FOREIGN AFFAIRS

JANUARY 12, 2024

The Right Way to Regulate AI

Focus on Its Possibilities, Not Its Perils

ALONDRA NELSON

The Right Way to Regulate AI

Focus on Its Possibilities, Not Its Perils

ALONDRA NELSON

Artificial intelligence "is unlike anything Congress has dealt with before," Senate Majority Leader Charles Schumer said in June 2023. The pace at which AI developers are producing new systems—and those systems' potential to transform human life—means that the U.S. government should start "from scratch," he declared, when considering how to regulate and govern AI. Legislators, however, have defied his wishes. Following OpenAI's late 2022 unveiling of ChatGPT, proposals for how to encourage safe AI development have proliferated faster than new chatbots are being rushed to market. In March 2023, Democratic legislators proposed moratoriums on some uses of AI in surveillance. The next month, a group of bipartisan lawmakers floated a bill to prohibit autonomous AI systems from deploying nuclear weapons. In June, Schumer debuted his own AI agenda, and then in September, a bipartisan group of senators reintroduced a bill for AI governance promoting oversight, transparency, and data privacy.

The race to regulate is partly a response to the platitude that government may simply be too sluggish, too brittle, and too outmoded to keep up with

fleet-footed new technologies. Industry leaders frequently complain that government is too slow to respond productively to developments in Silicon Valley, using this line of argument to justify objections to putting guardrails around new technologies. Responding to this critique, some government proposals encourage expeditious AI development. But other bills try to rein in AI and protect against dangerous use cases and incursions into citizens' privacy and freedoms: the Algorithmic Accountability Act that House Democrats proposed in September 2023, for instance, mandates risk assessments before technologies are deployed. Some proposals even seek to accelerate and put the brakes on AI development at the same time.

This commendable but chaotic policy entrepreneurship risks scattering government's focus and threatens to lead to a situation in which there is no clear governance of AI in the United States at all. It doesn't have to be this way. A tendency to slip behind the curve of technological innovation is not an inherent weakness of government. In fact, trying to outpace government regulation is the tech industry's deliberate strategy to circumvent oversight. Government has an irreplaceable role to play as a stabilizing force in AI development. Government does not have to be a drag on innovation: it can enable it, strategically stewarding science and technology investments to not only prevent harm but also enhance people's lives.

From its first days, U.S. President Joe Biden's administration has worked toward a more integrated technology policy agenda that addresses AI's widening uses, considering competition, privacy, and bias as well as how to safeguard democracy, expand economic opportunity, and mitigate an array of risks. But AI technology is changing rapidly, and much more must be done to quickly clarify the central goal of AI governance so that policymaking is not only reactive.

AI governance should reject choice architectures that cast the future as a rigid binary—between a vision of paradise or dystopia or between a false dilemma of pursuing efficiency or ensuring equity. Safety and innovation in AI are not mutually exclusive. Because new and emerging AI

technologies are so dynamic and used for so many purposes, however, they may elude conventional policy approaches. The United States does not need so many new AI policies. It needs a new kind of policymaking.

FALSE ANALOGY

To regulate AI, many policy advisers in the United States and beyond have first sought an analogy. Are AI systems more like a particle accelerator complex, a novel drug therapy, or nuclear power research? The hope is that identifying a parallel, even a loose one, can point to the existing governance strategy that should apply to AI, guiding current and future policy initiatives.

The economist Samuel Hammond, for instance, took inspiration from the massive twentieth-century U.S. effort to build and assess risks related to nuclear weapons. He has proposed a Manhattan Project for AI safety, a federal research project focused on the most cataclysmic risks potentially posed by artificial intelligence. The nonprofit AI Now Institute, meanwhile, has begun to examine the viability of a regulatory agency based on the U.S. Food and Drug Administration: an FDA-like regulator of AI would prioritize public safety by focusing on prerelease scrutiny and approval of AI systems as the U.S. government does with pharmaceuticals, medical devices, and the country's food supply.

Multilateral analogies have also been suggested. The German Research Center for Artificial Intelligence has advocated modeling AI governance on the European Organization for Nuclear Research (CERN), the intergovernmental body that oversees fundamental scientific research in particle physics. In May 2023, Sam Altman, Greg Brockman, and Ilya Sutskever—then co-leaders at OpenAI—recommended that an AI governance framework be modeled on the International Atomic Energy Agency; in this model, the United Nations would establish an international bureaucracy to develop safety standards and an inspection regime for the most advanced AI systems.

