TECHNOLOGY

AI: How Safe Is Safe Enough?

Risk assessment of artificial intelligence using proportionality and foresight, not government overregulation, maximizes benefits and minimizes risks.

BY RANDALL MAYES

The "hype cycle" diagram represents the adoption of an emerging technology from product launch to mainstream embrace. It is typically S-shaped: First there is a rush of early adopters, then a lull, and finally—if the product is successful—embrace by a broad group of consumers.

The use of artificial intelligence (AI) has not followed this pattern. Instead, it has lurched through cycles of euphoria and retrenchment, with each new capability triggering fresh rounds of excitement, disappointment, and alarm.

From the unveiling of the earliest AI at a 1956 Dartmouth conference, through 2006, AI systems were rules-based rather than data-driven. Programmers explicitly encoded knowledge as logical rules—"If X, then Y"—that the machines followed, implying a belief that human thinking operated roughly the same way. Along the way, investment in AI development experienced two "AI winters" with decreased interest and funding, largely because researchers overestimated their understanding of how the human brain functions when they tried to replicate it.

Things changed in 2006 when Canadian researchers tweaked the backpropagation algorithm—how artificial neural networks learn from their mistakes—to enable artificial neural networks to do "deep learning." With enormous amounts of data fed into computer systems, deep learning can infer patterns on their own through statistical methods. This simulates cognitive neural networks in the human brain. Along with the introduction of cloud-based data storage, the collection of massive amounts of data, and data analytics, deep learning has led to immense advancement in AI and its expanding

Randall Mayes is a technology analyst/futurist who has taught AI and worked with AI training data for two decades.

embrace by the public.

The next major development came from DeepMind, the UK AI research lab whose systems defeated the world's top chess and Go players. This intensified what once seemed like purely theoretical discussions about AI risks and regulation. Historically, the vision for some AI pioneers was intelligence augmentation—a man–machine symbiosis—while others, including Alan Turing, envisioned thinking computers and automation of the human experience. As the machines moved closer to fully replicating human thinking, it was only a matter of time before activists and governments began voicing their concerns. The unveiling of OpenAI's ChatGPT in November 2022 and GPT-4 in March 2023 heightened those concerns and sparked even more public use.

Moving forward with AI development and regulation, governments have three options they can take. The challenge for risk assessment is to determine which approach maximizes benefits (international competitiveness, economic development, and quality of life) while minimizing risks to humans and critical infrastructure.

PETAR MAKSIMOVIC / ALAMY

OPTION 1: SHUT IT ALL DOWN

Many AIs are large language models (LLMs). They are fed massive amounts of data and then, when asked a question, construct an answer one word at a time by using the data to predict what word should follow the previous word. Note the word "construct"—the AI is not retrieving a stored answer but instead assembling an answer on the fly.

This functioning has troubled some technologists. Eliezer Yudkowsky, the director of the Machine Intelligence Research Institute, has advised to "Shut it all down. We are not ready and ... we are not taking risks serious enough." Yudkowsky (2023) outlines concerns about out-of-control AI, drawing on a thought experiment proposed by Oxford University philosopher Nick Bostrom that provides a scenario where a robot is programmed to maximize paperclip production (Bostrom 2014). If the robot runs out of metal and is not aligned with human values, what will prevent it from entering your office and taking metal objects, even if it had to do so by force?

DeepMind's chess and Go software uses reinforcement learning to influence behavior, loosely analogous to hormonal rewards for certain behaviors in humans. These AI drives are goal-seeking systems that can learn and improve themselves (Omohundro 2008). Some AI critics fear rational systems will exhibit sub-goals such as acquiring more power and resources, which contribute to the performance of their primary goals, and then the systems could decide that humans are in the way of their goals.

Autonomous AI agents are an application of LLMs used to increase efficiency and reduce operating costs in the business sector by performing tasks, making decisions, and interacting with their environment intelligently. Lu et al. (2024) found that such LLMs cannot learn independently or acquire reasoning skills. Rather, they require continual refining by humans, suggesting they pose no existential threat to humanity.

