and inequality:** AI has the potential to significantly disrupt job markets by automating tasks previously performed by humans. As with any such innovation, new jobs will likely be created, but transitions can be challenging and potentially exacerbate socio-economic inequalities. There is also a concern that the benefits and profits from AI will be disproportionately concentrated, leading to increased economic disparity within or across countries.
- **Existential risk:** While the existential risks associated with AI are often the focus of speculative media and online discussions, the actual likelihood of a "Terminator scenario" remains unclear. Such risks involve situations where AI surpasses human intelligence, gaining the capacity to act independently and possibly in ways that are not aligned with human values and interests. While currently hypothetical, these longer-term risks warrant consideration due to their potential high-impact nature.

The breadth of risks associated with AI underscores the need for a multifaceted approach to AI governance. One-size-fits-all solutions are unlikely to be effective or practical given the diverse contexts within which AI operates. An appropriate governance approach for autonomous vehicles will not make sense for military drones or deceptive online content. This implies a need for careful, case-by-case evaluation of governance proposals, taking into consideration the unique risks and regulatory environment associated with each specific application of AI.

Many of the risks outlined above already fall within areas that are subjected to regulation, which lends itself to a categorization of AI technologies by those "born free" (emerging in previously unregulated domains) and "born captive" (arising in already regulated domains).[17] While many see the debate about existential risks as a distraction from more immediate-term risks,[18] existential risks nevertheless deserve to be evaluated carefully, and prioritized when the danger is shown to be of sufficient magnitude. Longer-term existential risks will likely require special and potentially more aggressive solutions compared to the near-term risks that are already becoming apparent in many areas.

It is also important not to forget one additional source of risk: governance challenges. There is a risk that regulatory and governance frameworks will themselves not keep pace with the rapid development and deployment of AI,[19] leaving government and regulators ill-equipped to handle the oncoming wave of technological change.

[17] Adam Thierer, *Evasive Entrepreneurs and the Future of Governance: How Innovation Improves Economies and Governments*, Washington, DC: Cato Institute, 2020. https://www.cato.org/books/evasive-entrepreneurs-future-governance.

[18] Editorial, "Stop Talking about Tomorrow's AI Doomsday When AI Poses Risks Today." *Nature,* June 27, 2023, https://www.nature.com/articles/d41586-023-02094-7. Jan Brauner and Alan Chan, "AI Poses Doomsday Risks—But That Doesn't Mean We Shouldn't Talk About Present Harms Too," *Time,* August 10, 2023, https://time.com/6303127/ai-future-danger-present-harms/.

[19] This is known as the pacing problem. See, Adam Thierer, "The Pacing Problem and the Future of Technology Regulation," Mercatus Center Expert Commentary, accessed July 27, 2023, https://www.mercatus.org/economic-insights/expert-commentary/pacing-problem-and-future-technology-regulation.

The need for a unified framework

While the drumbeat for AI regulation grows louder, the calls often lack specificity. Mere demands for "regulation" might bring attention to some issues, but they fail to provide actionable steps toward viable solutions. At the same time, even more specific proposals often fall short because they lack evidence of both tangible problems and efficacy of solutions.

Private entities are beginning to establish their own internal frameworks for managing AI-related risks. Tech giants such as Google and Microsoft have published statements articulating their commitment to AI alignment values.[20] Microsoft has a Responsible AI office.[21] Google has a Secure AI Framework.[22] OpenAI has taken the step of launching an "alignment" project, which relates to ensuring that AI systems act and have goals that align with human values. That project is called "Superalignment," and as part of the effort OpenAI is dedicating a substantial 20 percent of the company's computing (i.e. processing) power to AI safety and alignment research.[23] Associations of AI researchers from the private sector and academia have put forth proposals for AI safety.[24]

These self-regulatory efforts merit commendation, particularly in light of the fact that the nature of AI risks is still being worked out. Private firms often have potent incentives to resolve problems caused by their own operations, especially if these issues pose financial or reputational risks. This sort of "regulation by markets" tends to be more sensitive, faster and more finely-calibrated than the often inefficient, slow, and widely-cast regulation of governments.

Governments worldwide are also increasingly recognizing a need for structured AI governance. The US National Institute of Standards and Technology has published guidance on managing AI risk.[25] The publication followed an executive order from the Trump administration, calling for increased investment in AI and the development of a national strategy on AI.[26] As of this writing, the National Telecommunications and Information Administration in the Commerce Department is producing a report on AI policy development.[27] The Office of Management and Budget is working on guidance to federal agencies on responsible AI practices.[28] The UK is also actively pursuing AI regulation,[29] creating an AI safety task force and producing a white paper putting forth AI principles.[30] The European Parliament and Canada have both proposed draft laws for AI regulation.[31]

China is another important player. The country already has regulations on the books related to use of personal data, disinformation, and fomenting social unrest.[32] China is using advanced surveillance techniques, such as facial recognition, throughout parts of the country.[33] More generally, the Chinese Communist Party is looking to become a global leader in AI innovation and to gain a strategic technological, economic and military advantage in this space.

Meanwhile, back home in the US, the Office of Science and Technology Policy (OSTP) has put forth a Blueprint for an AI Bill of Rights,[34] and legislative

[20] Google AI, "Google AI Principles – Google AI," accessed July 27, 2023, https://ai.google/responsibility/principles/. Microsoft AI, "Microsoft Responsible AI," accessed July 27, 2023, https://www.microsoft.com/en-us/ai/responsible-ai.

[21] Microsoft AI, "Microsoft Responsible AI."

[22] Royal Hansen and Phil Venables, "Introducing Google's Secure AI Framework," Google Blog, June 8, 2023, https://blog.google/technology/safety-security/introducing-googles-secure-ai-framework/.

[23] Jan Leike and Ilya Sutskever, "Introducing Superalignment," OpenAI Blog, June 5, 2023, https://openai.com/blog/introducing-superalignment.

[24] Anderljung, Barnhart, et al., "Frontier AI Regulation."

[25] National Institute of Standards and Technology, "AI Risk Management Framework (AI RMF 1.0)," January 26, 2023, https://www.nist.gov/itl/ai-risk-management-framework.

[26] Executive Order No. 13,859 84 *Federal Register* 3967 (February 14, 2019).

[27] National Telecommunications and Information Administration, "AI Accountability Policy Request for Comment," 88 *Federal Register* 22433, April 13, 2023, https://www.federalregister.gov/documents/2023/04/13/2023-07776/ai-accountability-policy-request-for-comment.

[28] Krishan, "White House Federal Agency AI Guidelines."

[29] Astha Rajvanshi, "Rishi Sunak Wants the U.K. to Be a Key Player in Global AI Regulation." *Time*, June 14, 2023, https://time.com/6287253/uk-rishi-sunak-ai-regulation/.

[30] Department for Science, Innovation and Technology, "A pro-Innovation Approach to AI Regulation," UK Department of Science, Innovation, and Technology, March 2023, https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach.

[31] Adam Satariano, "E.U. Takes Major Step Toward Regulating A.I.," *The New York Times*, June 14, 2023, https://www.nytimes.com/2023/06/14/technology/europe-ai-regulation.html. Government of Canada, "The Artificial Intelligence and Data Act (AIDA) – Companion Document," accessed September 5, 2023, https://ised-isde.canada.ca/site/innovation-better-canada/en/artificial-intelligence-and-data-act-aida-companion-document.

[32] Matthew Hutson, "Rules to Keep AI in Check: Nations Carve Different Paths for Tech Regulation," *Nature,* August 8, 2023.

[33] Paul Scharre, *Four Battlegrounds: Power in the Age of Artificial Intelligence,* W. W. Norton & Company, 2023.

[34] The White House, "Blueprint for an AI Bill of Rights," October 4, 2022, https://www.whitehouse.gov/ostp/ai-bill-of-rights/.

initiatives are actively underway in Congress, including the proposed creation of an AI regulation commission, and priority-setting for AI regulation from the senate majority leader, Chuck Schumer. The think tank community is also abuzz producing policy papers on AI.

