

# The AI rules that Congress is considering, explained

**V** [vox.com/future-perfect/23775650/ai-regulation-openai-gpt-anthropic-midjourney-stable](https://www.vox.com/future-perfect/23775650/ai-regulation-openai-gpt-anthropic-midjourney-stable)

Dylan Matthews

August 1, 2023



*Simone Virgini for Vox*

Dylan Matthews is a senior correspondent and head writer for Vox's Future Perfect section and has worked at Vox since 2014. He is particularly interested in global health and pandemic prevention, anti-poverty efforts, economic policy and theory, and conflicts about the right way to do philanthropy.

**FUTURE  
PERFECT**

Finding the best ways to do good.

Part of The rise of artificial intelligence, explained

AI is getting seriously good. And the federal government is finally getting serious about AI.

The White House announced a suite of artificial intelligence policies in May. More recently, they brokered a number of voluntary safety commitments from leading AI companies in July. That included commitments to both internal and third-party testing of AI products to ensure they're secure against cyberattack and guard against misuse by bad actors.

Senate Majority Leader Chuck Schumer outlined his preferred approach to regulation in a June speech and promised prompt legislation, telling his audience, “many of you have spent months calling on us to act. I hear you loud and clear.” Independent regulators like the Federal Trade Commission have been going public to outline how they plan to approach the technology. A bipartisan group wants to ban the use of AI to make nuclear launch decisions, at the very minimum.

But “knowing you’re going to do something” and “knowing what that something is” are two different things. AI policy is still pretty virgin terrain in DC, and proposals from government leaders tend to be articulated with lots of jargon, usually involving invocations of broad ideas or requests for public input and additional study, rather than specific plans for action. Principles, rather than programming. Indeed, the US government’s record to date on AI has mostly involved vague calls for “continued United States leadership in artificial intelligence research and development” or “adoption of artificial intelligence technologies in the Federal Government,” which is fine, but not exactly concrete policy.

That said, we probably are going to see more specific action soon given the unprecedented degree of public attention and number of congressional hearings devoted to AI. AI companies themselves are actively working on self-regulation in the hope of setting the tone for regulation by others. That — plus the sheer importance of an emerging technology like AI — makes it worth digging a little deeper into what action in DC might involve.

You can break most of the ideas circulating into one of four rough categories:

- **Rules:** New regulations and laws for individuals and companies training AI models, building or selling chips used for AI training, and/or using AI models in their business
- **Institutions:** New government agencies or international organizations that can implement and enforce these new regulations and laws
- **Money:** Additional funding for research, either to expand AI capabilities or to ensure safety
- **People:** Expanded high-skilled immigration and increased education funding to build out a workforce that can build and control AI

## New rules

---

Making new rules for AI developers — whether in the form of voluntary standards, binding regulations from existing agencies, new laws passed by Congress, or international agreements binding several countries — is by far the most crowded space here, the most consequential, and the most contested.

On one end of the spectrum are techno libertarians who look warily on attempts by the government to mandate rules for AI, fearing that this could slow down progress or, worse, lead to regulatory capture where rules are written to benefit a small handful of currently dominant companies like OpenAI. The Electronic Frontier Foundation and the R Street Institute are probably the leading representatives of this perspective in DC.

Other stakeholders, though, want extensive new rulemaking and legislating on a variety of AI topics. Some, like Sens. Josh Hawley (R-MO) and Richard Blumenthal (D-CT), want sweeping changes to rules around liability, enabling citizens to sue AI companies or prosecutors to indict them if their products cause certain harms.

One category of proposals deals with how AI systems interface with existing rules around copyright, privacy, and bias based on race, gender, sexual orientation, and disability. Think of it more like AI ethics rather than AI safety.



Federal Trade Commission chair Lina Khan stands poised to be a leading figure in regulation concerning the near-term implications of AI.

*Michael M. Santiago/Getty Images*

**Copyright:** The US Copyright Office has issued rulings suggesting that most texts, images, and videos output by AI systems cannot be copyrighted as original works, as they were not created by a human. Meanwhile, large models like GPT-4 and Stable Diffusion rely on

massive training datasets that usually include copyrighted texts and images. This has prompted myriad lawsuits and provisions in the European Union's AI Act requiring model builders to “publish information on the use of training data protected under copyright law.” More regulations and laws from either US agencies or Congress could be forthcoming.

