Will Artificial Intelligence Undermine Nuclear Deterrence?
History Tells Us That's Unlikely

By: Rohin Sharma

Arguably, the balance of power and the threat of mutually assured destruction is what kept the peace between the U.S. and Soviet Union during the Cold War. Despite being in conflict across the globe, both sides knew a potential escalation could destroy their respective societies. The perception (even unfounded) that an adversary could massively retaliate likely prevented a larger scale conflict.

In the age of Artificial Intelligence (AI), however, does the same calculus exist? Could AI set the conditions where a would-be belligerent might perceive they have a considerable advantage? Could this translate into aggressive action by that belligerent which would undermine the deterrent balance? What can be done to mitigate the risk of this possible scenario?

This article will present the case that AI will not impact traditional theories of nuclear deterrence. While AI may play a role in conventional conflict, it is unlikely that the strategic nuclear balance of power will be significantly altered. Countries that historically relied on a minimum deterrent (India, China, and North Korea), may find their arsenals somewhat vulnerable, however, the risks regarding preemptive strikes would still be exceedingly high. As a corollary, AI will also impact nuclear command, control, and communications (NC3), although whether this will be destabilizing remains to be seen.

This article will begin by briefly summarizing traditional elements of deterrence theory. It will then discuss the impact of AI on each of these facets, showing that they are unlikely to undermine models of mutually assured destruction. The article will then cover framework for how AI can be rationalized to ensure strategic stability. The article will conclude with potential recommendations for emerging nuclear powers.

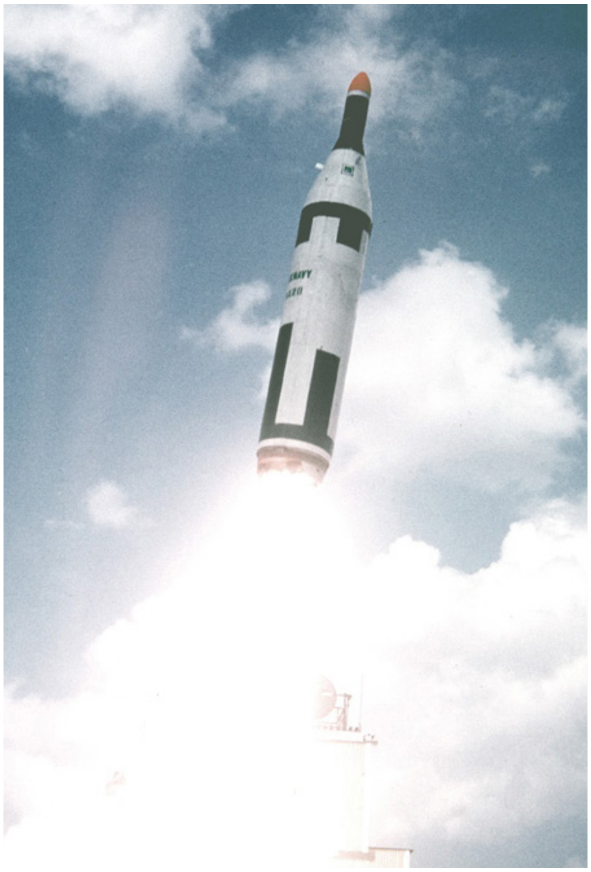
Tenants of Nuclear Deterrence

The early nuclear age ushered in the myth of the nuclear revolution which asserted that nuclear weapons would significantly alter warfare.[1] Due their destructive capability, even minor states that possessed them would be able to deter larger states from asserting their will. The aspects of the Melian Dialogue, where strong states do what they can and weak states do what they must, were no longer applicable.[2] A minimum deterrent was asserted by numerous theorists at the start of the Cold War, including Robert Oppenheimer, and advanced by Mao Zedong and various Indian leaders after they detonated their first devices.[3,4] This doctrine promoted the idea that nuclear weapons would decrease the number of conflicts and there was no need for an expensive deterrent such as those developed by the US and Russia.

However, advances in technology, even in the 1950s, undermined this doctrine. During the late 1950s, leaders in both Moscow and Washington feared that their nuclear weapons program was susceptible to a damaging first strike, even with nascent intelligence, surveillance and reconnaissance systems. Based on these technologies, theorists surmised that an adequate deterrent had several requirements.

