ADVANCE SOCIAL SCIENCE ARCHIVE JOURNAL
Available Online: https://assajournal.com
Vol. 5 No. 01 Jan-Mar 2026.Page#. 6538-6548
Print ISSN: 3006-2497 Online ISSN: 3006-2500
Platform & Workflow by: Open Journal System

Artificial Intelligence, Nuclear Deterrence, and Strategic Stability in South Asia: A Security Dilemma Perspective

Dr. Muhammad Naveed Ul Hasan Shah
Assistant Professor, Department of Political Science and International Relations, University of Central Punjab, Lahore, Pakistan
muhammad.naveed1@ucp.edu.pk
Izna Zaman Khattak
BS International Relations, Department of Political Science and International Relations, University of Central Punjab, Lahore, Pakistan
iznazamanofficial@gmail.com
Muhammad Bilal Malik
BS International Relations, Department of Political Science and International Relations, University of Central Punjab, Lahore, Pakistan
bilalmalikir@gmail.com

Abstract
Artificial Intelligence (AI) is increasingly transforming the concept of nuclear deterrence in one of the most volatile strategic equations in the world the India-Pakistan nuclear dyad. The article explores the changing South Asian perceptions of vulnerability and resolve through the power of Artificial Intelligence (AI) based intelligence, surveillance and reconnaissance (ISR) systems, autonomous weapons systems, machine assisted command and control systems and cyber-AI convergence. Based on Robert Jervis' security dilemma theory, the article concludes that AI is not just putting new weapons into the arsenal, but at the same time deepening the structural insecurity of the region's high concentration of actors, locations, and doctrine, which are separated only by a narrow margin. When one State invests in the development of the "early warning" or "counterforce" targeting using artificial intelligence, it is interpreted as an offensive and threatening act by the other State, which creates a vicious circle of mistrust, diminishing the time for decision-making and increasing the chance of unintentional or accidental escalation. The analysis synthesizes the international academic research on the topic of AI and nuclear risk with the unique characteristics of South Asian deterrence short missile ranges, sub-conventional warfare, and doctrinal opacity. Finally, policy recommendations on 'confidence building measures restraint in AI-nuclear entanglement' and 'bilateral and trilateral risk-reduction' dialogue between India, Pakistan and China are given.

Keywords: *Artificial Intelligence, Nuclear Deterrence, Strategic Stability, Security Dilemma, South Asia, India-Pakistan Relations, Autonomous Weapons, Command and Control*

Introduction

The potential impact of the new technology on the risk of nuclear war is not a new one, but becomes uniquely relevant when it comes to South Asia. The relationship between India and Pakistan is not bipolar like the one that existed during the early Cold War era, but has the following features: geographic contiguity, an unresolved territorial dispute over Kashmir, a significant conventional military imbalance in India's favour, and a history of both sides using force that is below the nuclear threshold. The consequence of such structural features is that innovations that could affect the speed, accuracy or readability of military operations have

impacts on nuclear stability that diverge in some important respects from their counterparts in more institutionally developed or geographically dispersed nuclear relationships.

Nuclear deterrence in South Asia has always been a thin fence that holds up of mutual vulnerability, doctrinal confusion and close geography. Even with the acquisition of their nuclear weapons in 1998, India and Pakistan have engaged in repeated sub-conventional and conventional wars at their border; and the missile flight time is measured in minutes, not the half hour or more that characterized the relationship between the Cold War superpowers. Artificial Intelligence is now entering this already fragile environment, with its potential to speed up the process of merging intelligence, enhance targeting accuracy and make military command systems more resilient, but also with the risk of shrinking the window for decision making, blurring the lines between conventional and nuclear systems, and creating new opportunities for miscalculation.

The Defence Artificial Intelligence Council, the Defence AI Project Agency and other initiatives are clearly a part of the national effort to apply machine learning to various aspects of defense such as surveillance, logistics and decision-making. In response, Pakistan has established its own institutional investments, such as a special centre for artificial intelligence and computing and enhanced development and collaboration with China in the field of unmanned systems. The result is not a technological parity, but a parity of learning; each modernization is read as a proof of hostile will by both sides, and the logic of both sides' reading is very close to the logic of the security dilemma, which was first formulated in the international relations literature almost half a century ago.

