Generative Artificial Intelligence and Deterrence Stability between India and Pakistan

By

Anum A. Khan

Imagine a scene. It is the Indian military command center on the night of May 28, 2028. A high-ranking Indian military officer sends a message to a subordinate commander; Pakistan has initiated a preemptive strike. Prepare for a retaliatory strike. This is not a drill.

Missile flight time between both India and Pakistan is a few minutes. Hence, commanders do not have time for verification and decide to launch a retaliatory strike on warning.

In Pakistan, a high-ranking Pakistani military officer contacts Pakistan's prime minister to tell him India has initiated a nuclear strike, and an immediate response is required. The prime minister immediately orders a counterstrike. Pakistan proceeds to counterstrike protocols.

Missiles begin crossing borders between India and Pakistan as the world hurtles towards a nuclear catastrophe in South Asia. The world's worst fears are realized. It is only later realized that the high-ranking Indian military officer's message to an Indian commander was made by a nuclear command-and-control system embedded with generative artificial intelligence (AI).

The Role of Generative AI in Modern Warfare

The use of generative AI, which is best exemplified by ChatGPT, by an adversary, or third party, can impact deterrence stability between India and Pakistan. In a worst-case scenario, it could trigger a nuclear war, like the scene mentioned above. Should generative AI be embedded in nuclear command-and-control systems, it is theoretically possible that hyper-personalized messages, like the one described above, could be generated in a system that was not originally designed to allow such an outcome. The potential for a tailored message, for specific individuals, is possible.

Traditionally, content and information generation were based on human creativity. However, at present, generative artificial intelligence is being considered as a tool for digital transformation. Generative AI is a deep learning model which has the capability to generate synthetic media including text, video, and audio—dependent on training.

Perception Building and the Escalation of Crises

Generative AI can also play a distinctly different role in leading to nuclear crisis. In a deterrence framework, the development of a narrative and perception are crucial. Similarly, communication of deterrence threats is also achieved through perception-building of an adversary—concerning a state's capabilities and credibility. In the specific case of India and Pakistan, India is creating misperceptions regarding Pakistan's nuclear program. Recently, Pakistani Ambassador Munir Akram, at the United Nations Security Council's high-level debate, stated that India is actively pursuing a state-led disinformation campaign to malign Pakistan.

He was referring to recent EU Disinfolab's reports on how India is using fake nongovernmental organizations, think tanks, media outlets, and international organizations to malign Pakistan. Similarly, it is possible that India may use Generative AI to manipulate public

opinion, thereby, creating misperceptions among Pakistan's public. This could make the citizenry, as well as the government of both India and Pakistan, susceptible to separate facts from fiction—especially in a crisis situation. Hence, AI-driven narratives could push both countries to a crisis leading to warfighting.

Misinformation and Crisis Management Challenges in South Asia

Misinformation through generative AI can make crisis management and escalation dominance between India and Pakistan much more difficult. According to Heather Williams and Alexi Drew, Herman Kahn's escalation ladder has transformed into a web across domains with multiple actors and alliances, especially in the age of social media. In South Asia, due to mated warheads and cannasterization of missiles by India, coupled with geographical contiguity with Pakistan, nuclear use may become ultimate escalation within minutes—without a ladder or a web. Social media could be a threat multiplier and can become lethal with the use of generative AI. Although, tweets are unlikely to independently start a crisis, tweets from government officials of India and Pakistan could be fabricated through utilizing hacking and generative AI to escalate an ongoing crisis.

Risks of Generative AI in Command-and-Control Systems

While Indian nuclear weapons may have an effective command and control where nuclear weapons have nuclear safety and security mechanisms, there was a "Brahmos missile crisis" in 2022. Inadvertent or accidental, the launch of a missile could trigger a reactionary strike by Pakistan as the missile is dual-capable. Furthermore, this fog of war could undermine strategic stability. As a Group Captain was blamed for the launch, many commanders like him, who may be capable of launching nuclear missiles, even on Indian nuclear submarines, could be manipulated via high-level orders created through generative AI.

False Flag Operations and Strategic Escalation

India has also opted for adventurism against Pakistan through its false flag operations and surgical strikes. Pakistan has adopted a conventional strategy of quid-pro-quo plus as a conventional response to such Indian limited strike. However, malicious false flag operations through disinformation and fake news by India to garner public support for electoral gains and domestic politics can result in crisis escalation in such a competitive strategic environment. It is due to mistrust within the adversarial relationship of India and Pakistan.

The Need for Confidence-Building Measures (CBM) in AI

There is a dire need for risk reduction and confidence-building measures (CBM) between India and Pakistan in the domain of AI. However, the appetite for CBMs and dialogue on the Indian side is non-existent. India and Pakistan could establish bilateral AI incident reporting and mitigation centers to counter the use of AI as a tool for inducing false alarms and other escalatory actions. Furthermore, there is a need for India and Pakistan to further opt for unilateral AI security and regulation measures, especially for inter-organizational and inter-state strategic communications. If leading states like the United States, China, and Russia opt for an AI arms

control agreement, less powerful states, including India and Pakistan, may be persuaded to follow suit. This will help establish a universal AI arms control regime regarding the use of AI for military purposes.

Generative AI and Deterrence Stability: A Scenario for De-escalation

It is the night of May 28, 2028, at India's military command center. A high-ranking Indian military officer sends a message to a subordinate commander; Pakistan has initiated a preemptive strike. Prepare for a retaliatory strike. This is not a drill.

Flight time between India and Pakistan is a few minutes. Hence, commanders do not have time for verification and need to decide quickly whether to launch a retaliatory strike. He calls the command center to verify the message received.

India's prime minister is contacted immediately concerning the message received by one of his commanders. He picks up the hotline to Pakistan's prime minister and asks, "Have you launched nuclear weapons?"

Pakistan's prime minister responds, "Not at all. This is fake!"

India's prime minister informs the military command center. The crisis is averted, and nuclear weapons are not used.

Confidence-building measures, which were present in the second scenario, have the opportunity to change the outcome of a conflict. Given the potential for artificial intelligence to be used in nefarious ways, it is time to take action to avert such a crisis.

Ms. Anum A. Khan is an Associate Director at the Center for International Strategic Studies (CISS) Islamabad. She is also currently a Research Fellow at Comprehensive Test Ban Treaty Organization and a Project Associate of The Third Nuclear Age Project led by Prof. Andrew Futter. Views expressed are the author's own.