SURVIVABLE SECOND-STRIKE FORCES

One of the most important aspects of a nuclear force is to have second strike capability. Although doctrines differ, second strike forces give a nuclear power the ability to sustain an initial nuclear or conventional attack, and still provide a capable retaliatory effort. For a second strike capability to



ABOVE: CAPE CANAVERAL, Fla. (undated photo) – An unarmed Polaris A1 Missile successfully completes an early launch at the Cape. Today, the Sea-Based Strategic Deterrence mission continues. The current generation of submarine launched ballistic missile—the Trident II D5/D5LE—along with the Intercontinental ballistic missile and nuclear capable bombers create the nation's strategic nuclear deterrence capability. (U.S. Navy Photo/Released)

be considered credible, a nuclear state must give the perception that they can retaliate and cause unacceptable damage even after being attacked.

Analysts fear that technological advancements might lead an aggressor to pre-emptively strike an opponent's critical forces. While this has not happened with nuclear weapons, there are historical precedents with conventional weapons for this happening, to include Japan's attack on the US fleet at Pearl Harbor in 1941 or Israel's attack on the Egyptian Air Force at the start of the Six Day War in 1967. In this case, advanced technologies, strategic surprise, and poor defensive measures, were exploited to destroy an adversary's strategic assets. There is a fear that AI and other technologies could lead to a repeat of history, but this fear is likely unfounded.

NUCLEAR COMMAND AND CONTROL

In addition to survivable nuclear forces, any nuclear force must have a robust system for command and control. NC3 "is the exercise of authority and direction, through established command lines, over nuclear weapon operations by the President as the chief executive and head of state."[5] Elements of a robust NC3 system include defined decision-making at the top that allows for the National Command Authority to implement nuclear plans quickly and effectively. Moreover, NC3 consists of a system of early warning radars and satellites to alert of a potential nuclear strike. In addition, all these elements must be linked through a series of unclassified and secure communications channels. The command-and-control system is designed to provide rapid response in the event of a nuclear escalation.

Pre-Artificial Intelligence-Advances in Technology Undermine Nuclear Deterrent

While it's tempting to say that AI will have changed the previously stated dynamics regarding nuclear deterrent, analysts throughout history have (often incorrectly) made the same assessments regarding the technologies of their time. During the late 1950s, Albert Wohlsetter believed the US was in danger of nuclear decapitation and advocated increasing the number of strategic weapons to counter this vulnerability.[6] The theorist Bernard Brodie went further when, in 1959, he argued that "the supreme advantage of the initiative in launching an unrestricted thermonuclear war can hardly be contested, for the side possessing hope, reasonably in some circumstances."[7]

While Americans feared that there could be a decapitation strike, this analysis was based on inflated intelligence estimates on Russian capabilities. The shock of Sputnik, not unlike what happened to Iraq WMD estimates in the wake of 9/11, forced much of the intelligence to be politicized, implying that Soviet missiles had a range and accuracy that they never had. The fear of the missile gap caused a rapid increase in US nuclear weapons numbers, inflating capacity to beyond what was necessary.[8]

In 2006 and 2007, Keir Lieber and Darryl Press, Professors at Georgetown and Dartmouth, presented arguments for technology undermining nuclear deterrence. They made the case that advances in technology presented three issues that allowed for a first strike (either by the US or a potential adversary).

First, precision strike capabilities—far more capable than seen during the First Gulf War-would allow targeting with an accuracy not previously possible.[9] Second, advances in Intelligence, Surveillance, and Reconnaissance (ISR) would allow a nation state to identify and target infrastructure critical to a nuclear program. Finally, low yield nuclear weapons would allow for a precision strike without the collateral damage typically associated with higher yield nuclear weapons.[10] A nuclear weapons state with the capacity to accurately attack an adversary's critical infrastructure using a low yield nuclear weapon might be tempted to do so.

While Lieber and Press present a viable argument for an adversary that possesses a minimum deterrent nuclear arsenal, their theories lack credibility when applied to a robust nuclear power. First, they overemphasize how effective ISR systems are in detecting nuclear weapons programs (delivery systems, warheads, and storage). As of September 2017, the US did not have an accurate number or the locations of North Korea's nuclear weapons.[11] If ISR platforms cannot provide sufficient information on their own to account for quantity and locations of North Korea's nuclear weapons, it seems even less likely that it could be used to maintain accurate information about more sophisticated nuclear programs such as those in Russia and China. Secondly, small modifications to a nuclear program such as increasing mobility or launch on warning, could nullify a precision strike capability. Finally, the authors underestimate the political implications a nuclear weapons strike will have. Even a low yield device would be put pressure on a foreign leader to respond in kind, leading to a potentially dangerous escalation. Any robust nuclear power is unlikely to consider a first strike against another nuclear adversary since the conditions described by Lieber and Press, simply do not exist.