**Privacy:** Just as large AI companies have faced lawsuits for copyright violations in the construction of their models, so too have some plaintiffs argued that the mass web scraping necessary to collect the terabytes of data needed to train the models represents an invasion of privacy. The revelation in March of a since-patched data vulnerability that allowed ChatGPT users to access other users' chat histories, and even their payment information, raised further alarms. Italy even briefly banned the service over privacy concerns about the training data. (It's since been allowed back.) Policymakers have been focusing on similar issues in social media and online advertising for some time now, with common proposals including a full ban on using personal data to target ads, and FTC action to require “data minimization” in which websites can only collect data relevant to a narrow function of the site.

**Algorithmic bias:** In part because they draw upon datasets that inevitably reflect stereotypes and biases in humans' writing, legal decisions, photography, and more, AI systems have often exhibited biases with the potential to harm women, people of color, and other marginalized groups. The main congressional proposal on this topic is the Algorithmic Accountability Act, which would require companies to evaluate algorithmic systems they use — in other words, AI — for “bias, effectiveness and other factors,” and enlist the Federal Trade Commission to enforce the requirement. The FTC has said it will crack down using existing authority to prevent “the sale or use of — for example — racially biased algorithms”; what these enforcement actions might look like in practice is as yet unclear.

Another set of proposals views AI through a national security or extreme risk perspective, trying to prevent either more powerful rogue AIs that could elude human control or the misuse of AI systems by terrorist groups or hostile nation-states (particularly China) to develop weapons and other harmful products. A future rogue AI with sufficiently high capabilities, that humans cannot shut down or coerce into following a safe goal, would pose a high risk of harming humans, even if such harm is merely incidental to its ultimate goal. More immediately, sufficiently powerful AI models could gain superhuman abilities in hacking, enabling malign users to access sensitive data or even military equipment; they could also be employed to design and deploy pathogens more dangerous than anything nature has yet cooked up.

**Mandatory auditing, with fines against violators:** As with racial or gender bias, many proposals to deal with uncontrollable AIs or extreme misuse focus on evaluations and “red-teaming” (attempts to get models to exhibit dangerous behavior, with the aim of discovering weaknesses or flaws in the models) which could identify worrisome capabilities or behaviors by frontier AI models. A recent paper by 24 AI governance experts (including co-authors from



leading firms like Google DeepMind and OpenAI) argued that regulators should conduct risk assessments before release, specifically asking “1) which dangerous capabilities does or could the model possess, if any?, and (2) how controllable is the model?”

The authors call for AI firms to apply these risk assessments to themselves, with audits and red-teaming from third-party entities (like government regulators) to ensure the firms are following protocol. Regulators should be given regular access to documentation on how the models were trained and fine-tuned; in extreme cases, “significant administrative fines or civil penalties” from regulators for failing to follow best practices could be necessary. In less severe cases, regulators could “name and shame” violators.

In the nearer term, some in Congress, like Sens. Ted Budd (R-NC) and Ed Markey (D-MA), are pushing legislation to require the Department of Health and Human Services to conduct risk assessments of the biological dangers posed by AI and develop a strategy for preventing its use for bioweapons or artificial pandemics. These are fairly light requirements but might serve as a first step toward more binding regulation. Many biosecurity experts are worried that AIs capable of guiding amateurs through the process of creating deadly bioweapons will emerge soon, making this particular area very high-stakes.



Gary Marcus, left, and Sam Altman are among the prominent AI voices calling for a licensing regime.

*Andrew Caballero-Reynolds/AFP via Getty Images*

**Licensing requirements:** The attorney Andrew Tutt in 2017 proposed a more far-reaching approach than simply mandating risk evaluations, one instead modeled on tougher US regulations of food and pharmaceuticals. The Food and Drug Administration generally does not allow drugs on the market that have not been tested for safety and effectiveness. That has largely not been the case for software — no governmental safety testing is done, for example, before a new social media platform is introduced. In Tutt’s vision, a similar agency could “require pre-market approval before algorithms can be deployed” in certain applications; “for example, a self-driving car algorithm could be required to replicate the safety-per-mile of a typical vehicle driven in 2012.”

This would effectively require certain algorithms to receive a government “license” before they can be publicly released. The idea of licensing for AI has taken off in recent months, with support from some in industry. OpenAI CEO Sam Altman called for “licensing or registration requirements for development and release of AI models above a crucial threshold of capabilities” in testimony before Congress. Jason Matheny, CEO of the Rand Corporation and a former senior Biden adviser, told the Senate, “we need a licensing regime, a governance system of guardrails around the models that are being built.” Gary Marcus, an NYU professor and prominent voice on AI, urged Congress to specifically follow the FDA model as it ponders regulating AI, requiring pre-approval before deployment.