Currently, deep learning has limitations. It is narrow in scope and lacks reasoning capabilities. AI is undergoing a slow takeoff that was not intended. However, the trajectory could accelerate sharply if AI begins rewriting its own code and recursively improving its own capabilities—what Bostrom (2014) calls an intelligence explosion. By charting the pace of technological progress, AI pioneer and author Ray Kurzweil projects that the singularity—when AI acquires human-level intelligence—will take place in 2045 (Hemphill 2025). If AI agents do acquire the ability to develop their own motives, this would create an alignment or control problem.

Some would argue that just because something is unregulated doesn't make it unsafe, just as regulating something doesn't make it safe. Scientists developed genetic engineering of crops, and fatalities from that technology are non-existent even though—in the United States at least—regulation is light. The automobile and pharmaceutical industries are heavily regulated, and human fatalities from both are common and

tolerated, at least to some degree. Even with clinical trials, roughly 106,000 patients die annually in American hospitals from side effects of prescribed medication (Lazarou et al. 1998). Americans accept roughly 40,000 traffic fatalities annually, according to federal data.

In *Moral Machines: Teaching Robots Right from Wrong* (2008), Yale University ethicist Wendell Wallach and Colin Allen argue: "Would people have stopped the development of cars? Probably not. Most people believe that the advantages of cars outweigh the destructive potential." If we shut AI down, societies will not receive the benefits, and using this rationale would hold AI to a different standard than other technologies.

OPTION 2: FIT THE TECHNOLOGY PROBLEM TO THE BUREAUCRACY

Governments have the unique role of simultaneously promoting innovation through the development of emerging technologies while balancing economic development, international competitiveness, and the associated social effects. Carrying out this role requires considerable expertise.

Instead of creating a Department of AI in the executive branch, representatives of Google have argued for risk-based oversight rather than centralized control that would stifle the rapid development and deployment of new AI technology. They point out that the National Institute of Standards and Technology (NIST) has already launched the Artificial Intelligence Risk Management Framework to focus on different levels of risks in different industrial sectors of the economy. The NIST Management Framework is designed to augment existing risk practices and address new risks as they emerge (Bailey 2023). Any new agency will require resources, not just funding, but a new agency would likely be outpaced by the industry it seeks to regulate (Harris 2023).

Given the complexity of technology forecasting, it is not realistic to expect a centralized government body to adequately provide guidance for every emerging technology and related public policy issues. Virtual foresight groups can use existing personnel within current government planning organizations that collaborate with think tanks, academia, and the private sector to identify regulatory gaps and develop contingency plans (Fuerth & Faber 2012). As these technology fields develop, experts are called on to speak at the National Academy of Sciences, the President's Council on Bioethics, and Congress. Using this model not only informs lawmakers but also educates the public in the process.

Another disadvantage to the centralized regulatory approach is that it provides disproportionate risk assessment in terms of severity and probability. With a focus on regulations rather than addressing risks, it is an expensive and lengthy process to evaluate all the real and perceived risks, and as a result both human and financial resources are not utilized efficiently (Mahler 2022).

Undoing the OTA / In the legislative branch of government, lawmakers are typically not experts on the development of emerging technologies or their future implications. Instead, they rely on congressional staffers to advise them. The staffers typically rely on reports from management consulting firms and think tanks. Unfortunately, in this fast-paced environment, staffers have very little time to gain the background to adequately brief and advise lawmakers on complex science and technology issues.

Following the Cold War and the shift toward applied science, technology brought about several new opportunities and challenges. In the 1970s, the federal government became involved in long-term planning in areas other than defense. In 1972, Congress authorized the Office of Technology Assessment (OTA), and in 1975 the Congressional Research Service created a Futures Research Group to assist legislators with reports and research. While in operation, the OTA produced over 700 studies for policymakers.

It is an expensive and lengthy process to evaluate all the real and perceived risks, resulting in inefficient use of resources.

When the Reagan revolution descended on Washington in 1981, his administration pushed for more market-based solutions. As a result, in the early 1980s the Congressional Research Service eliminated the Futures Research Group. Then in 1995, Congress, headed by then-Speaker of the House Newt Gingrich, shut down the OTA as part of the Republican caucus's Contract with America, arguing the OTA was politically biased and that the private sector would do a better job.