The Impact of Artificial Intelligence Within Deterrence Models

AI is the latest technological advancement that may impact the deterrence models. AI can accelerate ISR and precision strike capabilities, manipulating mass data sets to accurately ascertain the location and disposition of an adversary's nuclear weapons capabilities. This may give the perception that an adversary can conduct a pre-emptive strike, destroying an adversary's nuclear assets in one blow. Furthermore, AI can also serve as a contemporary "Dead Hand," allowing for the executions of NC3 in the event of a decapitating strike.

INTEGRATION OF AI INTO INITIAL STRIKE

The largest impact AI will have on nuclear deterrence is to seamlessly integrate the Observe, Orient Decide, Act (OODA) loop. This will give a nuclear power the perception they can pre-emptively target an adversary's nuclear forces in one strike without fear of retaliation. There is fear that this could increase instability, increasing the probability that an aggressor would want to launch a pre-emptive attack.

Artificial intelligence could support an aggressor in two ways. First, AI could coordinate the thousands of sensors necessary for precision strike, even using predicative analysis to determine where certain assets may be located. Arguably, AI enabled sensors, combined with historical patterns, could accurately predict the location of submarines, mobile missile launchers, and nuclear capable aircraft, enabling states to maintain high confidence predictions of the locations of an adversary's nuclear arsenal.

Secondly, AI could be used to bring precision to strike hardened facilities. Underground assets (whether in silos or in mountains) have always been seen as a way to ensure a second-strike capability. However, AI can provide modeling data to undermine the protection, providing mapping and historical data to target supposedly survivable forces.

While these arguments appear to have merit, they do not stand up to scrutiny. The models of deterrence will still stand even with the introduction of AI. Historically, even advanced militaries with the best technology available have had difficulty spotting an adversary's strategic assets. During the Vietnam War, advancements in ISR and satellite technology did not spot the mobile surface-to-air missiles that proved so devastating.[12] Likewise, during the Gulf War, further advances in ISR could not spot the mobile SCUD missiles that were in an open desert.[13]

Even in the current era, the issue of concealment is still an issue. Second-rate Houthi rebels have hidden their missiles within a small "box" and frustrated superior Saudi forces.[14] Lebanese Hezbollah has missile capabilities that Israel, even with the use of advanced imagery and AI, cannot successfully target.[15] Finally, as of 2018, US intelligence officials have stated they don't know the numbers and locations of North Korean nuclear weapons, despite the DPRK having a modest program.[16]

MAINTAINING DETERRENCE IN THE AGE OF AI

AI will have a minimal effect on deterrence models for several reasons. First, the sensor amalgamation described in this paper is still subject to human and technological fallacies. Even the best sensor will not be able to determine the difference between a conventional or nuclear missile in flight, making any discrimination impossible. Furthermore, like the mobile surface-to-air missiles and the mobile Scuds, a potential adversary can use a system of decoys to undermine the overhead sensors, making them unable to determine what is a legitimate target.[17]

In addition, even with an advanced AI algorithm, it is unclear whether any political leader would engage in a risky preemptive strike. When targeting even a rudimentary nuclear power it is unlikely that a democratic head of state would order/initiate an attack based on the recommendations/output of an AI algorithm. For a Western leader to attack North Korea, which already has an established ICBM, the potential retaliatory effects would likely be too great. When applied to a South Asia scenario, where both India and Pakistan have semimature programs, the risk of a pre-emptive strike is further heightened. The risk of retaliation increases drastically when you start at look at the advanced Chinese or Russian nuclear programs.

AI AND NUCLEAR COMMAND AND CONTROL

While a pre-emptive nuclear strike is unlikely, AI can have a negative impact on NC3, particularly on launch on warning doctrines. Launch on warning emphasizes using nuclear weapons, preventing a decapitating preemptive strike. In order to execute a launch on warning posture, a nuclear power would need early warning satellites, the ability to communicate orders from an authority figure, and 24-hour alert for nuclear forces.[18]

Nuclear powers may see AI as a link in this process. During the Cold War, both sides considered solutions to communicate with distributed forces if the central authority was unable to provide direction. Emerging nuclear powers may see AI as a way to negate the need for a human in the loop, thereby enabling detection, analysis, and launch, all without human involvement. This pre-delegation of authority to algorithms would eliminate the human component from nuclear command and control.