Although these developments constitute positive strides in the sense that there is growing awareness that AI technologies are going to require policy changes, these activities represent a variety of approaches. Some of the responses are more sensible than others. Some of these documents or proposals aim to govern internal activities of firms or governments, rather than outline a regulatory framework. Many of the values expressed in these documents are broadly agreeable and uncontroversial, such as the need for pre-deployment testing of systems, and assurances of humans being able to maintain control over AI systems. However, the devil can be in the details.

For example, Microsoft's blueprint document for governing AI calls for building "safety brakes" into systems and defining a class of high risk AI systems,[35] but the potential impacts of these suggestions—if they were to be implemented—remains unclear. Other documents, including the proposed "AI Bill of Rights" or vision document from Sen. Schumer, provide little detail about specific, concrete policies. Others have proposed grandiose ideas as varied as an "AI Manhattan Project,"[36] a 6-month pause on AI research,[37] or the creation of an isolated island for AI research.[38] Yet these proposals also lack specificity. Moreover, there is a clear tension forming between approaches focused on protecting human rights and those focused on identifying and mitigating the most potent risks.[39]

Despite there being many advocates for AI regulation, few proposals are backed by solid evidence or clearly-defined goals. The common thread tying these proposals together is their lack of specifics. To be fair, even the critics of such proposals can be light on providing details.[40] Thus, just as there is an important need for more specific policy proposals, there is also a need for a rational, objective, and evidence-based framework with which to evaluate the proposals. The next section aims to present such a comprehensive framework, in hopes of guiding a more thoughtful discussion around AI regulation. Debate should be grounded in concrete facts about the real world rather than vague statements of abstract goals or ideological principles.

Regulation 101 principles for AI governance

To evaluate the various policy responses being proposed for particular AI risks, one should be thinking about the sort of evidence needed to guide policymakers. Fortunately, US regulators have accrued substantial experience in these matters over the years, resulting in a well agreed-upon set of best practices in the regulatory policy arena. This section will leverage these best practices, drawing on various executive orders and government directives instituted over the past four decades.

Importantly, these principles need not be restricted to regulatory proposals. They apply equally to legislative proposals, as well proposals from international bodies. Given the novelty of AI, Congress is likely to take steps to expand federal authority over AI in coming years. When regulatory agencies use their significant authority to regulate artificial intelligence,[41] they are expected to follow a rational and evidence-based decision-making process. These principles should apply to legislative decision-making too, which has the potential to influence regulatory outcomes for years to come.

---

35 Microsoft, "Governing AI," p. 13.

36 Samuel Hammond, "We Need a Manhattan Project for AI Safety," *Politico Magazine*, May 8, 2023, https://www.politico.com/news/magazine/2023/05/08/manhattan-project-for-ai-safety-00095779.

37 Future of Life Institute, "Pause Giant AI Experiments."

38 Ian Hogarth, "We Must Slow down the Race to God-like AI," *Financial Times*, April 13, 2023, https://www.ft.com/content/03895dc4-a3b7-481e-95cc-336a524f2ac2.

39 Nihal Krishnan, "Experts Warn of 'Contradictions' in Biden Administration's Top AI Policy Documents," FedScoop, August 23, 2023, https://fedscoop.com/experts-warn-of-contradictions-in-biden-administrations-top-ai-policy-documents/.

40 Marc Andreessen, "Why AI Will Save the World," Andreessen Horowitz, June 6, 2023, https://a16z.com/2023/06/06/ai-will-save-the-world/. Jeremy Howard, "AI Safety and the Age of Dislightenment," Fast.Ai, July 10, 2023, https://www.fast.ai/posts/2023-11-07-dislightenment.html.

41 Adam Thierer, "The Many Ways Government Already Regulates Artificial Intelligence," Medium, July 11, 2023, https://medium.com/@AdamThierer/the-many-ways-government-already-regulates-artificial-intelligence-74941254ed8d.

The key documents informing this framework include Executive Order 12866,[42] signed by then-President Bill Clinton in 1993, and the 2003 version of Office of Management and Budget Circular A-4,[43] which provides a framework for preparing regulatory assessments. Generally, the policy making process, from initial conceptualization to eventual realization, can be broken down into a straightforward four-step process:

Step 1: Demonstrate a problem exists

Step 2: Define the desired outcome

Step 3: Enumerate alternative solutions

Step 4: Rank alternatives

Despite the seeming simplicity of this four-stage process, it is essential to remember the complexity beneath each step and the potential for missteps. This highlights the need for thorough evidence collection at every stage to aid in decision-making. Despite these principles being well agreed-upon, regulators inconsistently adhere to them.[44] Diligent adherence to this process helps ensure the desired outcomes are achieved while minimizing the possibility of unintended consequences. The next sections of this report will explore each of these four steps in-depth.

Step 1: Demonstrate a problem exists

The opportunity for interventions in markets to improve welfare arises from what economists call a "market failure." Simply put, a market failure is a scenario where consumer or business needs that could be met at an acceptable price remain unfulfilled, typically due to various transaction costs such as information gaps or collective action problems. Consequently, the analyst's role involves detecting these gaps where public needs are going unmet, and identifying potential pathways to enhance welfare in a cost-efficient manner.

Mere assertions about market failure do not suffice as valid justifications for expensive marketplace interventions. As an illustration, a regulator might claim that misinformation propagated by AIs negatively affected a domestic election. But is this truly the case? There's a critical difference between conjecture based on superficial reasoning, and hard evidence gathered from detailed market analysis. Cherry-picking a few isolated examples is not robust evidence to justify regulatory action either, especially at the federal level. The problem must be proven to be substantial and typically systemic, rather than merely anecdotal.

Identifying a problem is just the start. It is also necessary to understand the problem thoroughly before attempting to devise solutions. This involves unraveling the root cause of the problem. If a leaky roof is causing a wet floor, the root problem (the leaky roof) should be addressed, rather than continually mopping up the symptom (the wet floor).

Consider the problem of AI bias, where AI systems, such as a facial recognition system, misidentify individuals of a certain race at a higher rate. A surface-level solution might be to mandate that any mistakes be corrected. This is like the "mopping up the wet floor" approach. Auditors could manually review the outputs of the system, correct misidentifications, and thereby ensure fair outcomes. But this doesn't solve the underlying problem. It only deals with the symptoms, as the AI system will continue to make biased predictions. The root cause of the problem in this case may be that the AI was trained on a dataset that did not adequately represent all types of individuals. To truly solve the problem (or "fix the leaky roof"), we need to focus on improving the data used to train the AI system.

Very often regulations themselves are the root cause of "market failures," because they act as barriers to the very entrepreneurial activity that aims to satisfy unmet consumer and business wants. For instance, occupational licensing regimes create barriers to work in certain professions. These restrictions make an industry less competitive and raise prices for consumers. Licensing of AI products is similarly likely to reduce competition, especially from smaller and upstart firms, as well as from open source technologies. Restrictions that limit competition in this way will likely lead to "government failures," which are simply market inefficiencies that have government itself as their root cause.

[42] Executive Order No. 12,866 58 *Federal Register* 51735, October 4, 1993, https://www.archives.gov/files/federal-register/executive-orders/pdf/12866.pdf.

[43] Office of Management and Budget, "Circular A-4: Regulatory Analysis," Washington DC, 2003, https://obamawhitehouse.archives.gov/omb/circulars_a004_a-4/.

[44] Jerry Ellig, "Evaluating the Quality and Use of Regulatory Impact Analysis: The Mercatus Center's Regulatory Report Card, 2008–2013," Mercatus Center Working Paper, 2016, https://www.mercatus.org/students/research/working-papers/evaluating-quality-and-use-regulatory-impact-analysis.

Therefore, regulators must delineate the problem, trace it back to its root cause, and provide empirical evidence supporting their theories. Policymakers shouldn't forget that their own past policies could well be part of the problem. Evidence should be verifiable, meaning the data can be confirmed, and any uncertainties should be acknowledged and quantified to the extent possible.