Since the OTA was disbanded, congressional science and technology issues and responsibilities have increased while congressional support staff was reduced. Between 2002 and 2018, the Government Accountability Office (GAO) started conducting technology assessments for Congress and published 157 science and technology reports. Think tanks including Brookings and the Center for American Progress lobbied Congress to restore the OTA to advise the legislative branch. The idea escalated when Facebook's Mark Zuckerberg appeared before a congressional committee and the hearing revealed just how uninformed lawmakers were about the company's business practices.

Democratic lawmakers have attempted to revive the OTA. In 2018, a vote in the House of Representatives to restore

funding failed, so in 2019 the GAO launched the Center for Strategic Foresight to enhance its ability to identify, monitor, and analyze emerging issues and their implications before they occur. Currently, Congress has insufficient funding, a lack of expertise, and structural impediments that prevent it from assessing technology effectively, according to Miesen & Manley (2021).

A congressionally mandated study recommended that the legislative branch create an advisory position on technology and congressional offices hire more individuals with technology forecasting experience (Thomas 2019). As the OTA no longer exists, the GAO has assumed the trend analysis and technology forecasting role to inform Congress and other agency officials who are ultimately responsible for foresight and filling regulatory gaps.

Alignment problem / A strategy popularized by Russell (2020) and Borg et al. (2024) is "Don't make AI safe, make safe AI," meaning that AI systems should be engineered to behave in ways that are aligned with human values. This approach requires that AI applications meet certain safety guidelines to receive government approval—that is, it follows the precautionary principle.

This strategy assumes that machines can have human values. But, as robot ethicist Sean Welsh points out, AI manipulates symbols and numbers from pixels that represent human emotions and values, and this is not sentience (Welsh 2024).

Even if machines could have values, which human values should they adopt? There is a lack of standardization of human values at the government and individual levels. The three major AI hubs—China, the United States, and the European Union—have digital data and AI initiatives based on their unique hierarchy of values that includes international competitiveness, economic development, and social effects (quality of life and externalities).

At the individual level, situations arise where it is necessary for humans—or a self-driving AI—to make decisions regarding life and death. The infamous Trolley Problem is a thought experiment that involves making snap judgment decisions for a given situation and choosing who will die. For example, should a car swerve to strike and kill pedestrians on the left, right, or straight ahead, or do nothing, resulting in killing the passengers in the vehicle? Through the interactive website Moral Machine, MIT computer scientist Iyad Rahwan surveyed over two million participants from 233 countries and found that rules for morality are not universal (Awad et al. 2018).

It is important to remember that there is no evidence that AI could actually be malevolent. Robots in *Terminator* scenarios are based on Hollywood science fiction rather than our understanding of machine intelligence. Harvard psychology professor Steven Pinker thinks that those in the AI safety movement misunderstand the nature of intelligence. As he explains in an interview in *Quillette*:

> Just because a machine is super-intelligent doesn't mean it has a desire to dominate or annihilate us. Due to evolution, these two traits are bundled in humans, but motivation is independent of calculation in AI. An intelligent system will pursue whatever goals are programmed into it, and keeping itself in power or even alive indefinitely need not be among them. (Johnson 2023)

OPTION 3: FIT THE BUREAUCRACY TO THE TECHNOLOGY PROBLEM

In contrast to Yudkowsky, venture capitalist and Netscape co-founder Marc Andreessen argues:

> Our civilization is built on technology. We have a problem of poverty, so we invent technology to create abundance. We believe in risk, in leaps into the unknown. (Andreessen 2023)

Historically, scientists have not fully understood the risks of the most important technological innovations at the time of their invention. Even with the externalities that accompanied the past industrial revolutions, humanity has benefited from an increased life span, a higher standard of living, and is arguably better off.

Genetic engineering and synthetic biology were once viewed much like AI is today, and they provide successful risk assessment case studies for informing AI's development. Those technologies were developed using the "pro-actionary" approach, which employs proportionality with an equal emphasis on risks and benefits. Restrictive measures are employed only if the potential effects of an activity have both significant probability and severity. Manufacturers are held liable for the safety of their products and regulators must demonstrate that they are not squandering resources and delaying social benefits to address minimal gains in safety.

Genetic engineering / In 1971, President Richard Nixon initiated the "War on Cancer." As part of that research push, molecular biologists Stanley Cohen and Herbert Boyer began researching the genetics of viruses to better understand how they enter the human germline. They discovered that viruses use enzymes that could be used for reproducing and recombining DNA for beneficial purposes, including making life-saving drugs.