**“Compute” regulation:** Training advanced AI models requires a lot of computing, including actual math conducted by graphics processing units (GPUs) or other more specialized chips to train and fine-tune neural networks. Cut off access to advanced chips or large orders of ordinary chips and you slow AI progress. Harvard computer scientist Yonadav Shavit has proposed one model for regulating compute. Shavit’s approach would place firmware (low-level code embedded into hardware) that can save “snapshots” of the neural networks being trained, so inspectors can examine those snapshots later, and would require AI companies to save information about their training runs so they can verify their activities match the information on the chip firmware. He would also have regulators monitor chip orders to ensure no one is purchasing a critical mass of unmonitored chips not subject to these regulations, just as biorisk experts have advocated monitoring gene synthesis orders to prevent the deliberate engineering of dangerous pathogens.

Export controls, like those the US placed restricting the sale of advanced chips to China, could also count as a form of compute regulation meant to limit certain nations or firms’ ability to train advanced models.

## **New institutions for a new time**

---

Implementing all of the above regulations requires government institutions with substantial staff and funding. Some of the work could be done, and already is being done, by existing agencies. The Federal Trade Commission has been aggressive, especially on privacy and bias issues, and the National Institute of Standards and Technology, a scientific agency that

develops voluntary standards for a number of fields, has begun work on developing best practices for AI development and deployment that could function either as voluntary guidelines or the basis of future mandatory regulations.

But the scale of the challenge AI poses has also led to proposals to add entirely new agencies at the national and international level.

**The National Artificial Intelligence Research Resource:** One new federal institution dedicated to AI is already in the works. A 2020 law mandated the creation of a task force to design a National Artificial Intelligence Research Resource (NAIRR), a federally funded group that would provide compute, data, and other services for universities, government researchers, and others who currently lack the ability to do cutting-edge work. The final report by the task force asked for \$2.6 billion over six years, though Congress has not shown much interest in allocating that funding as of yet. A bipartisan group in the House and Senate recently introduced legislation that would formally establish NAIRR and instruct the National Science Foundation to fund it.