The absence of an internationally coordinated research infrastructure poses a significant challenge for AI governance. Yet even conventional multilateral paradigms predicated on nation-state membership are unlikely

To regulate AI, many policy advisers in the United States and beyond have first sought an analogy.

to produce an effective way to govern competitive, for-profit industry efforts. AI companies are already offering products to a global and diverse customer base, including public and private enterprises and everyday consumers. And none of these analogies, including the U.S. domestic ones, reflect the fact that the data that enable AI systems' development have already become a global economic and political force. Further, all these potential models end up neglecting some critical domains on which AI will likely have a transformative impact, including health care, education, agriculture, labor, and finance.

The problem with reaching for a twentieth-century analogy is that AI simply does not resemble a twentieth-century innovation. Unlike the telephone, computing hardware, microelectronics, or many pharmaceutical products—technologies and products that evolved over years or decades—many AI systems are dynamic and constantly change; unlike the outputs of particle physics research, they can be rapidly deployed for both legitimate consumer use and illicit applications nearly as soon as they are developed. Off-the-shelf, existing governance models will likely be inadequate to the challenge of governing AI. And reflexive gestures toward the past may foreclose opportunities to devise inventive policy approaches that do not merely react to present challenges but anticipate future ones.

DROP AN ANCHOR

Instead of reaching to twentieth-century regulatory frameworks for guidance, policymakers must start with a different first step: asking themselves why they wish to govern AI at all. Drawing back from the task of governing AI is not an option. The past decade's belated, disjointed, and ultimately woefully insufficient efforts to govern social media's use of algorithmic systems are a sobering example of the consequences of passively hoping that social benefits will trickle down as an emergent property of technological development. Political leaders cannot again buy

the myth—peddled by self-interested tech leaders and investors—that supporting innovation requires suspending government's regulatory duties.

Some of the most significant challenges the world faces in the twenty-first century have arisen from the failure to properly regulate automated systems. These systems collect our data and surveil our lives. The indiscriminate use of so-called predictive algorithms and decision-making tools in health care, criminal justice, and access to housing causes unfair treatment and exacerbates existing inequities. Deepfakes on social media platforms stoke social disorder by amplifying misinformation. Technologies that went undergoverned are now hastening democratic decline, intensifying insecurity, and eroding people's trust in institutions worldwide.

But when tackling AI governance, it is crucial for leaders to consider not only what specific threats they fear from AI but what type of society they want to build. The public debate over AI has already shown how frenzied speculation about catastrophic risks can overpower people's ability to imagine AI's potential benefits.

Policymakers must start with a different first step: asking themselves why they wish to govern AI at all.

Biden's overall approach to policymaking, however, illustrates how viewing policy as an opportunity to enrich society—not just as a way to react to immediate problems—brings needed focus to government interventions. Key to this approach has been an overarching perspective that sees science, research, and innovation as offering both a value proposition and a values proposition to the American public. The administration's signal early policy achievements leveraged targeted public funding, infrastructure investment, and technological innovation to strengthen economic opportunities and ensure American well-being.

The 2022 Inflation Reduction Act, for instance, was not designed to merely curb inflation: by encouraging the production and use of advanced batteries, solar power, electric vehicles, heat pumps, and other new building technologies, it also sought to help address the climate crisis and

advance environmental justice. The 2022 CHIPS and Science Act promoted the revival of U.S. innovation by backing the development of a new ecosystem of semiconductor researchers and manufacturers, incorporating new opportunities for neglected U.S. regions and communities.

Government investments in science and technology, in other words, have the potential to address economic inequality. Like building a stock portfolio, it will take time for some of these investments to yield their full benefits. But this lodestar liberalism—anchored in values—has allowed the administration to forge bipartisan support in an otherwise fractious political milieu.

FLEXIBLE BENEFITS

The Biden administration has begun to make moves to apply the same approach to AI. In October 2022, the White House released its Blueprint for an AI Bill of Rights, which was distilled from engagement with representatives of various sectors of American society, including industry, academia, and civil society. The blueprint advanced five propositions: AI systems should be safe and effective. The public should know that their data will remain private. The public should not be subjected to the use of biased algorithms. Consumers should receive notice when an AI system is in use and have the opportunity to consent to using it. And citizens should be able to loop in a human being when AI is used to make a consequential decision about their lives. The document identified specific practices to encode public benefits into policy instruments, including the auditing, assessment, "red teaming," and monitoring of AI systems on an ongoing basis.

The blueprint was important in part because it emphasized the idea that AI governance need not start entirely from scratch. It can emerge from the same fundamental vision of the public good that the country's founders articulated centuries ago. There is no society whose members will always share the same vision of a good future, but democratic societies are built on a basic agreement about the core values citizens cherish: in the

case of the United States, these include privacy, freedom, equality, and the rule of law.

