Published on *The National Law Review* https://natlawreview.com

Tech Giants' Pledge: Responsible AI Development for a Safer Future [PODCAST]

Article By:

Theodore F. Claypoole

Pin Wu, Ph.D.

On Friday, July 21, 2023, the White House announced that Amazon, Anthropic, Google, Inflection, Meta, Microsoft, and OpenAI voluntarily committed to numerous principles for the responsible development of artificial intelligence (AI) technologies. The principles include safety, security, and trust, which impose various non-binding duties on the companies. The companies agreed to ensure AI products are safe before public release, to safeguard against threats, misuse and risks, and to earn people's trust. The commitments apply to generative AI models that are more powerful than the current industry frontier.

In particular, regarding safety, the companies committed to use internal and/or external red-teaming to verify AI safety against unintended use, societal risks, and national security (e.g., bio, cyber, etc.) concerns. The companies pledged to share information with the industry and the government regarding safety risks, dangerous capabilities, and attempts to circumvent safeguards.

To encourage AI security, the companies agreed to invest in cybersecurity to protect proprietary and unreleased AI model weights and to detect and report vulnerabilities in AI models and methods of using them.

To engender trust in AI, the companies committed to provide mechanisms for users to determine whether content is AI-generated, which may include watermarking AI-produced audio, video, and graphics. The companies committed to publicly report AI capabilities, limitations, and proper uses. The companies agreed to protect user privacy when collecting data for the AI technologies and to prioritize research into avoiding harmful bias and discrimination caused by AI technologies.

These commitments echo the principles in the "Blueprint for an AI Bill of Rights, Making Automated Systems Work for the American People," published by the White House Office of Science and Technology Policy in October 2022. The voluntary commitments by these seven leading technology companies form a beginning in developing and establishing future binding obligations to minimize risks of AI technologies and allow innovations to continue thriving, both in the U.S. and worldwide.

Industry leaders observed that these voluntary commitments align with strategies already employed by Amazon, Anthropic, Google, Inflection, Meta, Microsoft, and OpenAI, and the commitments are not enforced by government action.

National Law Review, Volume XIII, Number 336

Source URL:https://natlawreview.com/article/tech-giants-pledge-responsible-ai-development-safer-

future-podcast