



Written by [Luis Miguel](#) on July 22, 2023

# White House, Tech Firms Agree to Implement Voluntary AI Safeguards

As the capabilities of artificial intelligence continue to grow, policymakers are working with Big Tech to place guardrails around it.

Joe Biden on Friday hosted a White House meeting of CEOs and presidents from seven of the largest tech companies at which the leaders of the firms agreed to a number of rules for governing AI development.

It remains to be seen whether Congress will pass any AI-related legislation this session. In the meantime, the companies are holding themselves to the nonbinding guidelines as a momentary means for dealing with rising concerns about artificial intelligence.



Melpomenem/iStock/Getty Images Plus

As the outlet *Politico* [noted](#):

Amazon, Anthropic, Google, Inflection, Meta, Microsoft, and OpenAI all agreed to a set of eight rules Friday, which include external testing of AI systems before their release, investing in cybersecurity protection for unreleased models and using a watermarking system for AI-generated content. The list of attendees includes Microsoft President Brad Smith, Meta President Nick Clegg, and Google President Kent Walker.

The companies' commitment to these safeguards is meant to "underscore three key principles that must be fundamental to the future of AI: safety, security, and trust," [a] White House official said. The key AI principles highlight the administration's stated focus on protecting AI models from cyberattacks, countering deep fakes and opening up dialogue about companies' AI risk management systems.

The administration's involvement in the tech leaders' meeting was lauded by industry groups. Global software industry advocate BSA said in a statement: "Today's announcement can help to form some of the architecture for regulatory guardrails around AI."

The White House, meanwhile, is reportedly preparing an executive order on AI, though the specifics and a timetable for it are yet to be announced.

Back in March, over 1,100 professionals in artificial intelligence and related fields [signed an open letter](#) that called for a six-month moratorium on "giant AI experiments." Signatories include Elon Musk, Andrew Yang, and Apple co-founder Steve Wozniak.

The letter argues that "AI systems with human-competitive intelligence can pose profound risks to society and humanity" and that, as a result, innovations should be "planned for and managed with commensurate care and resources" in order to prevent an "out-of-control race to develop and deploy ever more powerful digital minds that no one — not even their creators — can understand, predict, or



Written by [Luis Miguel](#) on July 22, 2023

---

reliably control.”

The tech executives further contended that governments should step in if AI developers won't implement safeguards of their own will. In the signatories' view, the government should get involved by creating regulatory bodies, funding safety research, and providing economic support for citizens when AI replaces large swaths of human employment.

The unknowns behind the nature and capacity of artificial intelligence make it a controversial subject, and the populace is as awed at its potential for good as its risks.

In May, it was reported that an AI-driven drone in a virtual simulation by the Air Force chose to “kill” its human operator in order to comply with its goal, although the Air Force later came out and [denied that story](#). In China, researchers have created small, unmanned aircraft powered by AI that have [bested ones remotely controlled by humans](#) in a dogfight.

In Romania, the prime minister earlier this year added an AI [advisor to the cabinet](#). Ion, an artificial intelligence with a mirror-like chrome screen that can display words and even a digital face, is tasked with analyzing social media networks in order to inform policymakers “in real time of Romanians' proposals and wishes,” per Prime Minister Nicolae Ciucă.

Because Ion will ostensibly allow the Romanian government to have a clear understanding of the moods and thoughts of the people, Ciucă said, Romanians should consider it their duty to use the online website and Twitter account to give their opinion on political issues and current events. But could this technology be abused or manipulated?

In one shocking example from 2016 of the surprise turn AI can take away from the intention of those who create it, Microsoft developed a “smart” AI Twitter account named TayTweets that was designed to learn from user communication.

The idea sounded innocent enough — except online trolls quickly [turned TayTweets](#) into a Nazi within a matter of hours. Microsoft ended up deleting the account.

Of course, it could also be the government that would abuse artificial intelligence for use against the people. As *The New American* [previously reported](#), the Biden White House is spending over \$550,000 in public grant money to create an AI model that would be used in rapidly detecting “microaggressions” across social media.

The grant was made through Biden's \$1.9 trillion American Rescue Plan and gives researchers at the University of Washington the funds to develop a technology that would ostensibly shield internet users from discriminatory language.

Specifically, the researchers claim to be developing machine-learning models capable of reading through social media content to identify “implicit bias” and microaggressions, the latter being defined as language that might offend members of “marginalized” groups.

Artificial intelligence may potentially pose some dangers to society; but the greater danger is unchecked government power, which will use any means — including AI — to crush dissent.



## Subscribe to the New American

Get exclusive digital access to the most informative, non-partisan truthful news source for patriotic Americans!

Discover a refreshing blend of time-honored values, principles and insightful perspectives within the pages of "The New American" magazine. Delve into a world where tradition is the foundation, and exploration knows no bounds.

From politics and finance to foreign affairs, environment, culture, and technology, we bring you an unparalleled array of topics that matter most.



### What's Included?

- 24 Issues Per Year
- Optional Print Edition
- Digital Edition Access
- Exclusive Subscriber Content
- Audio provided for all articles
- Unlimited access to past issues
- Coming Soon! Ad FREE
- 60-Day money back guarantee!
- Cancel anytime.

**Subscribe**