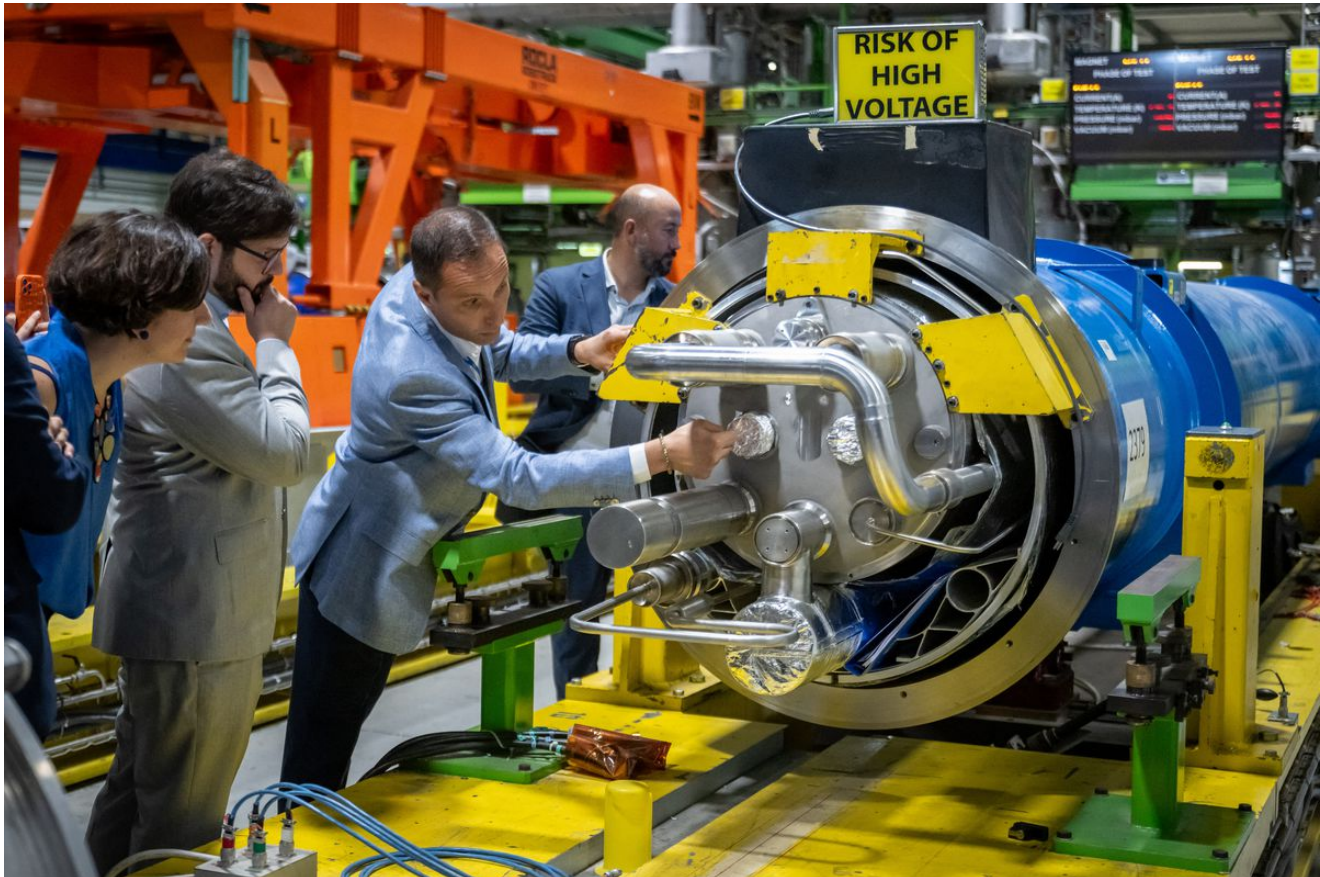
In contrast to technologies like nuclear power and even the early internet, AI is dominated by the private sector, not the government or universities. Companies like OpenAI, Google DeepMind, and Anthropic spend hundreds of millions of dollars on the server processing, datasets, and other raw materials necessary to build advanced models; the goal of NAIRR is to level the playing field somewhat, “democratizing access to the cyberinfrastructure that fuels AI research and development,” in the words of the National Science Foundation’s director.

**Dedicated regulator for AI:** While agencies like the FTC, the National Institute of Standards and Technology, and the Copyright Office are already working on standards and regulations for AI, some stakeholders and experts have argued the topic requires a new, dedicated regulator that can focus more specifically on AI without juggling it and other issues.

In their testimony before Congress, OpenAI CEO Sam Altman and NYU professor Gary Marcus both endorsed creating a new agency. Brad Smith, president of Microsoft, has echoed his business partner Altman and argued that new AI regulations are “best implemented by a new government agency.” Computer scientist Ben Schneiderman has suggested a National Algorithms Safety Board, modeled on the National Transportation Safety Board that investigates all airplane crashes, some highway crashes, and other transportation safety disasters.

Sens. Michael Bennet (D-CO) and Peter Welch (D-VT) have introduced legislation that would act on these suggestions and create a Federal Digital Platform Commission, charged with regulating AI and social media.

But others have pushed back and argued existing agencies are sufficient. Kent Walker, the top policy official at Google, had suggested that the National Institute of Standards and Technology (NIST) should be the main agency handling AI. (Notably, NIST does not have any regulatory powers and cannot compel tech companies to do anything.) Christina Montgomery, a top executive at IBM, similarly told Congress that taking time to set up a new agency risks “slow[ing] down regulation to address real risks right now.”



Chilean president Gabriel Boric, center, examines a gigantic magnet at a visit to CERN.  
*Fabrice Coffrini/AFP via Getty Images*

**CERN for AI:** Rishi Sunak, the UK prime minister, pitched President Joe Biden on setting up a “CERN for AI,” modeled after the Conseil européen pour la recherche nucléaire (CERN) in Geneva, which hosts large-scale particle accelerators for physics research and was the birthplace of the World Wide Web. Advocates like the computer scientist Holger Hoos argue that setting up such a facility would create “a beacon that is really big and bright,” attracting talent from all over the world to collaborate on AI in one location not controlled by a private company, making the exchange of ideas easier. (The sky-high salaries being offered to AI experts by those private companies, however, might limit its appeal unless this institution could match them.)



