

ASIAN ANGLE SUNDEEP WASLEKAR

# A nuclear war started by artificial intelligence sounds like sci-fi. It isn't



WE ARE IGNORING a spectre on the horizon. It's the spectre of a global nuclear war triggered by artificial intelligence. UN Secretary General Antonio Guterres has warned of it. But so far nuclear-weapons states have avoided talks on this cataclysmic threat.

They argue that there is an informal consensus among the five biggest nuclear powers on the "human in the loop" principle. None of the five say they deploy AI in their nuclear-launch command systems. This is true but misleading.

They use AI for threat detection and target selection. AI-powered systems analyse vast amounts of data from sensors, satellites and radars in real time, analyse incoming missile attacks and recommend options for response. The human operators then cross-check the threat from different sources and decide whether to intercept the enemy missiles or launch retaliatory attacks.

Currently, the response time available for human operators is 10 to 15 minutes. By 2030, it will be reduced to between five and seven minutes. Even though human decision-makers will make the final call, they will be swayed by the AI's predictive analytics and prescriptions. AI may be the driving force behind launch decisions as early as the 2030s.

The problem is that AI is prone to errors. Threat-detection algorithms can indicate a missile attack where none exists. It could be due to a computer mistake, cyber intrusion or environmental factors that obscure the signals. Unless human operators can confirm the false alarm from other sources within two to three minutes, they may activate retaliatory strikes. The use of AI in many civilian functions such as crime prediction, facial recognition and cancer prognosis is known to have an error margin of 10 per cent. In nuclear early-warning systems, it could be around 5 per cent. As the precision of image-recognition algorithms improves over the next decade, this margin of error may decline to 1-2 per cent. But even a 1 per cent error margin could initiate a global nuclear war.

The risk will increase in the next two to three years as new agentic malware emerges, capable of worming its way past threat-detection systems. This malware will adapt to avoid detection, autonomously identify targets and automatically compromise them.

There were several close calls during the Cold War. In 1983, a Soviet satellite mistakenly detected five missiles launched by the United States. Stanislaw Petrov, an officer at Sepukhov-15 command centre, concluded that it was a false alarm and did not alert his superiors who could launch a counter-attack. In 1995, Olenegorsk radar station detected a missile attack off Norway's coast. Russia's strategic forces were placed on high alert and then-president Boris Yeltsin was handed the nuclear briefcase. He suspected it was a mistake and did not press the button. It turned out to be a scientific rocket. If AI had been used for



ELON MUSK AND U.S. PRESIDENT DONALD TRUMP IN THE WHITE HOUSE'S OVAL OFFICE. Photo: AFP

determining the response in both situations, the outcome could have been disastrous.

Currently, hypersonic missiles use conventional automation rather than AI. They can travel at speeds of Mach 5 to Mach 25, evade radar detection and manoeuvre their flight path. Major powers are planning to enhance hypersonic missiles with AI to locate and instantly destroy moving targets, shifting the kill decision from humans to machines.

There is also a race to develop artificial general intelligence, which could lead to AI models operating beyond human control. Once this happens, AI systems will learn to augment and replicate themselves, taking over decision-making processes. When such an AI is integrated into decision-support systems for nuclear weapons, machines will be able to initiate devastating wars.

Humans have perhaps five to 10 years before algorithms and plutonium could reduce us to skeletons and skulls.

We need a comprehensive agreement among major powers to mitigate this risk, going beyond the

reiteration of the "human in the loop" slogan. Such an agreement must include transparency, explainability and cooperation measures; international standards for testing and evaluation; crisis-communication channels; national-oversight committees; and rules to prohibit aggressive AI models capable of bypassing human operators.

Geopolitical shifts have created an unexpected opportunity for such a treaty. Leading AI experts from China and the United States were involved in several track-two dialogues on AI risks, resulting in a joint statement by former US president Joe Biden and Chinese President Xi Jinping in November.

Elon Musk is a staunch advocate for the need to save humanity from the existential risks posed by AI. He may urge current President Donald Trump to transform the Biden-Xi joint statement into a pact.

It would require Russia to get on board. Until January of this year, Russia had refused to discuss any nuclear-risk reduction measures, including the convergence with AI, unless the Ukraine issue was on the table. With Trump engaging Russian President Vladimir Putin in dialogue aimed at improving bilateral relations and ending the Ukraine war, Russia may now be open to discussions.

The question is who will bell the cat. China may be able to initiate trilateral negotiations. Neutral states including Turkey and Saudi Arabia could also pave the way. This is a historic opportunity to make a breakthrough and save humanity from extinction. We must not let it go to waste for lack of political vision, courage and statesmanship. ■

*Sundeep Waslekar is President of Strategic Foresight Group, an international think tank, and the author of 'A World without War'.*

*This article was first published by the Asian Peace Programme, an initiative to promote peace in Asia, housed in the NUS Asia Research Institute.*

BELOW: PROTESTERS HOLD AN ANTI-NUCLEAR RALLY IN MELBOURNE, AUSTRALIA THIS MONTH. Photo: AFP