To summarize, the process of pinpointing a market failure involves the following steps:

- Identify a market or government failure
- Provide evidence that the market or government failure is real
- Present proof that the problem is severe or systemic rather than anecdotal
- Propose a theory explaining the genesis of the problem
- Furnish evidence to support the theory's validity

Step 2: Define the desired outcome

Defining the parameters of success is the starting point when regulators set out to establish the goals they are aiming to accomplish. This task entails outlining clear, actionable, and measurable objectives that reflect the desired state of the world post-intervention. The desired goal should generally relate to outcomes resulting from the use of AI within a specific context, as opposed to pertaining to a the use of a preferred technology.[45] Whether it's a specific reduction in AI-induced misinformation or a certain level of enhanced algorithmic fairness, a vivid picture of what success looks like serves as a lighthouse guiding the entire regulatory process.

Datasets play an integral role in tracking progress toward these defined goals. Deciding on the specific datasets to be used, the methods of collection, and the timing and frequency of data gathering, are all important factors that should be spelled out up front. This may involve selecting relevant indicators like frequency of misinformation reports, survey results on perceived fairness, or concrete measurements of data privacy infractions. The selected data should allow for continuous tracking and assessment of the efficacy of the regulatory interventions. This is important both for regulators themselves as they work toward achieving their objectives, as well as for the public in holding regulators accountable for their actions.

The journey toward successful outcomes often isn't a straight path. Instead, some degree of failure or course-correction is usually to be expected, and need not be viewed as a problem, at least early on. Setting milestones that act as guideposts can indicate whether the regulatory action is leading toward the envisioned success or if strategic adjustments are necessary along the way. Much like using signposts on a long journey, milestones provide a sense of progress, confirm the right direction, and offer a chance to adjust if necessary.

In summary, the second step of the process involves defining success, deciding on the data to track that success, and recognizing interim outcomes that indicate whether the path being followed is the right one. By attending to these matters carefully, regulators will be better equipped to navigate the complexities of the real world and be more likely to achieve the desired outcomes.

Regulators will be more likely to achieve success when they:

- Establish a clearly-defined desired outcome
- Determine the specific data that will be utilized to monitor progress toward success
- Identify interim outcomes that can indicate whether the policy or program is progressing successfully or if adjustments are necessary

Step 3: Enumerate alternative solutions

The process of achieving regulatory objectives requires considering a comprehensive set of alternatives. A common pitfall of regulators is to conduct an analysis of just one alternative—the chosen policy alternative—when better options were available. Each proposed solution should be underpinned by a coherent theory explaining how the solution is expected to achieve the desired outcome, along with empirical evidence that supports the validity of the theory. There will always be an element of uncertainty in this process, which should be acknowledged and quantified to the extent possible.

[45] Ryan Nabil, "Developing a Flexible, Innovation-Focused U.S. Approach to AI Regulation," Submission to the Office of Science and Technology Policy, Executive Office of the President, July 7, 2023, https://www.ntu.org/foundation/detail/letter-to-the-white-house-the-need-for-a-flexible-and-innovative-ai-framework.

Thus, when enumerating regulatory alternatives, policy makers should:

- Create a diverse list of potential solutions to address the identified problem, avoiding the common mistake of considering only one policy option
- Base solutions on a theory explaining how they will achieve the desired outcome
- Provide empirical evidence to justify each theory
- Acknowledge uncertainty and, to the extent possible, quantify it at each step

In shaping an effective response to AI risks, policy makers should explore a broad spectrum of alternatives rather than a narrow or limited set of options. Alternatives should range from maintaining the status quo (a "business as usual" approach) to more proactive strategies. This "no action" approach also serves as a baseline. Costs and benefits, to the extent they are quantified in an economic analysis of alternatives, will be evaluated against this scenario. The baseline analysis also overlaps with the problem analysis discussed earlier. It is part of forecasting future trends to determine whether a problem in existence is significant, getting worse, staying the same, or resolving itself with time.

Another alternative option available to policymakers is regulatory streamlining, which involves reducing or eliminating government regulation over certain sectors of the economy. For instance, less regulation in areas of AI development might spur innovation and open the door for smaller, nimbler players to compete in the market, thereby eliminating inefficiencies such as those relating to a lack of competition. Streamlining could be particularly important in the context of encouraging open source technologies.

Private governance is yet another option, where industry bodies, professional organizations, or user communities drive self-regulation. An example could be an industry-wide AI Ethics Code that all member companies commit to follow. Soft law mechanisms, such as government guidelines from bodies like the National Institute of Standards and Technology,[46] voluntary standards (such as an agreement made by the Biden administration with AI industry leaders),[47] or outcomes of multi-stakeholder talks, also provide potential pathways to self-governance or more flexible government oversight.

Existing regulations may already provide a foundation to build upon, thereby negating the need for new regulations. For instance, existing privacy laws, anti-discrimination policies, or consumer protection regulations might be applicable to certain AI-related problems. Or it is easy to imagine existing regulation of automobiles being sufficient to govern many applications of autonomous vehicles. Nevertheless, there can be gaps in existing legal frameworks that might need to be addressed.

Once these options are thoroughly explored, it can make sense to consider new regulations, beginning with those that provide the most flexibility in allowing individuals and firms to identify the solutions most tailored to their unique circumstances. For example, informational measures, such as requiring companies to disclose certain aspects of their AI algorithms' code, can allow more flexibility than dictating what is in the code itself. Bear in mind that even disclosure policies can be burdensome if they force companies to undergo intrusive audits or reveal sensitive proprietary information. Similarly, market-oriented approaches leverage economic incentives to encourage desired outcomes. These might include tradable permits for data usage or computing power, tax breaks or grants offered to companies that adopt responsible AI practices, or penalties imposed for breaches of ethical or privacy guidelines. Market-based approaches encourage creativity in finding cost-effective solutions, but also can create inefficiencies if they aren't designed carefully.

Finally, performance standards that focus on the outcome, can be superior to rules specifying design standards. This helps ensure that certain technologies or firms are not privileged over others. In the context of AI for autonomous vehicles, for example, it might be the case that advanced object recognition algorithms are needed to identify vehicles, pedestrians, cyclists, traffic signs, or other objects on the road. Requiring a certain level of accuracy from these systems could be more desirable than mandating that specific technologies, like LiDAR, be used in the design of autonomous vehicles.

[46] National Institute of Standards and Technology, "AI Risk Management Framework."

[47] The White House, "FACT SHEET: Biden-Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI," July 21, 2023, https://www.whitehouse.gov/briefing-room/statements-releases/2023/07/21/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-leading-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/?s=09.

Analyzing alternatives for autonomous vehicle safety

Let's consider an example of alternatives selection in the context of autonomous vehicle safety. Policymakers could consider a wide range of alternative solutions to enhance safety, such as 1) no policy action; 2) encouraging the formation of a self-regulatory organization that sets and enforces safety standards among its members; 3) implementing mandatory safety testing before deployment; 4) encouraging the creation of voluntary safety testing guidelines; 5) enforcing strict maintenance and upgrade schedules for AVs; 6) setting standards for the minimum acceptable performance of AVs under various conditions; 7) developing laws that govern liability in the case of accidents involving AVs; 8) establishing a federal database to record and analyze AV incidents; or 9) encouraging industry to develop its own incident database.

In each case, a theory and evidence should be offered to support the policy option. In contrast, a narrow range of alternatives might focus only on option 3, mandatory pre-deployment testing. The government might consider different criteria for testing within this option, but this would still constitute a narrow set of options relative to what is available.

Regulatory flexibility can also be considered along a number of other dimensions, including based on factors like geographic region or firm size. We have seen companies like OpenAI call for capabilities thresholds for regulations,[48] for example based on computing power. Smaller entities may indeed bear more significant burdens from regulation,[49] suggesting the need for policy thresholds for small businesses. Registration or certification requirements are similarly less burdensome than licensing regimes. Stringency can also vary across alternatives, as can enforcement methods, compliance dates, or statutorily-defined options. Altogether, a broad exploration of alternatives ensures that regulation is not one-size-fits-all, but rather is a finely-tailored response to the specific AI risk being addressed.