A recent paper from a team of AI governance experts at Google DeepMind, OpenAI, several universities, and elsewhere proposed specifically setting up a CERN-like project for AI safety. “Researchers—including those who would not otherwise be working on AI safety—could be drawn by its international stature and enabled by the project’s exceptional compute, engineers and model access,” the authors write. “The Project would become a vibrant research community that benefits from tighter information flows and a collective focus on AI safety.”

**IAEA for AI:** Top OpenAI executives Sam Altman, Greg Brockman, and Ilya Sutskever said in May that the world will “eventually need something like an IAEA for superintelligence efforts,” a reference to the International Atomic Energy Agency in Vienna, the UN institution charged with controlling nuclear weapons proliferation and governing the safe deployment of nuclear power. UN Secretary-General António Guterres has echoed the call.

Others have pushed back on the IAEA analogy, noting that the IAEA itself has failed to prevent nuclear proliferation to France, China, Israel, India, South Africa, Pakistan, and North Korea, all of which developed their bombs after the IAEA’s inception in 1957. (South Africa voluntarily destroyed its bombs as part of the transition from apartheid.) Others have noted that the IAEA’s focus on monitoring physical materials like uranium and plutonium lacks a clear analogy to AI; while physical chips are necessary, they’re much harder to track than the rare radioactive material used for nuclear bombs, at least without the controls Yonadav Shavit has proposed.

In the same paper discussing a CERN-like institution for AI, the authors considered a model for an Advanced AI Governance Organization that can promote standards for countries to adopt on AI and monitor compliance with those standards, and a Frontier AI Collaborative that could function a bit like the US National Artificial Intelligence Research Resource on an international scale and spread access to AI tech to less affluent countries. Rather than copying the IAEA directly, the aim would be to identify some specific activities that a multilateral organization could engage in on AI and build a team around them.

## **New funding for AI**

---

Implementing new regulations and creating new institutions to deal with AI will, of course, require some funding from Congress. Beyond the tasks described above, AI policy experts have been proposing new funding specifically for AI capabilities and safety research by federal labs (which would have different and less commercially driven priorities than private companies), and for the development of voluntary standards for private actors to follow on the topic.

**More funding for the Department of Energy:** To date, much federal investment in AI has focused on military applications; the Biden administration’s latest budget request includes \$1.8 billion in defense spending on AI for the next year alone. The recent House and Senate

defense spending bills feature numerous AI-specific provisions. But AI is a general-purpose technology with broad applications outside of warfare, and a growing number of AI policy experts are suggesting that the Department of Energy (DOE), rather than the Pentagon, is the proper home for non-defense AI research spending.

The DOE runs the national laboratories system, employing tens of thousands of people, and through those labs it already invests considerable sums into AI research. “[DOE] has profound expertise in artificial intelligence and high-performance computing, as well as established work regulating industries and establishing standards,” Divyansh Kaushik of the Federation of American Scientists has written. “It also has experience addressing intricate dual-use technology implications and capability as a grant-making research agency.” These make it “best-suited” to lead AI research efforts.

On the Christopher Nolan end of the scale, the Foundation for American Innovation’s Sam Hammond has suggested that a “Manhattan Project for AI Safety” be housed in the Department of Energy. The project would facilitate coordination between private-sector actors and the government on safety measures, and create new computing facilities including ones that are “air gapped,” deliberately not connected to the broader internet, “ensuring that future, more powerful AIs are unable to escape onto the open internet.”

**More funding for the National Science Foundation:** Another place in government that has already been funding research on AI is the National Science Foundation, the feds’ main scientific grantmaker outside of the medical sciences.

The Federation of American Scientists’ Matt Korda and Divyansh Kaushik have argued that beyond additional funding, the agency needs to undergo a “strategic shift” in how it spends, moving away from enhancing the capabilities of AI models and toward “safety-related initiatives that may lead to more sustainable innovations and fewer unintended consequences.”



National Institute of Standards and Technology director Laurie Locascio is not exactly a household name, but she could be among the most important figures shaping AI regulation going forward.

*Guillaume Paumier*

**More funding for the National Institute of Standards and Technology:** NIST is not exactly the most famous government agency there is, but its unique role as a generator of voluntary standards and best practices to government and industry, without any regulatory function, makes it an important actor at this moment in AI history. The field is in enough flux that agreement on what standards should be binding is limited.

In the meantime, NIST has released an AI Risk Management Framework offering initial standards and best practices for the sector. It has also created a Trustworthy & Responsible Artificial Intelligence Resource Center, designed to provide training and documents to help industry, government, and academia abide by the Risk Management Framework. Some in Congress want to mandate federal agencies abide by the framework, which would go a long way toward adoption.

