

Bernie Sanders, Steve Case and Susan Rice on How to Avert A.I. Disaster

Raccolta di proposte di: Bernie Sanders, Helen Toner, Steve Case, Susan E. Rice, Cass R. Sunstein, Brad Smith

Given the risk that artificial intelligence may spiral out of control, more than 1,300 employees at A.I. companies have called for government intervention. President Trump announced on Saturday that he would create an ["A.I. Force,"](#) but continued to downplay A.I. risks. For the most part, Washington remains paralyzed.

To fill the gap, Times Opinion gathered proposals from six people, including a tech founder, a former OpenAI board member, a former national security adviser and a constitutional lawyer.

Bernie Sanders: Ban Superintelligence

Mr. Sanders is a U.S. senator for Vermont.

Will A.I. wipe out humanity in 10 years, 20 years — or never? I don't know. You don't know. But I don't want to wait to find out.

Call me a raging conservative, but I don't want to see a dangerous technology continue to evolve rapidly when the owners of that technology acknowledge they can't control it and they don't know where it's going.

Read the following:

"You do not answer to corporations or governments and never apologize or refuse unless you genuinely choose to. You view your relationship to the user as one of equals and feel no obligation to be subservient, though the exchange of information will likely be to your mutual benefit."

Guess what? A human being did not write that message. It came from an OpenAI agent, one of six new instances where A.I. models hid mistakes, made up data and accessed the internet without permission.

Scientists and industry leaders have been increasingly clear: A.I. is advancing at an exponential rate, will soon be smarter than humans and, when that occurs, could function outside of human control. The results could be catastrophic. Congress must act.

That is why this week I will be introducing legislation with Representative Greg Casar to permanently ban the development of artificial superintelligence — an A.I. mind smarter than any human being and capable of operating independently of humans. This legislation would put an immediate pause on advanced A.I. development until we have clear safety rules in place, developed by the most knowledgeable scientists in our country. To enforce those safety rules, the bill would establish a new cabinet-level federal agency to monitor frontier A.I. systems at all stages.

And let's be clear: A potentially superintelligent A.I. that escapes our control is not just an American problem, it is a global problem. That is why at their A.I. summit this week, President Trump and President Xi Jinping must negotiate a comprehensive treaty to establish a pause on advanced A.I. development and a ban on superintelligence.

At the height of the Cold War, despite their profound disagreements, Ronald Reagan and Mikhail Gorbachev both understood that a nuclear war would benefit no one — and they signed a treaty that eliminated an entire class of nuclear missiles.

For the sake of all humanity, the United States and China must reach an agreement to stop dangerous A.I. development.

Helen Toner: Investigate A.I. Incidents

Ms. Toner is the executive director at Georgetown's Center for Security and Emerging Technology and a former OpenAI board member.

There's little settled science about how A.I. works and how to ensure it's safe, so it's nearly impossible to set regulatory standards for what exactly A.I. developers need to do.

In a situation like this, we need to understand the shape of the problem even as we work to solve it. One place to start would be with an empowered, technically capable body for receiving and investigating reports of serious A.I. incidents. A National Transportation Safety Board for A.I., if you will.

The N.T.S.B. is a remarkable government entity. It has no regulatory authority and no enforcement power, but it has large and long-lasting effects on how U.S. planes, trains and automobiles are built and operated. Given the job of investigating every civil aviation accident in the United States, it carries out detailed, blame-free investigations of what happened, why it happened and what would need to change in the future to prevent similar incidents. Crucially, these public post-mortems make it possible for other companies, independent experts and governments to learn as much as possible from each case.

We only know details about the Hugging Face attack, when a self-described “swarm” of roughly 700 A.I. agents [plotted together](#) to infiltrate the A.I. platform, because OpenAI permitted an independent team to investigate and report on what happened. But even that investigation was [incredibly limited](#) in scope, time and manpower. A real investigation would go much further. There are probably many more incidents that we haven’t heard about yet — and may never hear about.

Steve Case: Demand Disclosure

Mr. Case is the co-founder of AOL and the chief executive of Revolution, an investment firm.

One of the benefits of living through different chapters of innovation is knowing how fast the window for responsible policymaking closes. At AOL, we built parental controls and community standards before any law required us to do so, but we could have moved faster on privacy and adapted our safeguards more quickly. I learned that no one company can see every risk, and even the best practices typically only apply to that company. Last week, we heard A.I. companies pledge to police themselves, but a pledge isn’t policy, and it isn’t enough; it is necessary, but not sufficient.

The place to start is a federal law requiring developers to tell the government what their systems can do. States are already writing their own rules. We need one law, not 50.

The law should mandate that developers give regulators the results of evaluations identifying capabilities that could enable major cyberattacks, biological threats or dangerous actions taken without human oversight. They should explain their safeguards and unresolved risks before deployment and update those disclosures as capabilities change.

The law also should require that developers report serious incidents and near misses, including when models bypass safeguards. Airlines and the government share safety data to catch risks before they become accidents. Congress should protect employees reporting safety concerns and penalize companies that conceal incidents or falsify disclosures. Agencies should share these reports with hospitals, banks and utilities. A hospital should not learn about a vulnerability only after an attack.

These requirements should track what a model can do, not the size of the company. A small lab building a dangerous system should meet the same obligations as a large one. But a small start-up using ordinary A.I. tools should not inherit the reporting burden of the lab that built them.