Table 1: Best practices for selecting regulatory alternatives

Alternatives should be broad, rather than narrow, and should include:	Regulators can provide flexibility by considering:
• No action/business as usual • Regulatory streamlining • Private governance • Soft law (e.g. industry standards, multi-stakeholder agreements, voluntary agreements) • Reliance on existing regulations	• Information measures rather than regulation • Market-oriented approaches rather than direct controls • Performance standards rather than design standards • Registration or certifications requirements rather than licensing • Different requirements for different geographic regions • Different requirements for different-sized firms • Different degrees of stringency • Different enforcement methods • Different compliance dates • Different choices defined by statute

Source: Author's assessment; and Jerry Ellig and James Broughel, "Regulatory Alternatives: Best and Worse Practices," Policy Brief (Arlington, VA: Mercatus Center at George Mason University, 2012).

[48] Sam Altman, Greg Brockman, Ilya Sutskever, "Governance of Superintelligence," OpenAI Blog, May 22, 2023, https://openai.com/blog/governance-of-superintelligence.

[49] James Broughel, "The Disproportionate Burden of Federal Regulation on Small Businesses," Submission to the US Senate Committee on Small Business & Entrepreneurship, August 23, 2023, https://cei.org/congressional_testim/the-disproportionate-burden-of-federal-regulation-on-small-businesses/.

Step 4: Rank alternatives

To identify the optimal approach, it is necessary to evaluate and rank alternatives systematically. Three analytical methodologies recommended for this purpose are cost-benefit analysis, cost-effectiveness analysis, and risk-risk analysis. Each has its own advantages and shortcomings, so let's take each in turn.

Cost-benefit analysis is a tool with a long history in regulatory decision-making.[50] It involves a detailed quantification of the costs and benefits associated with a given policy action, where costs and benefits should be measured in dollars for comparison. While government cost-benefit analysis can be an ideologically-driven exercise that lacks a clear economic rationale, cost-benefit analysis, in theory, offers a pragmatic method by which to weigh the potential advantages and disadvantages of various policies. In addition to considering the aggregate costs and benefits of policy responses, it is also important to consider the distribution of costs and benefits, as some groups like small businesses can be disproportionately affected. Cost-benefit analysis can become challenging in the context of existential risks, as almost any cost can appear worth bearing in return for a small chance of saving humanity.

Cost-effectiveness analysis provides another approach, useful when it is agreed certain outcomes or endpoints should be prioritized over others. For example, if a policy aims to reduce instances of fraud, regulators might evaluate alternatives on the basis of the cost-per-expected-instance of fraud prevented. This method allows regulators to compare the relative expense of achieving the prioritized outcome across the various different policy alternatives. This approach can make sense when dealing with existential risks where the primary goal is to avoid a disastrous outcome, and the cost comparison between different alternatives becomes the primary concern.

Finally, risk-risk analysis can be employed,[51] which is a methodology that involves comparing the risks a policy intervention mitigates against risks increased by enacting the policy. For example, costly regulations that raise energy prices can have negative repercussions for health and mortality.[52] These health opportunity costs—which come in the form of increased risk—can partially or even fully negate the risk-reducing benefits of regulations. Risk-risk analysis enables the analyst to gauge the policy's overall impact on risks.

Each of these methodologies offers a unique perspective, allowing regulators to weigh different factors when deciding on the best regulatory approach for AI.[53] Regulators should also seek public input, in case there are important details regulators missed in their analysis. By using a combination of these tools, policymakers can gain a broader perspective on the overall impacts of various policy alternatives, thus enhancing the likelihood of implementing an effective solution.

In sum, alternatives should be ranked according to one or more of the following methods:

- Cost-benefit analysis
- Cost-effectiveness analysis
- Risk-risk analysis

AGI and existential risk: Lessons from Kyoto

As we saw in the last section, existential risks bring with them unique challenges, such as difficulty in conducting a cost-benefit calculation. This doesn't rule out other forms of economic analysis. Nor does it negate the relevance of the basic process of identifying a problem and enumerating alternative solutions. It does mean some additional thinking is necessary.

In grappling with the existential risks potentially posed by AGI, contemporary discourse has shown some uncanny parallels with the climate change discussions of the 1990s. Uncertain, speculative, and potentially even benign, the nature of AGI risk remains deeply contested. In many ways, we stand at a crossroads similar to that faced by society during the early global warming debates.

---

[50] Executive Order 12,291 46 *Federal Register* 13193, February 17, 1981. For an example of a CBA that can serve as a model, see James Broughel and Michael Kotrous, "The Benefits of Coronavirus Suppression: A Cost-Benefit Analysis of the Response to the First Wave of COVID-19 in the United States," *PLOS ONE* 16, no. 6, June 3, 2021, https://doi.org/10.1371/journal.pone.0252729.

[51] For an example of a risk assessment that can serve as a model, see James Broughel and Andrew Baxter, "A Mortality Risk Analysis for OSHA's COVID-19 Emergency Regulations," *Journal of Risk and Financial Management* 15, no. 10, October 21, 2022, https://doi.org/10.3390/jrfm15100481.

[52] James Broughel, "The Lethal Impact of Rising Energy Prices," Competitive Enterprise Institute OpenMarket Blog, June 26, 2023, https://cei.org/blog/the-lethal-impact-of-rising-energy-prices/.

[53] Interestingly, the conclusions of these analyses often converge, given the tendency for opportunity costs to rise over time.

In 1997, the Competitive Enterprise Institute published a book edited by Jonathan H. Adler, which explored the pros and cons of the United States signing on to the Kyoto Protocol, an international climate change agreement that was being debated at that time.[54] In the concluding chapter, Fred Smith, founder of CEI, provides a lens through which to view existential risks.[55] While the chapter was motivated by the question of what to do about global warming, it also provides a helpful way to evaluate the governance of AGI and existential risks more broadly. Smith's framework delineates three categories of questions pertaining to risk: scientific, economic, and political. By answering these questions, a rational approach to responding to existential risks becomes clearer.

Scientific evaluation

On the scientific front, the primary questions concerning climate change in the 1990s were whether global warming was occurring or not, and what role humans had in contributing to any warming. While today, scientists tend to agree the planet is warming, it was not that long ago that this was unclear, and even claims of global cooling were not uncommon.[56] The parallel question for AGI is whether artificial general intelligence is likely to ever become a reality. Some risks are already apparent and therefore incontrovertible. Generative AI, which is used to generate music, text, images or other forms of content, is already available and presenting tangible risks in certain areas. AGI risks, by contrast, which are usually assumed to drive risks of an existential character, are different. When it comes to an artificial intelligence that exceeds human level capabilities in most domains, a range of unanswered questions persist such as whether such technology is even achievable. Thus, the scientific feasibility of AGI technology, the timeframe within which it may become a reality, and the nature of the potential risks stemming from it are all matters of intense scientific debate.

Economic evaluation

Economic questions in Fred Smith's framework probe the magnitude of the problem at hand. Here, the significance of AGI risks might be categorized into three broad domains: existential risks, moderate risks, and benign AGI. Each of these outcomes warrants different policy responses. For moderate risks, existing regulatory frameworks may be adequate, with no necessity for groundbreaking new solutions, though in some cases new laws or agencies might make sense. In the case where AGI is benign, it is likely that few if any new policies or interventions are required, other than perhaps regulatory streamlining. For risks to be existential, it will have to be demonstrated that widespread and irreversible harm could occur, and furthermore that such harm is relatively likely under certain conditions. With existential risks, a more distinct and, in some cases radical, set of solutions might be required. These might range from bans on certain activities to strict licensing or audit regimes for advanced AI firms and technologies. Some have gone so far as to recommend global surveillance measures to combat existential risks from AGI.[57] Such radical proposals should never be entered into lightly, given the significant tradeoffs involved for civil liberties, and given that a global totalitarian regime itself constitutes a form of existential threat to civilization.[58] Still, when risks are of an existential character, extreme solutions can sometimes make sense.

