

[Home](#) > [Publications](#) > [Articles](#)

AI & Global Governance: AI and Nuclear Weapons – Promise and Perils of AI for Nuclear Stability

[Frontier Technology \(https://cpr.unu.edu/keyword/frontier-technology\)](https://cpr.unu.edu/keyword/frontier-technology)

ARTICLE

2018.12.07

[Recommend](#)

[Tweet](#)



United Nations in the Era of AI

Dr Vincent Boulanin is a Senior Researcher at SIPRI. He joined SIPRI in 2014, where he works on issues related to the production, use and control of emerging military and security technologies, notably autonomous weapon systems and cyber-security technologies.

The conversation on how AI could impact warfare is still very young. In the past few years, much of this dialogue has revolved around the debate on Lethal Autonomous Weapons Systems (LAWS, aka [killer robots \(https://www.theverge.com/2018/8/27/17786080/united-nations-un-autonomous-killer-robots-regulation-conference\)](https://www.theverge.com/2018/8/27/17786080/united-nations-un-autonomous-killer-robots-regulation-conference)), which currently takes place within the framework of the [UN Convention on Certain Conventional Weapons \(https://www.unog.ch/80256EE600585943/\(httpPages\)/4F0DEF093B4860B4C1257180004B1B30?OpenDocument\)](https://www.unog.ch/80256EE600585943/(httpPages)/4F0DEF093B4860B4C1257180004B1B30?OpenDocument) (CCW). As its name suggests, the CCW focuses on conventional weapons, so it has a major blind spot on issues related to nuclear weapons and strategic stability. The transformative potential of AI, however, is also relevant for nuclear weapons and doctrines. AI could even be a driver of great 'entanglement' between the two areas. This convergence of AI and nuclear weapons deserves greater scrutiny.

AI and nuclear weapons: an old connection

The connection between AI and nuclear weaponry is not new. In fact, AI has been part of the nuclear deterrence architecture for decades. As early as in the 1960s, the United States and the Soviet Union saw that the nascent field of AI could play a role in the development and maintenance of their retaliatory capability – that is, the capability to respond to a nuclear attack, even by surprise. They pursued the [development of AI systems \(https://www.businessinsider.com/russias-dead-hand-system-may-still-be-active-2014-9\)](https://www.businessinsider.com/russias-dead-hand-system-may-still-be-active-2014-9) that could make their command and control process more agile and give decision-makers more time to focus on what really mattered: deciding whether to launch a nuclear strike or not. Early application of AI included automating threat detection, logistical planning for the transmission of launch orders, and missile targeting and guidance.

RECENT ARTICLES

ARTICLE



Women's Rights in a Post-2020 World: What Could (or Should) be Next for Gender at the Security Council

ARTICLE



The Importance of Decentralized Development Cooperation for Migrants in a Post-pandemic World

ARTICLE



On Women's Rights, the Security Council Should Stay Strong

RELATED PROJECTS

PROJECT

The New Geopolitics of Converging Risks: The UN and Prevention in the Era of AI

Early on, nuclear-armed states not only identified the appeal of AI for nuclear deterrence, they also saw its limitations. Given the dramatic consequences that a system failure would have, they were reluctant to hand over higher-order assessments and launch decisions to AI systems. A human had to remain 'in the loop'. The Soviet Union is the only country that pursued the development of a fully-automated command and control systems for nuclear weapons. However, this system, known as the [Dead Hand](https://www.wired.com/2009/09/mf-deadhand/) (<https://www.wired.com/2009/09/mf-deadhand/>), was meant to be activated only in the exceptional case of a decapitating attack on the Soviet nuclear command and control.

AI and nuclear warfare toolbox

What might change with the current AI renaissance, which is seeing breakthroughs in the areas of machine learning and autonomous systems? Recent advances in AI could be leveraged in all aspects of the nuclear enterprise. Machine learning could boost the detection capabilities of extant early warning systems and improve the possibility for human analysts to do a cross-analysis of intelligence, surveillance, and reconnaissance (ISR) data. Machine learning could be used to enhance the protection of the command and control architecture against cyberattacks and improve the way resources, including human forces, are managed. Machine learning advances could boost the capabilities of non-nuclear means of deterrence: be it conventional (air defence systems), electronic (jamming) or cyber.

Autonomous systems could be used to conduct remote sensing operations in areas that were previously hardly accessible for manned and remotely-controlled systems, such as in the deep sea. Autonomous unmanned systems such as aerial drones or unmanned underwater vehicles could also be seen by nuclear weapon states as an alternative to intercontinental ballistic missiles (ICBMs) as well as manned bomber and submarines for nuclear weapon delivery. These would be recoverable (unlike missiles and torpedoes) and could be deployed in ultra-long loitering periods – days, months or even years. At least one nuclear-armed state is already considering that possibility: In 2015, Russia revealed that it was pursuing the development of a nuclear-armed unmanned submarine, called [Status-6](https://nationalinterest.org/blog/buzz/status-6-why-russias-100-megaton-nuclear-torpedo-truly-terrifying-33661) (<https://nationalinterest.org/blog/buzz/status-6-why-russias-100-megaton-nuclear-torpedo-truly-terrifying-33661>).