In the 1970s, genetic engineering was an emerging technology with an unknown safety record, and little was known about viruses. The top researchers in the field assembled at the Asilomar Conference in 1975 and agreed to a voluntary moratorium on the lab procedure until they better understood the risks. Once scientists learned how to carry out the procedure safely, the Food and Drug Administration, Environmental Protection Agency, and the Department of

Agriculture were given the responsibility of determining if the recombinant DNA products are safe for humans and the environment.

Today, genetic engineering is a common lab procedure used for manufacturing fruits and vegetables, erythropoietin for chemotherapy patients, and synthetic insulin that was formerly obtained from animals and cadavers (Miller 2023). We are now able to prevent tropical fruits infected by funguses from becoming extinct and help life forms survive in hostile conditions such as climate change.

In some cases, adhering to the precautionary principle is worse than the perceived risks of the activity (Sunstein 2002). Boycotting recombinant DNA products delays medical treatments and results in commercial losses. From June 1998 to August 2003, the European Union had a moratorium on imports of all genetically modified foods and feed products, which cost American farmers roughly $300 million annually. Twelve nations filed a lawsuit, and the World Trade Organization ruled that the moratorium was illegal.

The Asilomar model provides a successful framework for assessing risks. Since the 1970s, genetic engineering and recombinant DNA have not experienced problems inside or outside the lab.

Synthetic biology / Following the September 11, 2001, terrorist attacks, US President George W. Bush used the term "Axis of Evil" to describe foreign governments that sponsored terrorism and sought weapons of mass destruction. In addition to fighting several terrorist groups, the war on terror focused on three nations: Iraq and its anthrax and nerve gas, Iran and its nuclear program, and North Korea and its missiles and biological weapons.

Intelligence reports indicated that North Korea was in possession of 13 biological agents, including smallpox and anthrax, that were tested on political prisoners. If North Korea had delivered a nuclear bomb by missile or a biological weapon by aerosolization device fitted on a drone, the United States was not prepared. To counter those scenarios, Congress passed and Bush signed the 2004 Project BioShield Act, which authorized $5.6 billion for purchasing and stockpiling vaccines.

With advances in molecular biology, scientists transformed biology into an engineering discipline known as synthetic biology, enabling them to design living systems from standardized genetic "parts" (promoters, genes, regulatory sequences) that can be assembled in predictable ways. In 2010, President Barack Obama asked his Bioethics Commission to review the emerging field and identify appropriate ethical boundaries to maximize public benefits and minimize risks. This resulted in a report, "New Directions: The Ethics of Synthetic Biology and Emerging Technologies," that chose a course of "prudent vigilance" as opposed to new regulations or a moratorium. In 2012, the UK Synthetic Biology Roadmap Coordination Group produced a similar report. After weighing the options, these experts determined that the benefits of this technology outweigh the risks, and the best way to develop responses to bioterrorism is to perform synthetic biology research.

The traditional approach to making vaccines, developed in the 1940s, had scientists culture and grow a virus such as influenza in chicken eggs to stimulate the production of antibodies that were made into vaccines and distributed. This is a time-consuming process. Ideally, vaccines reach consumers before the influenza season begins, protecting against outbreak. When an outbreak of an unexpected influenza strain occurs, effective vaccines are not available until the virus has mutated or a pandemic has run its course.

Early this century, scientists argued that by developing synthetic biology, they could produce quick-response vaccines for deadly viruses and bioattacks. In 2010, the US nonprofit genomic research organization J. Craig Venter Institute formed Synthetic Genomics in collaboration with drugmaker GlaxoSmithKline and assembled a bank of synthetically constructed virus vaccines ready to go into production if the World Health Organization identified an outbreak strain. Today, we have stockpiles of vaccines in the event of future viral infections. With hindsight it is easy to recognize that the COVID-19 pandemic left the medical community in panic mode and the opportunity costs for not having a plan of action were far more expensive.

In *A Dangerous Master: How to Keep Technology from Slipping Beyond Our Control* (2015), Wallach argues that the President's Bioethics Commission blew an opportunity to regulate synthetic biology and suggested that more testing and delaying would achieve superior results. However, after weighing the options for developing synthetic biology, the United States and United Kingdom determined that the benefits outweigh the risks. Since the precautionary approach will not deter rogue actors, it is the best way to counteract bioterrorism.