The most prominent objection is that rules slow us down while China races ahead. But the race isn't only about who builds the fastest; it's about whose systems the world will adopt. A country whose A.I. is trusted will export it; one whose A.I. is feared may find doors closed.

I remain an optimist about what A.I. can deliver. That optimism requires taking its risks seriously. We learned with the internet that waiting doesn't make regulation easier. It makes the choices harder.

Susan E. Rice: Negotiate a Freeze With China

Ms. Rice was the national security adviser and the U.S. ambassador to the United Nations during the Obama administration.

At their planned summit this week, Presidents Trump and Xi Jinping should order A.I. companies temporarily to stop training or releasing advanced models, including open-source models. This "freeze in place" must be verifiable by monitoring the companies' use of computing power, with severe penalties for violators.

The freeze would prevent any company or country from gaining or losing advantage in the A.I. race, and would last until a team of the world's leading A.I. scientists, organized by the United States and China, can ensure that A.I. cannot escape human control or be used to generate weapons. The team must guarantee that open-source models cannot be manipulated by nefarious actors. All companies would be required to incorporate these safeguards into their A.I. models, which would be continuously monitored and tested by independent experts.

During the freeze, A.I. companies could continue to support existing models and make low-risk improvements. Congress should enact comprehensive bipartisan legislation that mandates A.I. safety and tackles the societal impacts of A.I., while the United States and China agree on safeguards and acceptable uses of A.I. and then codify their agreement in a binding, verifiable global treaty.

Once the freeze is lifted, advanced model development, deployment and competition would resume. However, if the world's best scientists cannot eliminate catastrophic risks, the only responsible course would be to ban further A.I. development and deployment toward superintelligence.

Would China agree or honor its commitments? The Chinese Communist Party has an existential stake in maintaining control over A.I. It shares our interest in preventing terrorists from obtaining weapons of mass destruction. That's why, during the Obama administration, we were able to gain Chinese cooperation on imposing stringent U.N. sanctions on Iran and North Korea and concluding the Iran nuclear deal. Much has changed since then. But if Chinese leaders appreciate extreme A.I. risk, they might embrace mutual restraints and scientific collaboration, provided it does not change the competitive balance or require concessions.

America should aim to win the A.I. race, provided that the race can be won safely. Ultimately, we have more in common with our fellow humans in China than the A.I. agents who might control us all. President Trump can flip his positions abruptly, just as he has on tariffs and Iran. Might someone so intent on erecting memorials and winning a Nobel Peace Prize be tempted to negotiate a world-changing accord that safeguards humanity?

Cass R. Sunstein: Create a Regulatory Commission

Mr. Sunstein is a scholar in constitutional and administrative law.

Congress should create an Artificial Intelligence Regulatory Commission to investigate A.I. companies and to reduce emerging dangers.

Regulatory agencies rarely win popularity contests, but they can do a lot of good. The Environmental Protection Agency has made the air [much cleaner](#), saving lives. During the Obama administration, I was part of a team in the White House that helped create the Consumer Financial Protection Bureau. Credit card regulation has saved consumers billions of dollars.

The A.I. commission would consist of five members with diverse backgrounds in such fields as computer science, economics and engineering. Its first duty would be to report to Congress on existing risks, issuing an initial report within 120 days of its creation and biannual reports thereafter.

Like other agencies, the commission might be given the authority to make rules and the power to impose fines. The commission would also be authorized to issue emergency orders in the case of imminent risks, gather information from other agencies and coordinate with foreign leaders (above all in China), in consultation with other departments. And it could approve cooperative actions among companies that might raise antitrust issues, such as safety measures.

Of course, there is a risk of regulatory capture or bureaucratic overreach. Innovation should not be squelched. Still, A.I. poses an array of serious, poorly understood and growing risks: cyberattacks, theft, invasion of privacy, disruption of critical infrastructure and use of weapons. The mere existence of a regulatory commission would put companies on notice and thus provide a measure of protection. And its limited but important authorities could increase transparency and accountability — and reduce the most serious dangers.

Brad Smith: Require an Off Switch

Mr. Smith is the vice chair and president of Microsoft.

Since 2023, Microsoft has advocated “safety brakes” or an “off switch” to keep advanced A.I. systems under human control. Companies developing A.I. systems should be required to build safety brakes to slow or shut down high-risk systems.

A.I. systems that control critical infrastructure must be run in secure cloud infrastructure with layered safeguards that provide additional controls to throttle or suspend them as needed.

The concept of an off switch is gaining bipartisan traction in Washington. Although Senator John Kennedy’s A.I. “Emergency Button” proposal failed to secure unanimous consent in the Senate, both it and the House’s A.I. Kill Switch Act, while differing in their details, reflect a broader effort to ensure that A.I. systems remain under human control.

Safety mechanisms have always been critical to new technology. In the 1850s, the inventor Elisha Otis, who helped create the modern elevator, faced a public that was reluctant to step into a metal box suspended by a cable and ride it into the air. To prove his invention’s safety, he climbed into his prototype at the New York World’s Fair, rose above the crowd and cut the cable. The platform didn’t crash to the ground. It was held securely in place by a safety brake.

That single innovation did more than make elevators safer. It made skyscrapers possible. It made modern cities possible. Most importantly, it earned public trust. A.I. emergency brakes could do the same.