Political evaluation

Political questions assess the extent to which policy can realistically solve the problem. Four outcomes are possible: an unachievable policy outcome, a policy outcome achievable but only at an unreasonable cost, a policy outcome achievable at a reasonable cost, and a policy not required because AI is benign (in essence, the desired outcome is free). What constitutes a reasonable level of cost is obviously up for debate. In the environmental context, Smith pointed out that even once everyone agrees climate changes poses a significant problem, regulatory solutions mustn't be so expensive they are impossible to implement.

[54] Jonathan H. Adler, ed. *The Costs of Kyoto: Climate Change Policy and its Implications*, Washington, D.C.: Competitive Enterprise Institute, 1997.

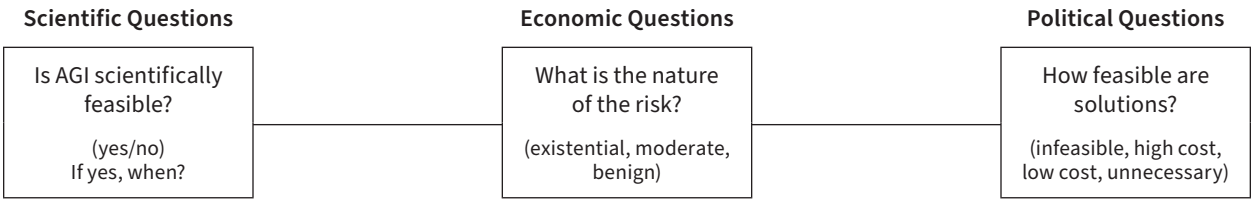
[55] Smith, "Opportunity Costs in Global Warming Debate."

[56] Peter Gwynne, "The Cooling World," *Newsweek*, April 28, 1975, "Science: Another Ice Age?" *Time*, June 24, 1974.

[57] Nick Bostrom, "The Vulnerable World Hypothesis," *Global Policy* 10, no. 4, 2019: 455-476, https://nickbostrom.com/papers/vulnerable.pdf.

[58] Maxwell Tabarrok, "Enlightenment Values in a Vulnerable World," *Effective Altruism Forum*, July 18, 2022, https://forum.effectivealtruism.org/posts/A4fMkKhBxio83NtBL/enlightenment-values-in-a-vulnerable-world.

Figure 1: Decision steps for assessing existential risks posed by artificial general intelligence



Much like we saw in the previous section, Smith's Kyoto Framework prompts us to recognize the reality of a problem, assess the potential harm of its effects, and evaluate the feasibility of potential solutions. Smith is also careful to warn us that we must be vigilant about the dangers of Type I and Type II errors. In other words, at each step there is a possibility we will treat benign risks as dangerous, high costs as low ones, and so on, and vice versa. Thus, there is always uncertainty about whether our assessment is correct and whether the information we have is reliable. Figure 1 illustrates the relevant questions policymakers should be asking when considering the nature of possible existential risks from AGI.

Insurance options: Prevention vs. resilience

Once the nature of a risk is characterized, only then should one move on to explore the various options available to policymakers to address the risk. Policy options for addressing risks fall into two broad categories, according to Smith: a prevention strategy that relies on regulation or "top-down" solutions, and a resilience strategy that relies on markets or "bottom-up" approaches. Prevention strategies tend to leave society poorer, while the resilience strategy tends to leave us richer, which enables society to combat a broader array of risks than it might be able to otherwise due to its greater wealth. Still, prevention should not be dismissed outright as a preventive approach is sometimes more effective.

When AGI risks are existential, this implies society may only have one chance at designing a successful policy response.[59] On the face of it, this makes it appear as though more stringent preventive policy responses are warranted, such as licensing requirements, bans, or the use of strict police powers or surveillance. However, even when an AGI risk is existential in nature, it is not inherently clear that prevention is always more effective than resilience, as preventive measures can constitute a kind of double-edged sword.

For instance, AI pauses or bans may prevent or slow down well-intentioned actors and firms from developing advanced AGI, freeing society from some risks these entities would unleash on society. At the same time, such policies might give malicious actors, both domestic or at the international level, the upper hand, thereby unleashing an even more dangerous set of risks.[60] These entities might include rogue states, such as North Korea, or non-state actors, like terrorist or hacker groups, focused on stealing advanced AI technology or developing their own technology for nefarious purposes.

Thus, there is always some probability that prevention will be superior to resilience, but there is also some probability the opposite is true—that prevention increases existential risk. The relative risks of the two approaches must be weighed against one another.

Similar principles apply in the case of moderate AGI risks. It is not inherently obvious whether prevention or resilience is superior. A careful weighing of costs and benefits and the corresponding risks of alternative approaches has to be conducted. However, bottom-up market solutions always have an advantage in that these approaches tend to promote wealth creation. In the case of benign AGI, market solutions will likely consistently outperform regulatory ones, since regulations would only succeed in making society poorer for few if any corresponding gains.

[59] Eliezer Yudkowsky, "Pausing AI Developments Isn't Enough. We Need to Shut It All Down," *Time*, March 29, 2023, https://time.com/6266923/ai-eliezer-yudkowsky-open-letter-not-enough/.

[60] James Broughel, "How Regulating AI Could Empower Bad Actors," *Forbes*, June 22, 2023, https://www.forbes.com/sites/jamesbroughel/2023/06/22/how-regulating-ai-could-empower-bad-actors/?sh=1ec5084c7ef5.

Table 2: Possible Scenarios for AGI Risks

The Science	The Economics	The Politics	Optimal Insurance Option
AGI is not scientifically feasible	N/A	N/A	Resilience
AGI is scientifically feasible	Risk is existential	Desired outcome is infeasible	Resilience
		Desired outcome is feasible at high cost	Prevention or resilience
		Desired outcome is feasible at low cost	Prevention or resilience
	Risk is moderate	Desired outcome is infeasible	Resilience
		Desired outcome is feasible at high cost	Prevention or resilience
		Desired outcome is feasible at low cost	Prevention or resilience
	Risk is benign	Desired outcome is freely available	Resilience

Source: Author's assessment.

Interestingly, it is never the case that prevention strategies unambiguously outperform resilience ones based on this framework. However, resilience strategies outperform prevention strategies in 4 out of the 8 scenarios considered here (see table 2). An unbiased approach therefore avoids presumption in favor of prevention over resilience. Yet, such presumptions are precisely what has been observed from advocates of AI regulation, who are inclined to presume from the outset that certain policy responses are warranted. If anything, table 2 demonstrates the resilience approach should be the default.

Further, there can be good reasons to uphold a presumption of liberty, which aligns more closely with the resilience insurance option. When policy options are technologically or politically infeasible, a presumption of liberty makes sense over preventive controls out of respect for citizens' rights. The same may be true when solutions are available at relatively equivalent cost from either the prevention or resilience strategy. All told, the resilience strategy has some significant advantages, but as noted, the best solution will depend on the context and the particular problem at hand, along with the feasibility of solutions.

To return to the Kyoto question, it should be noted that America did not end up joining the treaty, but did end up meeting roughly what would have been its emissions targets,[61] without many of the interventions the treaty envisaged.

[61] Anthony Watts, "USA Meets Kyoto Protocol Goal – without Ever Embracing It," Watts Up With That?, April 5, 2013, https://wattsupwiththat.com/2013/04/05/usa-meets-kyoto-protocol-without-ever-embracing-it/.

Understand we live in a world of tradeoffs and avoid Pascal's mugging

In addition to the lessons in preceding discussions, there are several insights that warrant further elaboration. First, regulatory interventions create risks of their own, which must be weighed against the risks of inaction. This is related to the notion of the "precautionary principle" in regulatory policy. The precautionary principle states that it is better to be safe than sorry. However, critics, such as Harvard law professor Cass Sunstein or think tank scholar Gregory Conko,[62] correctly recognize that the precautionary principle is incoherent and offers no particular guidance, as it overlooks the risks that may arise from taking action to prevent risks. The problems with the precautionary principle become evident when we consider the idea of opportunity costs. There is no truly safe option because we live in a world of tradeoffs. These tradeoffs must be balanced if we are to approach risks rationally. The obvious risk that arises in the case of AI regulation is the risk of stifling innovation that could improve welfare and even save lives.

