Cyber-Nuclear Nexus: A Hazardous Duo

Introduction

With a tragic frequency, the greatest discoveries of mankind go hand in hand with its greatest horrors. The creation of a nuclear bomb under the Manhattan Project has changed the life of the human race forever, introducing new possibilities of warfare, unprecedented on the scale of destruction and casualties. This novel military technology brought with itself the brand-new dreadful scenarios of the future, where possible nuclear warfare would likely end up in total annihilation of life on Earth.

However, despite the undeniably huge amount of negative consequences associated with the emergence of nuclear weapons, this issue has also a positive aspect. This positivity is connected to the rational deterrence theory and lies specifically in a concept of Mutually assured destruction (MAD), based upon it. During the Cold War era, the MAD doctrine has been used as a military strategy and has proved to be useful in preventing a nuclear conflict between the rival blocks in the international system, governed by rational calculations of states. Moreover, the functioning of MAD logic has exceeded the Cold War and to this day serves as a stabilizer in the complicated Russo-American relationship and relations of the West with nuclear rogue states, such as North Korea.

However, the introduction of modern technologies, such as artificial intelligence (AI) can severely disrupt the fragile political balance, set by the promise of mutually assured destruction by massive reciprocal nuclear strikes. Today's AI capabilities can shift the tactical offense-defense balance and destroy the stability, established by nuclear deterrence (Maas, 2019). The impact of AI on nuclear strategic stability can be colossal and initiate a sequence of catastrophic events with fatal consequences. In general, the emergence of military AI, potentiated by machine learning (ML), expands the range of offensive cyber capabilities and is a potentially destabilizing component to the established balance of power in the international system. However, the military AI's destructive potential appears especially dangerous when it comes to nuclear weapons.

Nuclear command-and-control systems

Nowadays, the advanced AI technologies are used in leading powers' nuclear command-and-control systems, where both operation and cybersecurity areas hugely depend on AI's expert systems and ML algorithms. Such AI technologies are important elements to these systems, as their use brings many advantages, mainly in expanding the alternatives for response in the event of a crisis, therefore creating prospects for de-escalation. Both nuclear weapons and their command-and-control systems need protection from kinetic and digital attacks. However, even secured by the AI technologies to some extent, these command-and-control systems together with their human controllers can still be exposed to the offensive use of the same technologies by malicious actors (Cimbala, 2016).

The current reliance of nuclear states on AI technologies for strategic warning threatens that possible interference of some alien actors in the nuclear command-and-control system of any state would endanger nuclear strategic stability and nuclear peace (Fitzpatrick, 2019). Judging from the current development of AI technologies and particularly the latest achievements in the sphere of synthetic media, such as creating deepfakes based on deep machine learning, the possibility of nuclear escalation by third parties increases. In this scenario, a non-state actor creates a deepfake of a quality so high, that the global community fails to identify its artificial nature, therefore believing the fabricated story and falling into a provoked crisis between nuclear powers. During such escalation deadlock, the nuclear arsenals of the hypothetical adversaries would be put on high alert, ready for immediate retaliation, as demanded by the AI algorithms, therefore posing a direct threat of nuclear war (Fitzpatrick, 2019). In the case of such nuclear crisis, the absence of necessary communication channels between the parties, which were destroyed by either information or cyber warfare, would be crucial negative factors that are capable of undermining crisis management (Cimbala, 2016).

The use of AI technologies by malevolent third parties is not the only threat regarding the nuclear weapons - another grave danger associated with the AI participation in nuclear command-and-control systems lurks in the possible biases of AI programs. These errors can be created by manipulating the input data of the AI programs that are indeed driven by it - a step, which would pervert the programs' output and undermine their entire algorithm. Also, the assessment of the AI programs can be negatively affected by non-malign components, associated with the human design, which can result in inserting personal biases of the human curators in the system. These biases can lead to nuclear command-and-control systems' incomprehensible behavior, leaving no opportunity for human operators to understand the logic behind its calculations. This distortion of the AI-driven situational-awareness algorithms, responsible for the functioning of strategic-warning systems, can significantly trigger the crisis escalation. Interestingly, such distortion can result both from the unintentional insert of the trainer's personal bias and from the intentional injection of the skewed data during the algorithms' developmental phase (Fitzpatrick, 2019).

Nuclear facilities

When talking about the impact of AI on nuclear weapons, the so-called 'peaceful atom' deserves to be mentioned, as nuclear facilities operating under AI-driven management software can also be subject to various negative factors, like malicious activities of the third parties, human error, etc. Any interruption in the functioning of a nuclear power plant is potentially dangerous, as evident from the past glitches at Chernobyl and Fukushima nuclear power plants, which have led to the greatest nuclear catastrophes in modernity.

According to Fitzpatrick, the vulnerability of nuclear power plants can be seen in a hypothetical scenario of an attack that involves not only a technical aspect but also social engineering to target human operators. In this case, a third party launches a wave of cyber-attacks performed via digital noise and containing minor elements of a real threat – a strategy, which would cause an AI-driven system to react with false positives. These safety alerts generated by AI software would require the enactment of an emergency

protocol, leading to reactor shutdown. After no actual structural damage of the facility would be identified by specialists, the AI software provider would be appealed to fix the error in its product by increasing the damage threshold for the sensing functions that react to a threat by automatically shutting down the facility. The provider's reconfiguration of the AI software's ML algorithms would weaken the defensive character of the system, making it vulnerable to the exact type of attacks, which were used to trigger the system's false-positivity safety alerts. Besides, the operation of a nuclear power plant at full capacity or beyond engineering parameters because of easing the sensing indicators of software could destabilize the reactor and lead to a nuclear disaster. This scenario shows how easily 'peaceful' nuclear infrastructure like civilian nuclear reactors can be weaponized when human operators are manipulated by the malware into changing the AI algorithms in a way, that would favor malicious third parties (Fitzpatrick, 2019).

Conclusion

The cyber-nuclear nexus potentiated by the technological ascent of AI and related vulnerability of nuclear command-and-control systems as well as nuclear infrastructure appears as a hazardous tandem, which may threaten security in different ways. The international community is responsible for finding a way to control the rise of military AI, which is otherwise extremely likely to diminish nuclear deterrence, jeopardize the existing strategic stability, and put the world under the threat of the nuclear apocalypse and total annihilation of the human race.

References:

CIMBALA, Stephen J. Nuclear Deterrence in Cyber-ia. *Air & Space Power Journal* [online]. 2016, **30**(3), 54-63. ISSN 1555385X.

FITZPATRICK, Mark. Artificial Intelligence and Nuclear Command and Control. *Survival (00396338)* [online]. 2019, **61**(3), 81-92. DOI: 10.1080/00396338.2019.1614782. ISSN 00396338.

MAAS, Matthijs M. How viable is international arms control for military artificial intelligence? Three lessons from nuclear weapons. *Contemporary Security Policy* [online]. 2019, **40**(3), 285-311. DOI: 10.1080/13523260.2019.1576464. ISSN 13523260.