Policymakers Must Prepare for the Advent of AI Disinformation

In discussions of artificial intelligence, commentary often focuses on how this technology may automate jobs, revolutionize war, or even lead to the collapse of humanity. However, a more immediate issue deserves serious consideration from national security policymakers. This is the potential for AI advancements—specifically generative AI—to turbocharge disinformation and flood our information ecosystem with untruths.

On the one hand, a new era of AI-powered disinformation could allow malign actors—particularly foreign adversaries—to manipulate the information environment and shape public discourse more easily, as Russia did successfully in the 2016 U.S. election. On the other hand, and perhaps more concerningly, the proliferation of content produced using this technology could erode public trust in the information we consume entirely, undermining the social fabric that holds societies together.

National security policymakers in the United States must recognize the threat these advancements pose to national and international security and prioritize addressing it. Given America's role as a global leader in the development of AI, U.S. policymakers have a responsibility to coordinate an international response to the coming era of disinformation and work with partners to prevent the further breakdown of our shared reality that this technology threatens.

At the heart of this challenge is generative artificial intelligence. This type of AI helps create hyperrealistic content—including text, images, audio, and video—by learning from large datasets. Interest in generative AI has spiked in recent months largely because of its application in ChatGPT, a sophisticated chatbot developed by OpenAI that has exploded in popularity since its public release in late 2022.

Many people have found value in these AI tools. However, as a dual-use technology, generative AI can easily be abused. The potential harm that this technology poses, particularly to our information environment, is becoming increasingly clear. Researchers have raised the alarm for years about the risks of "deepfakes"—highly realistic, AI-generated depictions of real people doing or saying something that did not occur. Until recently, however, this technology was fairly primitive and not accessible to the general public. Today, with the proliferation of generative AI tools, it is becoming cheaper and easier to create synthetic media that seems authentic, lowering the barriers to entry for anyone who wants to spread disinformation.

This emerging crisis is closely linked with the market-driven arms race that is taking place in the AI industry. Many companies are moving aggressively to develop AI in pursuit of market dominance, despite concerns about how this technology could be abused.

With the rapid public deployment of these tools, AI-assisted disinformation is already beginning to appear online. While some cases have been good-spirited, there are multiple examples of politically motivated actors using AI technology to deceive and manipulate. AI-generated content has appeared in pro-Chinese influence operations and, more recently, in fictitious reports of an explosion at the Pentagon spread by Russian state media.

It is not hard to imagine how this technology could fuel genuine geopolitical crises and instability should it continue to advance and be publicly deployed without constraint. Consider, for instance, how the proliferation of deepfake audio and video content of U.S. political candidates during the 2024 presidential election could lead to the complete breakdown of trust in the political process and its outcome and, in turn, widespread civil unrest.

A future with AI-powered disinformation may be inevitable, but there is still time to mitigate the disruption that this technology causes.

The policymaking community in the United States is already somewhat aware of the issue. In 2021, the National Security Commission on Artificial

Intelligence reported that AI could “increase the magnitude, precision, and persistence of adversarial information operations.” Additionally, the National Institutes of Standards and Technology (NIST) issued guidance for developing responsible AI systems earlier this year. More recently, the White House has taken steps toward developing a national AI strategy. To date, however, no enforceable action has been taken to shape or regulate the future of AI.

Although leaders like OpenAI’s Sam Altman and Google’s Sundar Pichai have called for regulation, some AI experts believe it is not feasible due to coordination problems. However, this position discounts the cases where norms and regulations born out of safety and ethical concerns have successfully regulated the development of emerging technologies. For example, the Asilomar Conference in 1975 produced guidelines for the safe development of gene editing technology, while international non-proliferation pacts have largely halted the spread of nuclear weapons around the world.

Another concern, particularly in the national security community, is whether regulating AI could harm the interests of the United States and its allies in the new era of strategic competition with China. According to some analyses, anxieties around AI are overblown and it is essential to do what it takes to win the AI arms race.

That said, safeguarding against the unfettered proliferation of AI technology would not necessarily undermine U.S. or allied interests but could, in fact, advance those interests. China reportedly relies heavily on overseas workers and technology transfer in its efforts to develop AI. Thus, regulatory or normative efforts to slow AI advances by the United States could slow those made by China as well.

In addition to recognizing the threats that AI poses to public discourse and, in turn, collective security, the national security community in the United States should consider a few specific actions.

First, the United States must address the issue at home. In the immediate term, Congress should ensure the responsible development and public deployment of generative AI tools, specifically those that empower next-generation disinformation. While compelling, a lot of what generative AI produces today is still clunky or flawed in some way. However, recent advancements are making these flaws increasingly difficult to spot. In the absence of action, AI tools capable of producing synthetic media that is indistinguishable from the real thing could soon be publicly available. Thus, enforceable standards for companies developing this technology should be put in place, building on the NIST framework, before the horse leaves the proverbial barn.

U.S. policymakers and social media companies should also work more closely together on addressing AI-powered disinformation through technical mitigations. Most importantly, the federal government should explore ways to help social media companies detect and remove synthetic audio-visual media as well as synthetic text. The U.S. Department of Defense has already started to fund technology to defend against AI-enabled disinformation, but more action is needed. Additional government funding could help develop new detection capabilities, while the development of common technical standards for companies would allow for greater transparency across platforms.

Alongside these measures, the federal government should work to increase awareness of and resilience against AI-powered disinformation among the general public. Just as cybersecurity largely hinges on individual behavior and proper cyber hygiene, strong media literacy is important for resisting disinformation. As some researchers have suggested, one way to improve resilience is to provide communities most vulnerable to information manipulation—such as members of historically marginalized groups—with tools to identify fake content. This could be done by partnering with state and local governments and by coordinating with foreign partners on best practices.

Finally, the United States must lead an effort alongside close allies to develop international standards and norms to regulate the safe development, deployment, and use of generative AI. Ideally, these would follow standards agreed upon domestically but necessarily require alignment with the international community. The European Union would be a natural partner here, having already indicated its own concern for this technology through its Artificial Intelligence Act.

Given its concerns over generative AI, there may even be potential to gain China's buy-in on an international agreement. In recent months, the Chinese government has pursued efforts to regulate deepfake technology and require chatbots like ChatGPT to toe the party line due to concern over the chatbot's uncensored replies.

In any case, the need for urgent action cannot be ignored. We are already seeing how this technology is accelerating the erosion of our shared reality. With generative AI expected to advance exponentially in the coming years, this trend will only get worse. And if policymakers wait too long to act, the consequences could be dire. For that reason, leaders in the United States must take this issue seriously to protect American national security as well as the security of the international system.

Tristan Paci is a research intern for the Australian Strategic Policy Institute in Washington, DC. He is a recent master's graduate of the School of Foreign Service at Georgetown University, where he focused on technology policy and security in the Indo-Pacific.

Image: Shutterstock.