AI: FROM EXISTENTIAL RISKS TO SOLVING COMPLEX PROBLEMS

In a pair of AI conferences in 2015 and 2017 organized by the Future of Life Institute and modeled after the 1975 Asilomar Conference, participants developed a set of principles for AI governance. Principle 21 states, "Risks posed by AI systems, especially catastrophic risks, must be subject to planning and mitigation efforts commensurate with their expected impact."

Although both the United States and the European Union are using the risk-based approach for economic efficiency rather than the precautionary principle for AI, proportionality does not necessarily manage risks. Rather, it ensures legislative proportionality (Mahler 2022). This makes it necessary to forecast probable scenarios and develop plans of action, especially for high-impact scenarios.

Quantum computing cyberattacks / Quantum computing will significantly increase the speed of computer processors. Some experts believe that mainstream adoption could take only a decade. This will require new methods of cybersecurity. In response, the US government has created post-quantum cryptographic standards for vendors and agencies that support critical infrastructure.

A probable future scenario is that rogue actors could use AI powered by quantum processing to initiate cyberattacks targeting critical infrastructure including dams, banking, utilities, and power plants. Attribution to the source must be established, and then two major policy questions must be considered: What constitutes a hostile act, and what is the appropriate response to a hostile act? A military AI arms race between the United States and China could cross a red line and will require diplomacy and may require law enforcement and military interventions to avoid World War III.

AI-inspired bioweapons / Researchers are using generative AI algorithms to develop new drugs using novel biological sequences purchased from commercial synthetic biology vendors. To manufacture a deadly toxin or pathogen, rogue actors can likewise order a genetic sequencer from a vendor. For security reasons, the vendors use screening software to compare incoming orders with known toxins or pathogens, and a close match will set off an alert. Using novel genetic sequences generated by AI, Microsoft has discovered vulnerabilities in the screening systems (Regalado 2025). Microsoft has alerted the government, which works closely with the vendors. The vendors have patched their systems, but some molecules can still escape detection.

Wild cards and black swans / Wild cards are low probability, high impact events, and black swans are unknown unknowns, which are impossible to forecast. A major goal for government forecasters is to reduce the public's susceptibility to such scenarios, which requires crisis decision making. Developing human–machine collaborations—including "digital twins" and quantum simulations in the metaverse that combine human reasoning and brute-force machine intelligence—could assist with these events that have eluded human forecasters.

An analysis of the available alternatives reveals that neither option provides risk-free benefits, which presents an AI dilemma. However, risk assessment of AI using proportionality and foresight maximizes benefits and minimizes risks.

CONCLUSION

Artificial intelligence presents humanity with a genuine dilemma: The risks of moving too fast are real, but so are the risks of moving too slowly. The history of transformative technologies—from the automobile to genetic engineering to synthetic biology—suggests that the question is never whether a technology carries risk, but whether its benefits—properly managed—outweigh those risks. For AI, the answer is almost certainly yes, provided that government keeps pace with development.

The three options examined here are not equally viable. A country shutting down AI development cedes the field to adversaries less scrupulous about safety and denies humanity the potential gains in medicine, productivity, and quality of life that AI promises. Fitting the technology to existing bureaucratic structures risks regulatory capture, resource misallocation, and the kind of disproportionate, precautionary overreach that cost American farmers hundreds of millions of dollars during the EU's genetically modified crop moratorium without making anyone safer.

The third path—fitting the bureaucracy to the technology problem—offers the most promising framework. It draws on tested models: the Asilomar process, which used voluntary expert consensus to manage recombinant DNA research; the synthetic biology experience, which demonstrated that prudent vigilance outperforms both paralysis and prohibition; and the risk-based governance frameworks now being developed by the NIST and others. What it requires, above all, is foresight infrastructure: the ability to identify emerging risks before they become crises, and to develop contingency plans for high-impact scenarios including quantum cyberattacks, AI-enabled bioweapons, and unpredictable black swan events.