This article poses a particular question: what implications does the militarization of artificial intelligence have for strategic stability and nuclear deterrence between India and Pakistan and which theory(s) best explains the dynamics? It suggests that the security dilemma theory developed by Robert Jervis, initially to explain how defensive military measures can be interpreted as offensive and thus fuel unintended spirals of hostility, is the most parsimonious and empirically grounded theory to explain the destabilising impact of AI in South Asia. The article goes through a number of stages. It first conducts a review of literature on AI and nuclear risk, both at a global and regional level. Then it develops the theoretical framework, and applies it systematically to four categories of AI-nuclear entanglement command, control and communications; intelligence, surveillance and reconnaissance, and counterforce temptations; autonomous weapons and automation bias; and cyber-AI convergence.

Literature Review

The number of studies on the subject of nuclear stability and AI has increased significantly in the last ten years, but they are still mainly concentrated on the United States, China, and Russia. In subsequent work, the authors considered norms for “transparency” and “restraint” as confidence-building tools that could help mitigate the dangers of AI-powered military systems even in cases where disarmament may be impractical (Horowitz & Scharre, 2021).

One stream of research has been concerned with the aspects of coercion and signaling in automation. Empirical survey research has discovered that nuclear threats supported by automated systems can have more credibility, precisely because they imply less human control over the launch decision, with worrying consequences for crisis bargaining. Concerning the application of automation and automation bias in national security domains, related work has investigated the impact of automation bias on decision-making and how rapidly human judgment is likely to be supplanted during a crisis (Horowitz & Kahn, 2024). Other arguments have been put forth by Michael Horowitz, who has suggested that lethal autonomous weapons

systems have a primary impact on deterrence and stability through altering “how” wars are fought and making war “faster and more reversible.”

The multi-year study by the Stockholm International Peace Research Institute (SIPRI) on AI, strategic stability and nuclear risk also found that the impacts of AI are not necessarily stabilizing, but depend critically on how states incorporate machine learning into their existing nuclear systems (Boulanin et al., 2020). The potential for nuclear war to rise with the advent of AI is less likely to be caused by intentional first use but more likely by accident, false alarm, and loss of assured second-strike capability (Geist & Lohn, 2018). Kenneth Payne has described anticipated shifts to strategic affairs brought about by artificial intelligence as an "AI revolution" that will be as big as past changes in military technology, but has warned that this new form of cognition will have different consequences than human judgment and the principles of classical deterrence theory (Payne, 2018).

Several scholars also have studied the individual technological sub-components of this transformation. Jürgen Altmann and Frank Sauer investigated the potential of autonomous weapon systems to disrupt strategic stability by shortening engagement time frames relative to human engagement (Altmann & Sauer, 2017). Paul Scharre's report on the world's march toward battlefield autonomy chronicled how militaries are already using machine-speed targeting systems with logic that is increasingly outside of human decision cycles. Elsa Kania has suggested that AI competition between great powers is creating an emerging security dilemma scenario whereby states are competing to not be left behind in a field of applications whose military implications are not fully understood (Kania, 2018). More recent commentary has focused on the institutional choices that states must make regarding what level of authority to vest in machines - rather than the intrinsic merit of any single algorithm - are necessary for nuclear stability in an AI era (Depp & Scharre, 2024); and a trilateral study of the United States, China, and Russia has called for a renewed arms control dialogue which explicitly considers the intersection of nuclear weapons and artificial intelligence (Stokes et al., 2025).