An example of how the precautionary principle can go awry comes from regulation of genetically modified organisms in Europe, which delayed golden rice from being brought to market. Golden rice is a rice that has been genetically engineered to produce beta-carotene and address vitamin A deficiency. Some estimates suggest regulatory setbacks delayed a final product from coming to market by as much as ten years,[63] preventing mass production of a life-saving grain and costing countless lives in the developing world. The example highlights how the precautionary principle itself is not actually precautionary, in the sense that it too involves risks.

It is important to note, however, that in some cases a precautionary approach is sensible. When risks are fairly likely to occur and also devastating and irreversible, these are precisely the moments when the most stringent oversight mechanisms can make sense. For example, if it is likely that certain artificial intelligence software would lead to easily accessible methods of producing dangerous pathogens, then legal restrictions on the ability to access such software is reasonable. These kinds of scenarios are when advocates of strict restrictions on AI development stand on the firmest ground.

Another insight is that regulators must remain vigilant of the potential for regulatory capture, where special interest groups co-opt regulatory agencies to serve their own interests rather than those of the public.[64] There is a long history of regulatory capture in the technology space, especially at agencies like the Federal Communications Commission.[65] Powerful interest groups with access to regulators will try to shape rules for their own benefit at the expense of rivals. This problem is likely to be especially relevant when a handful of industry players dominate the market, as is the case with Big Tech today. Relatedly, there is a need to understand the tendency for regulatory accumulation. As new rules are added without repealing old ones, this accumulation can lead to unnecessary complexity, which will benefit larger incumbent firms since they are the best positioned to absorb the compliance costs.

The preservation of constitutional rights is yet another critical consideration for AI regulation. AI regulation will need to respect the ability of members of the public to express themselves, because the First Amendment's protections will have to be upheld. In this sense, AI debates are likely to be similar to those surrounding social media in recent years. That said, there are exceptions to the First Amendment, for example in the case of "incitement to imminent lawless action, true threats, fraud, defamation, and speech integral to criminal conduct."[66] Going further, there is likely to be an array of legal liability and copyright issues that will need to be worked out with AI. Like with many other issues, there will be few one-size-fits-all solutions here. These issues will have to be dealt with on a case-by-case basis and by a range of authorities ranging from courts to regulatory agencies to legislative bodies.

---

[62] Cass R. Sunstein, "Beyond the Precautionary Principle," *University of Pennsylvania Law Review* 151, no. 3, 2003: 1003-1058, Gregory Conko. "Throwing Precaution to the Wind: The Perils of the Precautionary Principle," Competitive Enterprise Institute, September 1, 2000, https://cei.org/publication/throwing-precaution-to-the-wind-the-perils-of-the-precautionary-principle/.

[63] Ed Regis, *Golden Rice: The Imperiled Birth of a GMO Superfood,* Johns Hopkins University Press, 2019.

[64] Adam Thierer, "Regulatory Capture: What the Experts Have Found," Technology Liberation Front, December 20, 2010, https://techliberation.com/2010/12/19/regulatory-capture-what-the-experts-have-found/.

[65] Adam Thierer and Brent Skorup, "A History of Cronyism and Capture in the Information Technology Sector," *Journal of Technology, Law & Policy* 18, no. 2, 2013: 131-196.

[66] The Foundation for Individual Rights and Expression, "Artificial Intelligence, Free Speech, and the First Amendment," accessed July 31, 2023, https://www.thefire.org/research-learn/artificial-intelligence-free-speech-and-first-amendment.

Additionally, the insights above regarding sound decision making should not only apply to the regulatory process but also inform the legislative process. Congress should avoid passing vague and sweeping AI safety legislation that leaves difficult questions about AI policy up to regulators to answer. Instead, elected representatives in Congress should be expected to do the hard work of identifying problems, proposing alternative solutions, and justifying their solutions based on the evidence at hand. Absent such careful deliberation, legislators are likely to make mistakes or be unduly influenced by special interests, which will constrain regulators in the future in undesirable ways. It is better to get AI legislation right from the outset, even if it takes longer, rather than rush hasty legislation out the door that will have unintended consequences for years to come.

Finally, as already noted, existential risks pose unique challenges. These are risks that, while perhaps have a small probability of occurrence, carry consequences so severe that they threaten the very existence of humanity. Unfortunately, this unique set of circumstances is sometimes used as a sledgehammer to shut down any debate about optimal policy solutions. This phenomenon is known as Pascal's mugging,[67] where it is argued any cost is worth bearing if it reduces even the most remote chance of the end of humanity. A straightforward cost-benefit analysis can be misleading in such cases, because it suggests one should expend vast resources to prevent low probability risks of extreme consequence. To avoid being bankrupted while chasing phantom risks, it is better to focus on other aspects of decision making aside from the cost-benefit calculation. This includes demonstrating the existence of the problem empirically, and proving that the proposed solution will actually achieve the desired result. In most cases, when such evidence is demanded, the ability of Pascal's mugging to act as a trump card in policymaking is defused. In cases where it is not defused, there may be legitimate reasons to devote significant resources toward risk mitigation.

How current AI proposals stack up

In this section, we explore some of the proposals that have been put forth in recent months regarding AI governance and evaluate the extent to which these proposals adhere to the policy evaluation criteria outlined above.

Some of the more encouraging steps taken have been industry efforts at self-regulation, government guidelines related to best practices, and voluntary agreements made between industry and government. As noted earlier, companies like OpenAI, Microsoft, and Google already have self-imposed principles related to safe and responsible AI development, and the National Institute of Standards and Technology has industry guidance as well. These represent laudable attempts at establishing best practices.

That said, many of the guidelines and proposals lack specificity. For example, Microsoft's responsible AI principles include "fairness," "reliability and safety," "privacy and security," "inclusiveness," "transparency," and "accountability."[68] Similar principles were included in Sen. Schumer's AI policy framework, which identified its central policy objectives as "security," "accountability," "foundations," "explain" and "innovation."[69] The White House's Blueprint for an AI Bill of Rights emphasizes the need for "safe and effective systems," "algorithmic discrimination protections," "data privacy," "notice and explanation," and "human alternatives,"[70] which are only slightly more specific. According to the website Axios, the Schumer plan may result in legislation that is likely to focus on the following policy priorities: "The identification of who trained the algorithm and who its intended audience is, the disclosure of its data source, an explanation for how it arrives at its responses, and transparent and strong ethical boundaries."[71] Meanwhile, Sens. Blumenthal and Hawley have called for the creation of an independent oversight body to oversee an AI licensing regime.[72] At the time of writing, there are some of the most specific proposals we have seen with respect to what US legislation of AI technologies might look like.

[67] Nick Bostrom, "Pascal's Mugging," *Analysis* 69, no. 3, 2009: 443–45, https://doi.org/10.1093/analys/anp062.

[68] Microsoft AI, "Responsible AI Principles and Approach," accessed July 31, 2023, https://www.microsoft.com/en-us/ai/principles-and-approach.

[69] Schumer, "Schumer's SAFE Innovation Framework."

[70] The White House, "Blueprint for an AI Bill of Rights."

[71] Andrew Solender and Ashley Gold, "Scoop: Schumer Lays Groundwork for Congress to Regulate AI," Axios, April 13, 2023, https://www.axios.com/2023/04/13/congress-regulate-ai-tech.

[72] Blumenthal and Hawley, "Bipartisan Framework."

Elsewhere, scholars in the think tank community have identified areas where regulation of government use of AI might make sense.[73] Enforcing disclosure requirements for AI-generated political advertising, controlling the use of predictive policing algorithms, and setting boundaries on law enforcement's use of facial recognition tools are a few ideas, as well as carefully managing AI integration into critical public infrastructure.[74]

In the private sector, it could well be that firms adopt many practices related to transparency and security on their own even without regulation. It should perhaps therefore not be surprising that, in July of 2023, the Biden administration announced a voluntary agreement made with a number of leading AI firms.[75] These companies agreed, among other things, to test AI systems before release, share certain data with the government, and conduct research on AI risks. Even here, however, many questions are left unanswered, such as what kind of testing will be conducted, what the standards will be for what is considered safe or unbiased AI, and what private user information might be collected and shared.

