RSIS S. RAJARATNAM SCHOOL OF INTERNATIONAL STUDIES
Nanyang Technological University, Singapore

RSIS COMMENTARY

www.rsis.edu.sg No. 068 – 2 April 2025

Will AI Enhance Decision-Making in the Use of Nuclear Weapons?

By Alvin Chew and Asha Hemrajani

SYNOPSIS

AI has been integrated into nuclear weapon doctrines to facilitate efficient autonomous decision-making. While speed is often crucial in military operations, decisions to launch weapons of mass destruction, such as nuclear weapons, require deliberate human intuition and intervention that surpasses calculated assessments generated by AI.

COMMENTARY

Nuclear Weapons States (NWS) have been quick to incorporate AI into their nuclear doctrines, all hoping to have early adopters' advantage of the technology. However, the arcane knowledge of how AI functions, as well as the possibility of error, makes it too risky for nuclear decision-making. Hence, both the US and China have agreed that humans should be involved in matters of nuclear command, control and communications (C3). Even then, it remains perilous to incorporate AI as a decision-support tool for any potential nuclear launch. Agreements amongst NWS need to go beyond the vague "human-in-the-loop" rhetoric.

Secrecy of Nuclear Weapon Operations

Generative AI relies on Large Language Models (LLMs), which use advanced neural networks trained on massive amounts of text to predict and generate text. AI has been utilised in military applications for targeted precision strikes, as well as for intelligence gathering and surveillance. The gargantuan amounts of data and images collected can be rapidly analysed and accurately processed, enhancing decision-making in real-time operations.

Unlike conventional military operations, the launching of nuclear weapons is shrouded in secrecy. Furthermore, nuclear weapons, built for deterrence purposes, have not been used in conflicts since World War II. As history offers no case examples, LLMs will not have the benefit of learning from an abundance of open-sourced data regarding the catastrophic after-effects of nuclear weapon launches. LLMs will therefore be less effective when incorporated into the nuclear C3 structure.

Research conducted by the Stanford Institute for Human-Centered Artificial Intelligence compared five commercial LLMs that were used in military and diplomatic contexts. Due to the unavailability of real-world scenarios, simulations of nuclear crises were conducted to evaluate the effectiveness of AI models. Unquestionably, all the commercial LLMs tested demonstrated escalation risks – a characteristic of machine learning based solely on rational thinking.

However, nuclear deterrence – a core tenet of strategic stability – is executed based on a deep level of understanding of human psychology. Putin's strategy of "escalate to de-escalate" in the current Russia-Ukraine war would have breached the threshold of an autonomous nuclear launch if LLMs were to override or disproportionately influence human control in nuclear decision-making. AI systems are efficient in recognising patterns of events to arrive at a logical conclusion, but in their present stage of evolution, are incapable of penetrating the real intent of the deceptive human mind.

Physical Versus Cyber Domains

It is essential to define "human in the loop", i.e., the exact nature and degree of human involvement, in nuclear decision-making processes because AI cannot comprehend the consequences of a nuclear Armageddon. Hence, such high-impact events that will take place in the physical world cannot be left to AI, regardless of how robust the AI model and system are.

Furthermore, the question of liability cannot be transferred to an AI system. In a vicious cycle, any person tasked with launching a nuclear warhead will intuitively feel less pressured if the decision has been supported by AI, thereby likely lowering the threshold for launch. During the Cold War, there were instances when human intuition played a crucial role in averting a nuclear catastrophe. At the height of the Cuban Missile Crisis in 1962, a Soviet naval officer, Vasily Arkhipov, saved the world from World War III when he persuaded his submarine captain against firing a nuclear torpedo at pursuing US ships. Such cognitive pressures placed upon the human decision-maker can never be replicated in cyberspace, which is devoid of human instincts and emotions.

Conclusion

Recent years have seen the emergence of new initiatives, such as the REAIM Summit (Responsible AI in the Military Domain), which convenes governments, organisations, and experts to establish ethical guidelines and global norms for deploying AI in military contexts, including the use of nuclear weapons. Similarly, the Roundtable for AI, Security, and Ethics (RAISE) platform advocates stakeholder collaboration to ensure that AI technologies in defence and nuclear applications remain transparent, verifiable, and aligned with global security imperatives.

As AI systems become increasingly sophisticated and more widely used in military systems worldwide, it has become more imperative that proper governance over the use of AI in nuclear warfare is needed. Preliminary research shows that LLMs tend to lean towards escalatory outputs and decisions. Policymakers should acknowledge this initial research and insufficient information regarding model behaviour and consequently refrain from implementing LLMs for real-world decision-making in these contexts until further and more detailed research has been undertaken to study LLM behaviour under real-world conflict conditions, particularly regarding their inclination towards escalatory decisions.

If AI is to be used in nuclear weapon doctrines, the *LLM should be developed with a bias towards de-escalation*. It is hoped that international initiatives, such as REAIM and RAISE, will incorporate this philosophy into their work. By embedding a bias toward calm, measured responses, nuclear standoffs can be avoided, and alternative solutions can be encouraged. This, combined with strong oversight by human commanders, would help confirm that AI outputs in nuclear weaponry remain aligned with broader strategic, legal, and humanitarian principles.

Alvin Chew is a Senior Fellow at the S. Rajaratnam School of International Studies (RSIS), Nanyang Technological University (NTU), Singapore. Asha Hemrajani is a Senior Fellow at the Centre of Excellence for National Security (CENS) at RSIS.

S. Rajaratnam School of International Studies, NTU Singapore
Block S4, Level B3, 50 Nanyang Avenue, Singapore 639798