These long-standing values can—and must—still guide AI governance. When it comes to technology, policymakers too often believe that their approaches are constrained by a product's novelty and must be subject to the views of expert creators. Lawmakers can become trapped in a false sense that specific new technologies always need specific new laws. Their instinct becomes to devise new governance paradigms for each new tech development.

Lawmakers can become trapped in a false sense that specific new technologies always need specific new laws.

This instinct is wrong. Throughout history, the United States has reinterpreted and expanded citizens' rights and liberties, but the understanding that such entitlements and freedoms exist has been enduring. If policymakers return to first principles such as those invoked in the AI Bill of Rights when governing AI, they may also recognize that many AI applications are already subject to existing regulatory oversight.

Anchoring AI governance to a vision of the public good could diminish regulatory confusion and competition, stemming the flow of the sometimes contradictory bills lawmakers are currently producing. If it did, that would free both lawmakers and regulatory agencies to think more creatively in the areas in which policy innovation is truly needed. AI does pose unprecedented challenges demanding policy innovation. Already, the Department of Commerce's National Institute of Standards and Technology (NIST) has embarked on a different kind of policymaking when it comes to AI.

With a constitutional mandate to "fix the standard of weights and measures," NIST determines the proper standards to measure such things as length and mass, temperature and time, light and electricity. In 2021, Congress directed NIST to develop voluntary frameworks, guidelines, and best practices to steer the development and deployment of trustworthy AI systems, including ways to test for bias in AI training data and use cases.

Following consultations with industry leaders, scientists, and the public, in January 2023, NIST released its first AI Risk Management Framework 1.0. The "1.0" was meaningful. Versioning—think of Windows 2.0, 3.0, and so on—has long been commonplace in the world of software development to patch bugs, refine operations, and add improved features.

It is much less common in the world of policymaking. But NIST's use of policy versioning will permit an agile approach to the development of standards for AI. NIST also accompanied its framework with a "playbook," a practical guide to the document that will be updated every six months with new resources and case studies. This kind of innovation could be applied to other agencies. A more agile way of reviewing standards and policies should become a more regular part of the government's work.

THE OLD BECOMES NEW

The AI Bill of Rights and the NIST AI Risk Management Framework became the foundations of Biden's sweeping October 2023 Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. Running at 111 pages, it mobilizes the executive branch to use existing guidelines, authorities, and laws, innovatively applied, to govern AI. This sweeping mandate gives many key actors homework: industry leaders must provide insight into the inner workings of their most powerful systems and watermark their products to help support information integrity. The order directed the U.S. Office of Management and Budget to issue guidance on the federal government's own use of AI, recognizing that the government possesses extraordinary power to shape markets and industry behavior by setting rules for the procurement of AI systems and demanding transparency from AI creators.

But more must be done. AI governance needs an international component. In 2023, the European Union advanced significant new laws on AI governance, and the United Kingdom is moving to address AI regulation with what it calls a "light touch." The African Union has a regional AI strategy, and Singapore has just released its second national AI strategy in four years.

There is a risk that the world at large will suffer from the same glut of competing proposals that bedevils AI governance in the United States. But there are existing multilateral mechanisms that can be used to help clarify international governance efforts: with the UN Charter and the Universal Declaration of Human Rights, UN members states have already agreed to shared core values that should also guide AI regulation.

Democratic leaders must understand that disrupting and outpacing the regulatory process is part of the tech industry's business model. Anchoring their policymaking process on fundamental democratic principles would give lawmakers and regulators a consistent benchmark against which to consider the impact of AI systems and focus attention on societal benefits, not just the hype cycle of a new product. If policymakers can congregate around a positive vision for governing AI, they will likely find that many components of regulating the technology can be done by agencies and bodies that already exist. But if countries do decide they need new agencies—such as the AI Safety Institutes now being established in the United States and the United Kingdom—they should be imagined as democratic institutions that prioritize accountability to citizens and incorporate public consultation.

Properly constructed, such agencies could be a part of a broader governance infrastructure that not only detects how AI can infringe on rights and livelihoods but also scouts out how AI can proactively enhance them—by making dangerous jobs less perilous, health care more effective, elections more reliable, education more accessible, and energy use more sustainable. Although AI systems are powerful, they remain tools made by humans, and their uses are not preordained. Their effects are not inevitable.

AI governance need not be a drag on innovation. Ask bankers if unregulated lending by a competitor is good for them. Simply put, the ballast provided by proactive governance offers stability but also provides a controlled range of motion. First, however, policymakers must acknowledge that governing AI effectively will be an exercise in returning to first principles, not just a technical and regulatory task.