The AI firm Anthropic, which has made safety a priority, has proposed a \$15 million annual increase in funding for NIST, funding 22 additional staffers, to double staffing working on AI and build bigger “testing environments” where it can experiment on AI systems and develop techniques to measure their capabilities and possibly dangerous behaviors.

## **New people to take the lead on AI research**

---

A recent survey of AI researchers from the Center for Security and Emerging Technology (CSET), a leading think tank on AI issues, concluded that processors and “compute” are not the main bottleneck limiting progress on AI. The bottleneck for making intelligent machines is intelligent humans; building advanced models requires highly trained scientists and engineers, all of whom are currently in short supply relative to the extraordinary demand for their talents.

That has led many AI experts to argue that US policy has to focus on growing the number of trained AI professionals, through both expanded immigration and by providing more scholarships for US-born aspiring researchers.

**Expanded high-skilled immigration:** In 2021, the National Security Commission on Artificial Intelligence, a group charged by Congress with developing recommendations for federal AI policy, argued that Congress should dramatically expand visas and green cards for workers and students in AI. These are incredibly common proposals in AI circles, for clear reasons. The US immigration system has created major barriers for AI researchers seeking to come here. One survey found that 69 percent of AI researchers in the US said that visa and immigration issues were a serious problem for them, compared to 44 percent and just 29 percent in the UK and Canada, respectively.

Those countries are now using these difficulties to aggressively recruit STEM professionals rejected from the US. The analyst Remco Zwetsloot has concluded that these dynamics are creating “a consensus among U.S. technology and national security leaders that STEM immigration reform is now ... ‘a national security imperative.’”





Ilya Sutskever was born in Russia, grew up in Israel, and got his university degrees in Canada before moving to the US and cofounding OpenAI.

*Jack Guez/AFP via Getty Images*

**Funding for AI education programs:** Similarly, some policymakers have proposed expanding subsidies for students to gain training in machine learning and other disciplines (like cybersecurity and processor design) relevant to advanced AI. Sens. Gary Peters (D-MI) and John Thune (R-SD) have proposed the AI Scholarship-for-Service Act, which would provide undergraduate and graduate scholarships to students who commit to working in the public sector after graduation.

## **The ground is still shifting**

---

These four areas — regulation, institutions, money, and people — make up the bulk of the AI policy conversation right now. But I would be remiss if I did not note that this conversation is evolving quite rapidly. If you told me in January, barely a month after the release of ChatGPT, that CEOs of OpenAI and Anthropic would be testifying before Congress and that members would be taking their ideas, and those of unaffiliated AI risk experts, seriously, I would have been shocked. But that's the territory we're in now.

The terrain is shifting fast enough that we could be in an entirely different place in a few months, with entirely different leading actors. Maybe the AI labs lose influence; maybe certain civil society groups gain it; maybe the military becomes a bigger component of these talks.

All that makes now a particularly sensitive moment for the future of AI. There's an idea in tech policy called the Collingridge dilemma: When a technology is novel, it's easier to change its direction or regulate it, but it's also much harder to know what the effect of the technology will be. Once the effect of the technology is known, that effect becomes harder to change.

We're in the "unknown impact, easier to influence direction" stage on AI. This isn't an area of intractable gridlock in DC, at least not yet. But it's also an area where the actual technology feels slippery, and everything we think we know about it feels open to revision.

***You've read 2 articles in the last 30 days.***

---

### **Will you support Vox's explanatory journalism?**

Most news outlets make their money through advertising or subscriptions. But when it comes to what we're trying to do at Vox, there are a couple reasons that we can't rely only on ads and subscriptions to keep the lights on.

First, advertising dollars go up and down with the economy. We often only know a few months out what our advertising revenue will be, which makes it hard to plan ahead.

Second, we're not in the subscriptions business. Vox is here to help everyone understand the complex issues shaping the world — not just the people who can afford to pay for a subscription. We believe that's an important part of building a more equal society. We can't do that if we have a paywall.

That's why we also turn to you, our readers, to help us keep Vox free. If you also believe that everyone deserves access to trusted high-quality information, will you make a gift to Vox today?

Yes, I'll give \$5/month

We accept credit card, Apple Pay, and Google Pay. You can also contribute via

### **The rise of artificial intelligence, explained**

---