Tellingly, there is almost no empirical dimension to many of these discussions. Vague principles are held up as statements of priorities, but the extent to which problems currently exist, if they exist at all, is for the most part not addressed. Nor are solutions concrete enough to assess whether they are likely to solve potential problems. The unintended consequences likely to result from hasty regulation of AI on the basis of little or no information are obvious: the US is likely to fall behind as a global leader in the AI race. This could have important economic as well as national security implications, as our adversaries, most notably China, leap frog us in the innovation race.

State and local governments are also rushing to regulate AI. New York City has enacted an AI law regarding hiring discrimination practices.[76] That law requires companies that use AI software in hiring to notify job seekers in advance that an automated system will be used. Companies that rely on such automated hiring processes must also have independent auditors test the technology annually for bias in race, ethnicity, and gender. The law went into effect July 5, 2023 and covers applicants and employees who live in New York City, but could end up influencing practices outside of New York. California, meanwhile, is considering its own AI safety legislation.[77] The implication is there could end up being a patchwork of regulation across the various states if the federal government fails to act.

The European Parliament's draft AI Act establishes rules for AI based on technology and different levels of risk.[78] The law could end up requiring licensing of open-source AI alternatives and also the creation of sandboxes, which are "controlled environments established by public authorities, for AI testing before deployment."[79] In both the case of the proposed EU law and the New York City law, the costs of these provisions could well end up being substantial. For example, the proposed EU law could require summaries be produced of copyrighted materials used to train AI algorithms.[80] These provisions alone could end up being unworkable and cause leading technology companies to leave the EU rather than comply.[81]

Meanwhile, in the United States legislation to create an AI commission to study regulatory questions largely dodges difficult questions and leaves specifics up to the commission. A commission is not necessarily a bad idea in cases where regulation is needed and solutions are politically unworkable. In this case,

[73] Brent Skorup, "AI Alignment and Future Threats," Discourse Magazine, April 13, 2023, https://www.discoursemagazine.com/culture-and-society/2023/04/14/ai-alignment-and-future-threats/.

[74] Adam Thierer, "The Schumer AI Framework and the Future of Emerging Tech Policymaking," R Street Institute Commentary, June 27, 2023, https://www.rstreet.org/commentary/the-schumer-ai-framework-and-the-future-of-emerging-tech-policymaking/.

[75] The White House, "FACT SHEET: Biden-Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI," press release, July 21, 2023, https://www.whitehouse.gov/briefing-room/statements-releases/2023/07/21/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-leading-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/.

[76] James Barron, "How New York Is Regulating A.I." *The New York Times*, June 22, 2023, https://www.nytimes.com/2023/06/22/nyregion/ai-regulation-nyc.html.

[77] Billy Perrigo, "Exclusive: California Bill Proposes Regulating AI at State Level," *Time,* September 13, 2023, https://time.com/6313588/california-ai-regulation-bill/.

[78] European Parliament, "Artificial Intelligence Act," June, 2023, https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.html.

[79] Chanley T. Howell Kendall Spencer, "EU Paves the Way for U.S. in the Regulation of A.I." Foley & Lardner LLP blog, June 8, 2023, https://www.foley.com/en/insights/publications/2023/06/eu-paves-way-us-regulation-ai.

[80] Supantha Mukherjee, Foo Yun Chee, and Martin Coulter, "EU Proposes New Copyright Rules for Generative AI." Reuters, April 28, 2023, https://www.reuters.com/technology/eu-lawmakers-committee-reaches-deal-artificial-intelligence-act-2023-04-27/.

[81] Michelle Toh, "'Serious Concerns': Top Companies Raise Alarm over Europe's Proposed AI Law," CNN, June 30, 2023, https://www.cnn.com/2023/06/30/tech/eu-companies-risks-ai-law-intl-hnk/index.html.

however, it is not clear legislation is warranted yet. Agencies already have authority to regulate AI in many cases, and a commission "to regulate AI" will invariably end up looking for ways to justify its own existence, even if some recommended regulations are not actually needed. Rather than form a commission, it may make more sense to follow a wait and see approach that allows AI-related risks to emerge over time so they can be addressed as the need arises.

This may not work for existential risks, although presumably there will be early signs of problems before an AI apocalypse occurs. Unfortunately, most proposals to address such "x-risks" range from completely lacking any evidentiary basis to being so outlandish they are obviously unworkable, politically or economically.

For example, the recommendation to engage in a six-month pause, championed in an open letter by the Future of Life Institute,[82] aims to prevent an unrestrained AI arms race. But does it pass the first step of proving a substantial problem? The letter calls for limits on the growth of computational power, but assumes that more processing power equates to increased risk, when there is little evidence to support this assertion. The pause recommendation appears to be little more than a misguided attempt at applying the precautionary principle. Indeed, it is debatable how serious even the signers of the letter take its recommendations. For example, Elon Musk signed the letter but has since launched a new AI company.[83]

The recommendation for international treaties or oversight bodies akin to the International Atomic Energy Agency (IAEA) has also been made.[84] Yet the IAEA has only a mixed track record of success in controlling nuclear proliferation, as evidenced by countries that present continuing challenges, such as Iran and North Korea. Policymakers would need to evaluate whether the lessons and limitations of the IAEA's approach even make sense in an AI context.

Regarding new agencies or licensing bodies, something called for by Sens. Hawley and Blumenthal,[85] as well as Microsoft,[86] among others, economists widely recognize the potential for inefficiencies and unintended consequences, particularly when it comes to regulatory capture, with licensing regimes.[87] Occupational licensing is well known to create barriers to entry,[88] thereby reducing competition in an industry. In return, often little if any benefit in the form of quality or safety improvements is achieved.[89] Licensing regulation also tends to be regressive and at times discriminatory against foreigners and people of color.[90] An occupational licensing regime in the context of AI would likely hobble the nascent open source movement, which has the potential to democratize AI development and reduce the scope of influence of the big tech titans that currently dominate this market. Something similar can be said of an FDA-approval model for algorithms, which has also been suggested.[91] The primary aim of such proposals may be to choke off innovation from the open source sector as a means to stymie competition.[92]

Any proposal that involves firms requesting a permission slip from the government before being allowed to enter the market should be met with extreme skepticism unless advocates can present a clear path to success, offer flexibility to small businesses and open source technologies, and

[82] Future of Life Institute, "Pause Giant AI Experiments."

[83] Max Zahn, "Elon Musk Launches His Own AI Company to Compete with ChatGPT," ABC News, July 13, 2023, https://abcnews.go.com/Business/elon-musk-launches-ai-company-compete-chatgpt/story?id=101210078.

[84] Associated Press, "OpenAI's Sam Altman calls for an international agency like the UN's nuclear watchdog to oversee AI," Euronews, June 7, 2023, https://www.euronews.com/next/2023/06/07/openais-sam-altman-calls-for-an-international-agency-like-the-uns-nuclear-watchdog-to-over.

[85] Blumenthal and Hawley, "Bipartisan Framework."

[86] Dina Bass, "Microsoft Calls for a New U.S. Agency and Licensing for AI," *The Seattle Times*, May 25, 2023, https://www.seattletimes.com/business/microsoft/microsoft-calls-for-a-new-u-s-agency-and-licensing-for-ai/.

[87] George J. Stigler, "The Theory of Economic Regulation," *The Bell Journal of Economics and Management Science* 2, no. 1, 1971: 3–21, https://doi.org/10.2307/3003160.

[88] US Department of the Treasury, Council of Economic Advisers, and US Department of Labor, "Occupational Licensing: A Framework for Policymakers," July, 2015.

[89] Patrick A. McLaughlin, Matthew D. Mitchell, and Anne Philpot, "The Effects of Occupational Licensure on Competition, Consumers, and the Workforce," *Mercatus on Policy* (Arlington, VA: Mercatus Center at George Mason University, 2017).

