

By 2040, artificial intelligence could upend nuclear stability

April 24 2018

A new RAND Corporation paper finds that artificial intelligence has the potential to upend the foundations of nuclear deterrence by the year 2040.

While AI-controlled doomsday machines are considered unlikely, the hazards of [artificial intelligence](#) for nuclear [security](#) lie instead in its potential to encourage humans to take potentially apocalyptic risks, according to the paper.

During the Cold War, the condition of mutual assured destruction maintained an uneasy peace between the superpowers by ensuring that any attack would be met by a devastating retaliation. Mutual assured destruction thereby encouraged strategic stability by reducing the incentives for either country to take actions that might escalate into a [nuclear war](#).

The new RAND publication says that in coming decades, artificial intelligence has the potential to erode the condition of mutual assured destruction and undermine strategic stability. Improved sensor technologies could introduce the possibility that retaliatory forces such as submarine and mobile missiles could be targeted and destroyed.

Nations may be tempted to pursue first-strike capabilities as a means of gaining bargaining leverage over their rivals even if they have no intention of carrying out an attack, researchers say. This undermines strategic stability because even if the state possessing these capabilities

has no intention of using them, the adversary cannot be sure of that.

"The connection between nuclear war and artificial intelligence is not new, in fact the two have an intertwined history," said Edward Geist, co-author on the paper and associate policy researcher at the RAND Corporation, a nonprofit, nonpartisan research organization. "Much of the early development of AI was done in support of military efforts or with military objectives in mind."

He said one example of such work was the Survivable Adaptive Planning Experiment in the 1980s that sought to use AI to translate reconnaissance data into nuclear targeting plans.

Under fortuitous circumstances, artificial intelligence also could enhance strategic stability by improving accuracy in intelligence collection and analysis, according to the paper. While AI might increase the vulnerability of second-strike forces, improved analytics for monitoring and interpreting adversary actions could reduce miscalculation or misinterpretation that could lead to unintended escalation.

Researchers say that given future improvements, it is possible that eventually AI systems will develop capabilities that, while fallible, would be less error-prone than their human alternatives and therefore be stabilizing in the long term.

"Some experts fear that an increased reliance on artificial [intelligence](#) can lead to new types of catastrophic mistakes," said Andrew Lohn, co-author on the paper and associate engineer at RAND. "There may be pressure to use AI before it is technologically mature, or it may be susceptible to adversarial subversion. Therefore, maintaining strategic stability in coming decades may prove extremely difficult and all nuclear powers must participate in the cultivation of institutions to help limit nuclear risk."

RAND researchers based their perspective on information collected during a series of workshops with experts in nuclear issues, government branches, AI research, AI policy and national security.

"Will Artificial Intelligence Increase the Risk of Nuclear War?" is available at <http://www.rand.org>.

The perspective is part of a broader effort to envision critical security challenges in the world of 2040, considering the effects of political, technological, social, and demographic trends that will shape those security challenges in the coming decades.

Funding for the Security 2040 initiative was provided by gifts from RAND supporters and income from operations.

The research was conducted within the RAND Center for Global Risk and Security, which works across the RAND Corporation to develop multi-disciplinary research and policy analysis dealing with systemic risks to global security. The center draws on RAND's expertise to complement and expand RAND research in many fields, including security, economics, health, and technology.

Provided by RAND Corporation

Citation: By 2040, artificial intelligence could upend nuclear stability (2018, April 24) retrieved 21 May 2024 from <https://phys.org/news/2018-04-artificial-intelligence-upend-nuclear-stability.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.