The congressional capacity to perform this function has atrophied badly since the OTA was shuttered in 1995. Rebuilding that capacity—whether through a revived OTA, expanded GAO technology assessment, or virtual foresight networks drawing on academia, think tanks, and industry—is not optional. Lawmakers cannot regulate what they do not understand, and the gap between what Congress knows and what it needs to know is both troubling and widening.

The goal is not risk elimination. No technology worth having is risk-free. The goal is proportionality: matching the stringency of regulatory response to the severity and probability of harm, while preserving the innovation capacity that makes beneficial outcomes possible in the first place. Managed with that principle and paired with genuine foresight, AI is far more likely to extend and improve human life than to threaten it. *R*

READINGS

- Andreessen, Marc, 2023, "The Techno-Optimist Manifesto," Andreessen Horowitz, October 16.
- Awad, Edmond, Sohan Dsouza, Richard Kim, et al., 2018, "The Moral Machine Experiment," *Nature* 563(7729): 59–64.
- Bailey, Ronald, 2023, "Google Comes Out Against a Department of AI," *Reason*, June 16.
- Borg, Jana Schaich, Walter Sinnott-Armstrong, and Vincent Conitzer, 2024, *Moral AI and How We Get There*, Pelican/Penguin.
- Bostrom, Nick, 2014, *Superintelligence*, Oxford University Press.
- Fuerth, Leon S., and Evan M. H. Faber, 2012, *Anticipatory Governance:*

Practical Upgrades, National Defense University Press, October.

- Harris, Aubrey, 2023, "Will There Be a New Government Agency for AI?" *American Spectator*, July 28.
- Hemphill, Thomas A., 2025, "Augmented Humanity," *Regulation* 48(3): 57–59.
- Johnson, Matt, 2023, "There's Nothing Mystical About the Idea that Ideas Change History," *Quillette*, December 1.
- Lazarou, Jason, Bruce H. Pomeranz, and Paul N. Corey, 1998, "Incidence of Adverse Drug Reactions in Hospitalized Patients: A Meta-Analysis of Prospective Studies," *JAMA* 279(15): 1200–1205.
- Lu, Sheng, Irina Bigoulaeva, Rachneet Sachdeva, et al., 2024, "Are Emergent Abilities in Large Language Models Just In-Context Learning?" *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics* 1: Long Papers: 5098–5139.
- Mahler, Tobias, 2022, "Between Risk Management and Proportionality: The Risk-Based Approach in the EU's Artificial Intelligence Act Proposal," in Luca Colonna and Rolf Greenstein, eds., *Nordic Yearbook of Law and Informatics 2020–2021: Law in the Era of Artificial Intelligence*, Swedish Law and Informatics Research Institute, pp. 247–270.
- Miesen, Mike, and Laura Manley, 2021, "Building a 21st Century Congress: A Playbook for Modern Technology Assessment," Belfer Center for Science and International Affairs, Kennedy School of Government, Harvard University, June.
- Miller, Henry, 2023, "Synthetic Insulin and the Bureaucratic Mindset," *Regulation* 46(4): 2–3.
- Omohundro, Stephen M., 2008, "The Basic AI Drives," in Pei Wang, Ben Goertzel, and Stan Franklin, eds., *Artificial General Intelligence 2008: Proceedings of the First AGI Conference*, AGI 2008, March 1–3, 2008, University of Memphis, Memphis, TN. Vol. 171 of *Frontiers in Artificial Intelligence and Applications*, pp. 483–492. Amsterdam: IOS Press, 2008.
- Regalado, Antonio, 2025, "Microsoft Says AI Can Create 'Zero Day' Threats in Biology," *MIT Technology Review*, October 2.
- Russell, Stuart, 2020, *Human Compatible: Artificial Intelligence and the Problem of Control*, Penguin Books.
- Sunstein, Cass, 2002, "The Paralyzing Principle," *Regulation* 25(4): 32–37.
- Thomas, Will, 2019, "Study Complicates Campaign to Revive Congressional Technology Office," American Institute of Physics, November 22.
- Wallach, Wendell, and Colin Allen, 2008, *Moral Machines: Teaching Robots Right from Wrong*, Oxford University Press.
- Wallach, Wendell, 2015, *A Dangerous Master: How to Keep Technology from Slipping Beyond Our Control*, Basic Books.
- Welsh, Sean, 2024, "'Superintelligence,' Ten Years On," *Quillette*, July 2.
- Yudkowsky, Eliezer, 2023, "Pausing AI Developments Isn't Enough. We Need to Shut it All Down," *Time*, March 29.