Game-changing technologies?

Will the adoption of such systems fundamentally transform the field of nuclear strategy? The answer is no, at least not in the near-term, for three reasons.

First, these technologies reinforce rather than fundamentally alter the existing application of AI in nuclear force-related systems. Second, the field of nuclear weapon technology is renowned for its conservativeness – it has been historically slow at integrating new technologies. The US military, for instance, allegedly still uses [8-inch floppy disks](https://www.cnbc.com/2016/05/25/us-military-uses-8-inch-floppy-disks-to-coordinate-nuclear-force-operations.html) (<https://www.cnbc.com/2016/05/25/us-military-uses-8-inch-floppy-disks-to-coordinate-nuclear-force-operations.html>) to coordinate nuclear force operations. In that regard, machine learning and autonomous systems have some critical technical limitations that would make a rapid adoption unlikely in the near future. Machine learning systems operate like black boxes, which makes them potentially unpredictable, while the reliability of advanced autonomous systems is also technically hard to establish. Nuclear-armed states would have to crack difficult testing issues associated with the design of these systems to be confident that they can be used in a predictable and reliable manner and be certified for use. Third, the technology is not at the stage where it would allow nuclear-armed states to credibly threaten the survivability of each other's nuclear second-strike capability. Some experts have argued that a large-scale deployment of autonomous unmanned systems for remote sensing could make the [continuous at-sea deterrence obsolete](http://www.basicint.org/publications/david-hambling/2016/inescapable-net-unmanned-systems-anti-submarine-warfare) (<http://www.basicint.org/publications/david-hambling/2016/inescapable-net-unmanned-systems-anti-submarine-warfare>). In light of the current stage and development trajectory of AI technology and other key enabling technologies (such as sensor and power technology), this is bound to remain a very theoretical scenario for the foreseeable future.

Impact on strategic stability and nuclear risk

If recent advances in AI are unlikely to completely undermine the foundation of nuclear strategy, they could, without a doubt, have both a positive and negative impact on strategic stability.

On the one hand, recent advances of machine learning and autonomous could enhance stability as they provide nuclear weapon states with better information and better decision-making tools for time-critical situations, which would reduce the risk of miscalculation and accidental escalation. Moreover, they could generate new possibilities for the arms control community to monitor nuclear weapon-related developments and conduct verification operations.

On the other, the adoption – or even suspected adoption – of new AI capabilities by one or several nuclear-armed states could incentivize other states (be they nuclear-armed or not) to respond with destabilizing measures that could increase the likelihood of a nuclear conflict. This could include entering into an arms race, doubling down on the modernization of nuclear arsenals, renouncing a ['no](#)

[first use' policy \(https://www.cfr.org/backgrounder/no-first-use-and-nuclear-weapons\)](https://www.cfr.org/backgrounder/no-first-use-and-nuclear-weapons), increasing alert statuses, or further automating nuclear launch policies. Historical events like the [1983 Petrov incident \(https://www.bbc.com/news/world-europe-24280831\)](https://www.bbc.com/news/world-europe-24280831) (where the Soviet early warning systems wrongly detected a US nuclear attack) have also shown that AI technology could be the cause of an accidental or inadvertent escalation into a nuclear conflict.

Dealing with the risks

The risks posed by the convergence of AI and nuclear weapon technology are not necessarily new. Some of them have been known about for years. This means that solutions to address them may already exist. There might be no need to reinvent the wheel. 'No first use' policies, a commitment to lower the alert status of nuclear arsenals, as well as more openness about nuclear modernization plans and information-sharing via different dialogue tracks are measures that could clearly help to start mitigating the destabilizing potential of nuclear-related AI applications. It is essential that nuclear-armed states take note of the importance of this issue as part of the bilateral and multilateral talks on nuclear risk reduction.

This article has been prepared by Vincent Boulanin as a contributor to AI & Global Governance. The opinions expressed in this article are those of the author and do not necessarily reflect those of the Centre for Policy Research, United Nations University, or its partners.

Home > Publications > Articles

NAVIGATION

ABOUT

EXPERTS
RESEARCH

IMPACT
EVENTS

PODCASTS
VIDEOS

COMMUNITY

 FACEBOOK
 YOUTUBE

 TWITTER



UNITED NATIONS
UNIVERSITY
UNU-CPR

CONTACT US

DATA PRIVACY

TERMS OF USE

COPYRIGHT

767 Third Avenue, 35th Floor
New York, NY 10017
United States

[Designed and Developed
by the UNU Office of Communications
\(http://digital.unu.edu\)](#)