However, any use of AI would incur massive risk. During the Cold War, decades prior to modern AI, there were numerous examples of erroneous nuclear near misses. During the Carter administration, a software error mimicking a potential Russian attack

BELOW: Several Pershing II missiles are prepared for launching at the McGregor Range at White Sands Missile Test Range, New Mexico.[24]

nearly drove the U.S. to conduct a pre-emptive attack.[19] For their part the Russians believed NATO Able Archer exercise, conducted in 1983, masked a potential nuclear strike against the Soviet Union.[20] The machines and procedures of their day nearly lead to world to nuclear confrontation.[21]

In the Artificial Intelligence era, the risks are magnified significantly. A potential nuclear power might use pre-delegation authority to computer systems, taking the human out of the loop, similar to how the Soviets uses the "perimeter" system in the event of decapitation.[22] Even AI systems could misread outside signals, confusing an incoming conventional missile strike with a nuclear strike, or even a flock of birds as something more ominous.[23]

Recommendations for AI in the Nuclear Era

1. Do not rely on AI for pre-emptive strikes: As mentioned earlier, AI will likely not give an advantage to conduct a preemptive strike. History has shown that technologies have their weaknesses and AI is no different. Even for small nuclear powers like North Korea, where there are still unknown details about their program, the probability of eliminating an adversary's nuclear capability in one strike is miniscule. Therefore, since AI only enhances predictions (or something similar) more quickly than in the past, it would be foolish for a nuclear power to attempt a first strike on another nuclear power.

2. Build second strike forces: For any potential nuclear force, the introduction of AI reinforces the need for survivable second-strike forces, preferably relying on mobility (mobile missiles and submarines). While quantity and dispersion may make fixed assets difficult to preemptively strike, they still may be a tempting target for a would-be aggressor. Mobility will likely make it impossible for a pre-emptive strike, even with advanced sensors. The difficulties in doing so would even deter the most ardent aggressor.

3. Re-evaluate Launch on Warning Doctrines: By far, the largest risk of nuclear weapons in the age of AI is the use nuclear weapons without the human in the loop. Some nuclear powers might be tempted to use AI to launch nuclear weapons faster in the event a leadership decapitation, increasing the risk considerably. However, through either treaty or unilateral changes, nuclear powers should still consider removing some of their nuclear assets off launch on warning. At the very least, removing AI from current nuclear powers should be considered as part of any future arms control negotiations or confidence building measures.

Conclusion

For decades, theorists have posited that advances in technology would undermine traditional deterrence models. Today, there are many who view AI as a new development that will upset existing deterrence theory. However, like previous advances in technology, AI will unlikely alter the deterrence dynamics. Furthermore, keeping the "human in the loop" will assist in effective nuclear command, control, and communication. Leaders, particularly of nuclear states, must recognize that the challenges that existed before AI, continue to exist now that we have AI. The United States can lead in this effort by emphasizing caution and through continued research. ■

Rohin Sharma

is currently a policy analyst and strategist at the U.S. Army CWMD and Nuclear Agency as well as an adjunct professor at Georgetown University. He holds an M.A. in Security Studies from Georgetown and B.A. from Johns Hopkins. He held previous intelligence positions at the Department of the Army, Defense Threat Reduction Agency, and Department of Homeland Security. He is a former Army Officer and veteran of Operation Iraqi Freedom and Operation Enduring Freedom.

Notes

1. Kier Lieber and Daryl Press, "Power Politics in the Nuclear Age," chap. 1 in *The Myth of the Nuclear Revolution: Power Politics in the Atomic Age*, (Ithaca: Cornell University Press, 2020).

2. Keith Payne, "US Nuclear Weapons and Deterrence," *Air and Space Power Journal*, August 2015, https://www.airuniversity.af.edu/Portals/10/ASPJ/journals/Volume-29_Issue-4/V-Payne.pdf.

3. Christopher Ford, "Thinking About China's Nuclear Weapons Policy," New Paradigms Forum, August 9, 2024, https://www.newparadigmsforum.com/thinking-about-chinas-nuclear-weapons-policy.

4. Rajesh Rajagopalan, "India's Nuclear Doctrine Debate," US Nuclear Weapons and Deterrence, Carnegie Endowment for International Peace, June 30, 2016, https://carnegieendowment.org/research/2016/06/indias-nuclear-doctrine-debate.

5. Deputy Assistant Secretary of Defense for Nuclear Matters, "Nuclear Weapons Employment, Planning, Policy, and NC3," chap. 2 in Nuclear Matters Handbook 2020 [Revised] (Washington D.C.: Department of Defense, 2006), https://www.acq.osd.mil/ncbdp/nm//NMHB2020rev/chapters/chapter2.html.