[90] Tyler Boesch, Katherine Lim, and Ryan Nunn, "How Occupational Licensing Limits Access to Jobs among Workers of Color," Federal Reserve Bank of Minneapolis, March 11, 2022, https://www.minneapolisfed.org/article/2022/how-occupational-licensing-limits-access-to-jobs-among-workers-of-color.

[91] Ronald Bailey, "OpenAI Chief Sam Altman Wants an FDA-Style Agency for Artificial Intelligence," *Reason*, May 16, 2023, https://reason.com/2023/05/16/openai-chief-sam-altman-wants-an-fda-style-agency-for-artificial-intelligence/.

[92] Will Henshall, "The Heated Debate Over Who Should Control Access to AI," *Time*, August 25, 2023, https://time.com/6308604/meta-ai-access-open-source/.

establish easy options for repeal if something goes awry. For example, sunset provisions could be built-in that make licensing regimes hard to renew absent strong evidence the regime is succeeding. Yet nothing of the sort has been offered.

Something similar goes for audit proposals or for proposals for a secure "island" location where AI is studied under strict government oversight, which, in practice, might look something like the sandbox proposal in the draft EU AI law. Although regulatory sandboxes can be viewed as pro-innovation in markets where regulation is stringent,[93] such policies should generally not be looked to first when regulation is absent. Sandboxes should be seen as a solution to regulatory excess, not as a default regulatory regime, given the increased costs and limited options they entail for consumers. Benefits of sandboxes have to be substantial to justify these downsides.

Some AI regulatory proposals are frankly outlandish. A global mass surveillance state has been proposed by academic Nick Bostrom to monitor AI risks.[94] While this could theoretically help manage some risks, it opens a Pandora's box of civil liberties and privacy concerns and is also probably politically and economically infeasible. Proposals for a "Manhattan Project for AI" have gained some traction,[95] and some have called for large-scale nationalization of critical tech company resources as a means to control runaway AI.[96] Without a clearly defined objective, a theory of how to achieve success, or a mechanism for evaluating alternatives, such vague calls fall far short of our policy evaluation criteria. Overall, x-risk questions in particular could benefit from clearer thinking on the science, economics, and politics of the issues involved.

There are both pragmatic and extreme proposals being put forth in the area of AI regulation. What is noticeably absent from the discussions are solutions that are demonstrated to have an empirical basis, making it more likely they enhance welfare for citizens. This could change, of course. This is only just the beginning of the discussion, but many of the proposals today are based entirely on speculation. It is almost certainly the case that as AGI gets closer to being a reality, it will become more evident which risks are credible and which are merely the result of imaginations run amok.

Conclusions

AI, due to its multifaceted nature, demands a nuanced regulatory approach if we are to fully harness its potential while mitigating the risks. Considering the wide-ranging applications and possible impacts of AI, it becomes clear that any prospective regulations need to be precisely tailored to address specific problems rather than taking a broad-brush approach. Even in those instances where regulation is clearly warranted, an evidence-based approach should prevail.

Simultaneously, in certain sectors that have been born captive of regulation, there might be a stronger need to reassess the continued relevance and efficacy of existing regulations that have become obsolete in the face of changing technology. It might prove worthwhile to explore widescale regulatory streamlining as a potential pathway to stimulate innovation and growth across many sectors, especially as a means to open up AI development at smaller firms. Extreme proposals for new AI agencies or auditing bodies are likely to exacerbate industry concentration, bestowing advantages to the largest incumbents. Open source algorithms in particular are likely to be hurt by such policies.

In many areas, agencies already possess the regulatory authority needed to address AI-related issues, negating the need for new laws.[97] Indeed, an array of agencies have already begun to respond to the challenges and opportunities presented by AI.[98] Even when congressional action may be warranted, new laws should be subjected to rigorous scrutiny, as described in this paper.

[93] Ryan Nabil, "How the European Union Could Design a More Innovation-Friendly Approach to Artificial Intelligence," *Internationale Politik* (forthcoming).

[94] Bostrom, "Vulnerable World Hypothesis."

[95] Hammond, "We Need a Manhattan Project."

[96] Charles Jennings, "There's Only One Way to Control AI: Nationalization," *Politico Magazine,* August 20, 2023, https://www.politico.com/news/magazine/2023/08/20/its-time-to-nationalize-ai-00111862.

[97] Thierer, "Many Ways Government Already Regulates."

[98] Anthony E. DiResta, "The FTC Is Regulating AI: A Comprehensive Analysis," *Holland & Knight Alert*, July 25, 2023, https://www.hklaw.com/en/insights/publications/2023/07/the-ftc-is-regulating-ai-a-comprehensive-analysis.

Testing and impact assessments are also gaining traction as one of the preferred safety options before deploying advanced AI models.[99] Such assessments, whether voluntarily produced or mandated by law, and whether produced by government or industry, will need to be informed by empirical evidence as well as be backed up by economic and scientific theory. One area that requires significant urgent attention is academic research. There is a need for robust, well-constructed studies to inform public policy discussions. Currently, policy makers have very little information at their disposal with which to act. This sets the stage for exactly the kind of ready, fire, aim rulemaking that is likely to lead to poor decision making and bad results for citizens. There is evidence that the small literature that exists related to AI regulation is already getting off on the wrong foot.[100] Academics must strive to do better.

Policymakers must also remember that a default presumption of liberty and reliance on markets has historically proven successful when it comes to regulation of the internet and new technologies.[101] Therefore, the burden of proof should lie with those advocating for preemptive regulation. Too much of current discourse is driven by unsupported speculation driven by a fringe of the AI community focused on existential risks. While these risks cannot be ignored, doomsayers need to convincingly demonstrate that a problem exists, that regulation is superior to resilience, and that any associated public costs will bear fruit and are justified. If they can't, policymakers must not fall prey to scenarios where they are held hostage to mitigate largely nonexistent risks.

Currently, concrete evidence of existential risks posed by AI is scarce and AGI technology is still some way off from being a reality. More substantiated concerns exist in areas like misinformation, fraud, data security, and discrimination. In the grand scheme of things, these are not novel problems and many can be addressed using existing regulatory tools. Other threats relate to the potential for government abuse of AI systems and technology, and this should perhaps be where discussions of AI regulation start.

The AI community often prides itself on its rationalism. As such, it should welcome a rational, evidence-based approach to the development and evaluation of policy proposals. This paper has outlined such an approach, with the goal of informing robust, balanced, and effective AI governance.

---

[99] Anderljung, Barnhart, et al., "Frontier AI Regulation." Microsoft AI, "Microsoft Responsible AI Impact Assessment Template," June, 2022, https://blogs.microsoft.com/wp-content/uploads/prod/sites/5/2022/06/Microsoft-RAI-Impact-Assessment-Template.pdf.
S.3572 - 117th Congress (2021-2022): Algorithmic Accountability Act of 2022, https://www.congress.gov/bill/117th-congress/senate-bill/3572.

[100] James Broughel, "AI Economics Must Avoid The Ethical Mistakes Of Climate Economics," *Forbes*, July 20, 2023, https://www.forbes.com/sites/jamesbroughel/2023/07/20/ai-economics-must-avoid-the-ethical-mistakes-of-climate-economics/?sh=590369aaffdf.
Daron Acemoglu and Todd Lensman, "Regulating Transformative Technologies," working paper, July 6, 2023.
Charles I. Jones, "The AI Dilemma: Growth versus Existential Risk," working paper, 2023.

[101] Adam Thierer, "15 Years On, President Clinton's 5 Principles for Internet Policy Remain the Perfect Paradigm," *Forbes*, February 12, 2012, https://www.forbes.com/sites/adamthierer/2012/02/12/15-years-on-president-clintons-5-principles-for-internet-policy-remain-the-perfect-paradigm/?sh=fc4d77171703.

COMPETITIVE ENTERPRISE INSTITUTE

The Competitive Enterprise Institute promotes the institutions of liberty and works to remove government-created barriers to economic freedom, innovation, and prosperity through timely analysis, effective advocacy, inclusive coalition building, and strategic litigation.

COMPETITIVE ENTERPRISE INSTITUTE
1310 L Street NW, 7th Floor
Washington, DC 20005
202-331-1010