Comment

BY JENNIFER HUDDLESTON

In just a few years, artificial intelligence (AI) has gone from a niche technology to an everyday tool. It is transforming everything from access to information to scientific discoveries in ways that previously seemed like science fiction. But many people are uneasy about AI, and policymakers are heeding their concerns. Much of the discourse around AI policy now focuses on potential risks and the presumption that government constraints are needed and that such policy can and should take a single nationwide form.

AI is a general-purpose technology, but often the motivations for regulation are more about specific applications. Still, there may be some areas where policy certainty is needed to allow the market and investment to flourish.

Unfortunately, AI has become the new villain in many policy narratives that focus only on its potential downsides rather than on the benefits it is already delivering. Technologies we are used to, like talk-to-text and autocomplete, are no longer thought of as AI, yet they are. Certain technologies that consumers do not directly see, such as some tools that catch fraud or help respond to natural disasters, are not often recognized as AI, but they are too. Regulatory restraints can and do affect such tools as much as the generative AI chatbots that have become synonymous with AI.

JENNIFER HUDDLESTON is a senior fellow in technology policy at the Cato Institute.

Pacing problem / In the preceding article, Randall Mayes concludes, "The risks of moving too fast are real, but so are the risks of moving too slowly." Moving quickly toward regulation is a more European, "precautionary" approach that is calibrated to the technology of today, limiting not only potential harm but also future innovations and options that do not fit neatly into what the regulations permit. Of course, this regulatory straitjacket affects more than just AI.

As with many technologies before it, the "pacing problem" has shown that technology often moves faster than policy does. In many cases, this supposed problem is a benefit, allowing innovators and consumers to access new technologies with minimal government or regulatory intervention and allowing technology to evolve to meet the demands of consumers rather than bureaucrats. However, in some cases—particularly where industry is already regulated, such as health or transportation—this can indeed create problems as regulation is unable to adapt to the speed of technology.

The pacing problem also means policymakers should reconsider existing regulation. Since policy change often moves slowly, static policy can hinder innovation and technologies that it was not designed to accommodate. For example, AI may, in some cases, provide better and safer methods that call

into question whether existing regulations are still needed. The response should not be to force new technology to comply with regulations that it renders outdated, but rather to consider whether such regulations are needed in the first place. If not carefully considered, outdated regulation could discourage important innovation that it was never intended to affect.

Focus on harm / As noted, because of the broad nature of AI, there is unlikely to be a single regulatory approach. Much of the debate over AI regulation and technology is better understood in terms of the particular applications of the technology. As a result, often the true question should not be how to regulate AI generally, but rather how to prevent a specific harm or respond to a specific concern given a potential application of the technology. In some cases, regulators have already affirmatively stated that using AI to engage in an unlawful act, such as fraud or discrimination, does not undo the burdens on the individual. In other cases, AI may raise concerns about civil rights and civil liberties that call for careful consideration of how to restrain potential government abuse. As a result, what regulatory tools—if any—are appropriate is better considered in light of the application than the technology more generally.

When it comes to AI and other general-purpose technologies, regulation should be reserved for clearly agreed-upon, irreversible, and catastrophic harms, and it should be proportionally responsive to the likelihood of such risks. Policy should focus not only on preventing worst-case scenarios but also on creating an ecosystem that allows for best-case scenarios that could never be foreseen. Similarly, more adaptive frameworks, such as voluntary consultations and standards, can evolve with risks and technological changes more readily than top-down approaches based only on the technology of the day. As Mayes notes: "The goal is not risk elimination. No technology worth having is risk-free. The goal is proportionality: matching the stringency of regulatory response to the severity and probability of harm, while preserving the innovation capacity that makes beneficial outcomes possible in the first place."

The risks of over-regulation should not be underestimated. AI is moving exceptionally fast, and policy seems to move ever more slowly. This risks regulation that could lock in the technologies of 2026, preventing future improvements that might naturally alleviate such problems, even when regulation does move quickly. Regulators' crystal balls are rarely able to predict the future that innovators can provide consumers if allowed. R