6. Albert Wohlstetter, "The Delicate Balance of Terror" RAND, 1958, https://www.rand.org/pubs/papers/P1472.html.

7. Gordon A. Craig and Felix Gilbert, *Makers of Modern Strategy from Machiavelli to the Nuclear Age*, ed. Peter Paret, (Princeton: Princeton University Press, 1986), 709.

8. Arms Control Association, "The Missile Gap Myth and Its Progeny." Arms Control Today, November 5, 2011, https://www.armscontrol.org/act/2011-05/missile-gap-myth-and-its-progeny.

9. Kier Lieber and Daryl Press, "The Rise of U.S. Nuclear Primacy," *Foreign Affairs*, March 1, 2006, https://www.foreignaffairs.com/articles/united-states/2006-03-01/rise-us-nuclear-primacy.

10. Kier Lieber and Daryl Press, "The New Era of Counterforce: Technological Change and the Future of Nuclear Deterrence." *International Security*, March 2007, https://direct.mit.edu/isec/article/41/4/9/12158/The-New-Era-of-Counterforce-Technological-Change.

11. David Sanger, "How Would U.S. Verify That North Korea Is Disarming?" *New York Times*, March 9, 2018, https://www.nytimes.com/2018/03/09/us/politics/trump-north-korea-nuclear-missiles.html.

12. Mark Carlson, "Operation Spring High: Thuds vs. SAMs," HistoryNet, October 23, 2019, https://www.historynet.com/operation-spring-high-thuds-vs-sams/.

13. Thomas A. Keaney and Eliot A. Cohen, *Gulf War Air Power Survey: Weapons, Tactics, and Training*, (Washington D.C.: U.S. Government Printing Office, 1993) 285, https://media.defense.gov/2010/Sep/27/2001329801/-1/-1/0/AFD-100927-061.pdf.

14. Seth G. Jones, Jared Thompson, Danielle Ngo, Joseph S. Bermudez Jr., and Brian McSorley, "The Iranian and Houthi War against Saudi Arabia," Center for Strategic and International Studies, December 21, 2021, https://www.csis.org/analysis/iranian-and-houthi-war-against-saudi-arabia.

15. Tal Beeri, "Hezbollah's Missile and Rocket Arrays: A Tough and Complex Challenge for Israel," Alma Research and Education Center, May 6, 2021, https://israel-alma.org/2021/05/06/hezbollahs-missile-and-rocket-arrays-a-tough-and-complex-challenge-for-israel/.

16. Sanger, "How Would U.S. Verify."

17. James Johnson, "Rethinking Nuclear Deterrence in the Age of Artificial Intelligence," Modern War Institute, January 28, 2021, https://mwi.westpoint.edu/rethinking-nuclear-deterrence-in-the-age-of-artificial-intelligence/.

18. George Washington University National Security Archive, "The Launch on Warning: Nuclear Strategy and its Insider Critics," June 11, 2019, https://nsarchive.gwu.edu/briefing-book/nuclear-vault/2019-06-11/launch-warning-nuclear-strategy-its-insider-critics.

19. Daryl Kimball, "Nuclear False Warnings and the Risk of Catastrophe," Arms Control Today, December 2019, https://www.armscontrol.org/act/2019-12/focus/nuclear-false-warnings-and-risk-catastrophe.

20. Center for Arms Control and Non-Proliferation, "The Soviet False Alarm Incident and Able Archer 83," October 14, 2022, https://armscontrolcenter.org/the-soviet-false-alarm-incident-and-able-archer-83/.

21. Jaganath Sankaran, "A Different Use For Artificial Intelligence In Nuclear Weapons Command And Control," War on the Rocks, April 25, 2019, https://warontherocks.com/2019/04/a-different-use-for-artificial-intelligence-in-nuclear-weapons-command-and-control/.

22. Adam Lowther and Curtis McGiffin, "America Needs a Dead Hand More than Ever," War on the Rocks, March 28, 2024, https://warontherocks.com/2024/03/america-needs-a-dead-hand-more-than-ever/.

23. Bryce Farabarg, "Bad Idea: Integrating Artificial Intelligence with Nuclear Command, Control, and Communications," December 3, 2019, Center for the Strategic and International Studies, https://defense360.csis.org/bad-idea-integrating-artificial-intelligence-with-nuclear-command-control-and-communications/.

24. "Several Pershing II missiles are prepared for launching at the McGregor Range," National Archives Catalog, NAID: 6424504, https://catalog.archives.gov/id/6424504.