THE NEW YORK TIMES – 4 GIUGNO 2026

Trump Is Finally Facing Reality on A.I.

di Jen Easterly

President Trump has spent most of his second term resisting calls to impose federal oversight over artificial intelligence. But in recent weeks, his administration has shown signs of softening its hands-off approach.

Nearly two weeks ago, the president was set to sign an executive order at an event featuring American tech leaders, which would have asked A.I. companies to send new models to the government for review before releasing them to the public. The order was scrapped hours before it was to be signed, with Mr. Trump voicing concerns it would hamstring the private sector. Now, with less fanfare, he's signed a similar executive order that institutes a shorter review process.

Anti-regulation forces say the order goes too far. Pro-regulation groups say it doesn't go far enough. Having spent much of my career defending critical systems in government and industry, I welcome the move as recognition of an urgent new reality: The most powerful A.I. capabilities are becoming too consequential for national security to be released without meaningful coordination between the companies building them and the government responsible for protecting the country. It is, however, only a first step to building a stronger federal strategy for safeguarding Americans from threats posed by A.I.

In recent weeks, two leading American A.I. companies offered limited access to advanced cybercapabilities. Anthropic provided select entities early access to Claude Mythos Preview, which the company says has identified thousands of previously unknown vulnerabilities in critical software. OpenAI gave restricted early access to GPT-5.5-Cyber, providing vetted entities greater ability to conduct authorized security work under additional safeguards.

Both decisions reflect a kind of self-restraint rare in an industry where competitive pressure generally favors broader and faster deployment. They are a signal of how seriously these companies judge the risks of what they have built.

Those risks are real: The window between discovering a vulnerability and weaponizing it is rapidly compressing. That matters because the United States does not have a cybersecurity problem so much as a software quality problem. Much of the multibillion-dollar cybersecurity industry exists to compensate for technology built for speed, convenience and features — not security.

The software underpinning banks, hospitals, telecommunications networks, water systems and government services remains riddled with flaws and defects. Some have lingered for many years, and bad actors could exploit them to disrupt critical services, steal sensitive information or hold organizations hostage through ransomware.

While previous tools could probe for known weaknesses, new A.I. models can analyze vast amounts of software to uncover previously unknown vulnerabilities at scale. In some cases, they can identify flaws hidden in widely used software to the potential detriment of thousands of organizations before anyone realizes they exist.

On the flip side, A.I. may offer the best opportunity we have ever had to change that. These same models could help developers write more secure code and help defenders find and fix vulnerabilities before they are exploited by criminals or hostile governments.

A model capable of identifying dangerous vulnerabilities across widely used software could give defenders an extraordinary advantage. In the wrong hands, it could give adversaries the same advantage.

Mr. Trump's executive order is an attempt to get ahead of possible exploitation by creating a program for A.I. companies to voluntarily give the government a 30-day early-access window for reviewing new models before release. In addition, it establishes an A.I. cybersecurity clearinghouse within the federal government to coordinate vulnerability discovery, validation and remediation across industry and critical infrastructure. It also rightly recognizes the need to support those entities least equipped to defend themselves: rural hospitals, community banks and small utilities.

Critics will argue that any safeguards around frontier A.I. risk slowing American innovation at precisely the moment China is working aggressively to develop comparable capabilities. That concern is real. But the greater risk to American advantage is that we deploy these models carelessly and hand adversaries the same tools. A powerful model released without meaningful access controls is not an American asset; it is a global one.

Constraining China's access to powerful A.I. requires stronger controls on advanced computing infrastructure, better enforcement against circumvention and greater protection against theft of frontier-model capabilities. Responsibly governing what American companies are already building requires something different: government and industry working together before the most powerful systems are widely deployed.

Recognizing the right principle is only a first step. A voluntary framework cannot guarantee that government technical experts will be able to evaluate the most consequential capabilities of frontier models before those models are released, precisely when competitive pressure is greatest and a company has the strongest incentive to move quickly. And a principle enshrined in an executive order is only as durable as the administration that issued it.

It is time for Congress to act, not with rules that treat every A.I. start-up as a national security threat or stifle the many beneficial uses of artificial intelligence, but with focused, durable obligations for the most advanced models. A bipartisan draft bill reportedly expected this week from Representatives Jay Obernolte of California and Lori Trahan of Massachusetts is a promising sign that this work is already underway. Whatever its final form, obligations should include robust safety and security testing, secure development practices and responsible controls on deployment.

The United States already governs access to other elements with beneficial and harmful capabilities — pathogens, certain chemicals, nuclear materials — not to prevent legitimate research but because some capabilities require that the government know who has them and why before they are widely available. An A.I.

model with demonstrated offensive cybercapabilities is a dual-use technology in precisely that sense. Congress should treat it as one.

The United States should lead the world in artificial intelligence. But leadership cannot mean being the first to hand adversaries a tool capable of finding unknown vulnerabilities across the software running our hospitals, financial systems and power grid, and hoping we can manage the consequences later. The president's executive order is a good first move. The harder work — building an A.I. ecosystem that is innovative, trusted and resilient — still lies with Congress.