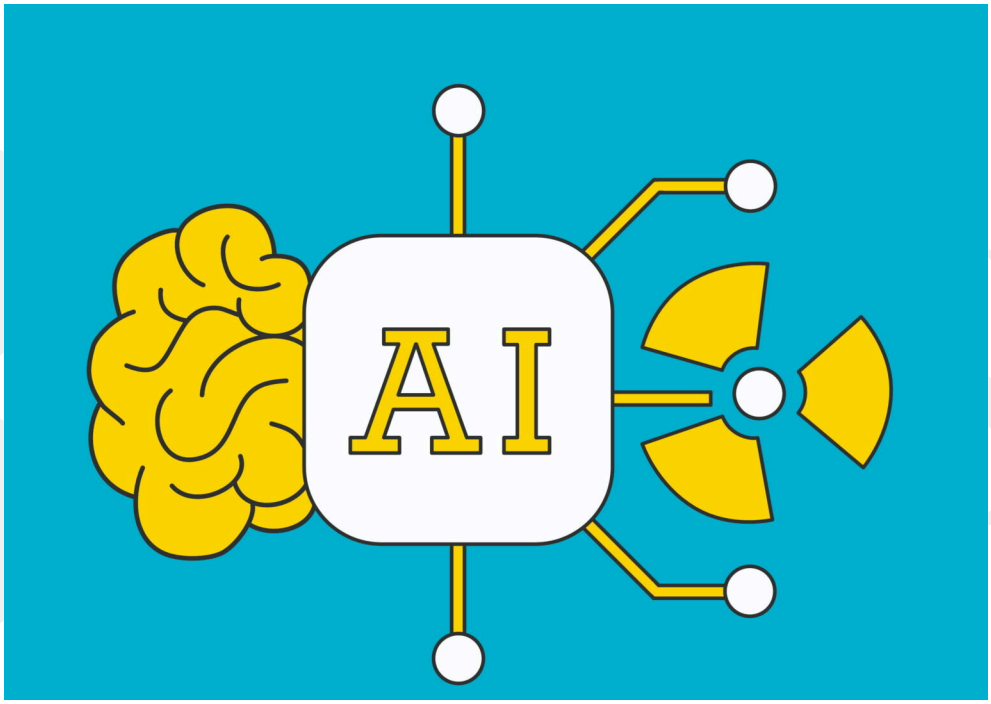
Miscalculation at Machine Speed?

27 OCTOBER 2024

Anton Meier



The nuclear balance between the United States and the Soviet Union was one, if not the main, reason the Cold War did not escalate into a full conventional and nuclear conflict. Now, the world again enters an era of great-power competition, and with it, the risk of nuclear confrontation rises. This time, the odds that a nuclear balance will lead to stability seem less promising. Certain actors, such as the Russian Federation, frequently threaten the use of nuclear weapons. This erosion of the nuclear taboo and fewer arms control treaties such as the INF Treaty and the Treaty on Open Skies are commonly known critical challenges to international security. However, the rapid advances in the development of artificial intelligence represent a significant yet often overlooked risk. The ability of AI and Machine Learning (AI/ML) systems with increasing computational power represents an opportunity to enhance military operations by enabling faster and more comprehensive data processing from multiple sources. This

includes nuclear deterrence which is a centrepiece of modern security architecture. While new technologies will not change the nature of statecraft and strategy per se, they will influence how states manage escalation.

On the tactical level, AI/ML capabilities can facilitate the detection of nuclear attacks. The North American Aerospace Defense Command has less than three minutes to assess and confirm initial indications from early-warning systems of an incoming attack. Any tool which can streamline this process and buy seconds of time would be invaluable to decision-makers. However, humans could end up as passive observers even when kept in the loop, because any human interaction would cost too much time. This is a challenge because while AI/ML systems can improve the processing of large quantitates of data, they still can misinterpret it. This could be caused by a technical failure or by an adversary intentionally delivering false information to the system. Additionally, with an increased dependence on AI/ML, nuclear command, control, and communications (NC3) structures could grow increasingly vulnerable to cyberattacks. The manipulation of battle networks through the delivery of false information or cyberattacks might compromise the ability to deploy nuclear weapons and thereby decrease deterrence.

On the strategic level, nuclear deterrence is primarily based on mutual assured destruction (MAD). MAD posits that a full-scale use of nuclear weapons by an attacker on a nuclear-armed defender with second-strike capabilities will result in the annihilation of both. The U.S. Department of Defense's Nuclear Matters Handbook states that: "counterforce targeting plans to destroy the military capabilities of an enemy force". Simply, put counterforce doctrine means the targeting of an opponent's conventional and nuclear military infrastructure with a nuclear strike. The objective is to limit or prevent the second-strike capability and therefore MAD. However, the improvement of AI/ML systems will lead to increasingly effective localisation of an opponent's nuclear weapons and the possibility of acquiring disarming strike capabilities: "Arsenals that are survivable today, however, can become vulnerable in the future." This development that reliable MAD might become undermined by advanced AI/ML-enabled counterforce targeting could disturb the nuclear balance and have escalatory effects.

If one state possesses the ability to destroy another state's second-strike capability, MAD is no longer guaranteed and hence the nuclear balance does no longer exist. In the situation that State A is confident that it located a significant number, or all, nuclear weapons of State B and can either destroy them before launching or intercept them, State B's deterrence failed. At the same time, State B also might be incentivised to escalate, as it fears being limited in employing its second-strike capabilities in the future because of State A's capability to destroy State B's nuclear assets. Catastrophic miscalculations could be the consequence. As Krabill and Johnson put it: "(E)ven a modicum of uncertainty about the effectiveness of AI-augmented cyber capabilities during a crisis or conflict would, therefore, reduce both

sides' risk tolerance, increasing the incentive to strike pre-emptively". AI/ML-based improved detection of nuclear weapons facilities is especially a challenge for Russian and Chinese defence planners, as both states rely primarily on mobile intercontinental ballistic missile launchers (ICBMs) for deterrence. These launchers are virtually impossible to harden but rely on their mobility for concealment. This concealment could be challenged through advanced AI/ML systems.

The USA are the NATO Ally with the largest nuclear arsenal and will, therefore, set the tone for any discussion on limiting the use of AI/ML in NC3. As nuclear deterrence is viewed as a central pillar of state survival, any measures which might limit its effectiveness are rightfully viewed critically by decision-makers. However, there is increasing advocacy of the dangers of AI-based targeting systems. Consequently, the US DoD was tasked in 2022 with a failsafe review to identify measures which could prevent the "unauthorized, inadvertent, or mistaken use of a nuclear weapon, including through false warning of an attack. Furthermore, the United States Congress introduced legislation prohibiting the use of federal funds for nuclear weapon systems based on AI/ML without meaningful human control. Nevertheless, further advocacy is necessary to regulate the escalation potential of AI/ML. Besides putting regulations on the use of AI in targeting and launching in place, a key challenge to tackle is, how these guidelines can be verified, as it is no longer sufficient to simply count warheads or launchers.

In conclusion, AI/ML systems could have both positive and negative implications for nuclear security. AI/ML capabilities will certainly decrease the fog of war and make nuclear weapons more capable. Whether this will have stabilising or destabilising effects in the long term is debatable. For now, however, AI brings uncertainty, which is certainly the poison of a nuclear balance.

DOI 10.69790/epis05.34366