The literature is sparse in South Asia, but is expanding. Recent scholarship has suggested that the strategic assurance dilemma is likely to arise in the future: How to make sure that the adversary does not become less convincing about its good intentions than it is about its ability to deter (Warren & Ganguly, 2026)? Some analysts have already looked into what potential disinformation or synthetic communications may be generated using generative AI tools that could lead to unintended escalation between the two states. A recent study that is specific to Pakistan places the country's emerging-technology landscape in a wider framework of post-nuclear-deterrence, suggesting that AI, lethal autonomous weapons, and related technologies pose dangers and opportunities to Pakistan's strategic position (Aurangzeb et al., 2026). This article extends and enriches the growing body of literature by using the security dilemma, a theory used in international relations, to streamline the study of the impact of AI on South Asian strategic stability.

Literature in the specific context of South Asia has tended to be descriptive and qualitative, using open-source defense publications, policy statements, and comparative analysis of cases as the basis for the analysis, rather than the quantitative or experimental approaches that are starting to emerge in the global literature on AI and nuclear risk. This article builds on the dominant qualitative approach common in the regional literature yet attempts to bring theory into the mix by following a single, well-specified international relations framework in a systematic manner. In so doing, it seeks to shift the focus of the South Asia-specific literature from mere description of the capabilities of AI to a more general explanation of why such capabilities have destabilising effects, which have been repeatedly observed by the region's analysts.

Theoretical Framework: Jervis's Security Dilemma

This article uses the following theoretical framework: The security dilemma theory put forward by Robert Jervis in his seminal article "Cooperation Under the Security Dilemma," 1978. In an anarchic international system without central authority to guarantee state security, what one state does to boost its own security sometimes has an unintended consequence of reducing the security of others, who then take steps to counterbalance the threat, setting off a spiral of mutual insecurity even if no state is aggressively inclined (Jervis, 1978). One of the key concepts in Jervis's model is the separation between military offensive and defensive postures and the difference between what offensive and defensive capabilities look like to an outsider. In the countries where there is no clear separation between offense and defense and where offense seems to be able to gain some advantage, the security dilemma becomes most pronounced as rational, purely defensive investments are indistinguishable from preparation for offense.

Jervis later applied the same reasoning specifically to the nuclear arena, suggesting that the logic of the security dilemma was weakened under particular conditions by the nuclear revolution: the volume of costs associated with war make it appear irrational to conquer under those circumstances (Jervis, 1989). However, Jervis emphasized that this stabilizing revolution was not inevitable; it required reliable retaliation forces, unambiguous signals and confidence in the survivability of one's own arsenal. Even in a nuclear relationship, the security dilemma logic is in effect if these conditions are not there or not accepted – in fact, they are not always in place in South Asia.

This article uses this framework to specifically explore the case of AI-enabled India-Pak military modernization, given that AI is systematically exacerbating the two conditions Jervis called “problem” and “balance” that need to be addressed: the offense-defense distinguishability problem and the offense-defense balance itself. The targeting data for offensive counterforce operations is also produced by AI-powered ISR systems, which are ostensibly defensive in nature, and used for early warning and situational awareness, making it hard for an adversary to tell if it is a defensive investment or if its systems are being prepared for offensive first-strike operations. Likewise, AI-powered decision cycles change the ratio of offense versus defense in favor of the offense, as the process of “fast” and “preemptive” decisions will come to outweigh “deliberate” ones, consistent with the predictions of Jervis' framework, which suggests arms races and crisis instability will grow. This article continues with the security dilemma logic by considering how each of these specific technologies is being used in the introduction of AI into the military systems of the South Asian region as an empirical instantiation of the security dilemma mechanism that Jervis described.

Artificial Intelligence and Nuclear Command, Control and Communication.

Nuclear command, control, and communications (NC3) systems are the institutional and technical infrastructure that supports and facilitates political authorization, transmission, and execution of nuclear launch decisions. Historically, however, states have been hesitant to entrust machines to perform more than advisory roles in the nuclear chain of command, with automation at the forefront of these systems playing a role, especially in early warning functions, which warn human operators of incoming missiles, and in communications infrastructure, which sends launch codes to delivery platforms.

The allure of the use of AI in NC3 for India and Pakistan is mostly due to the fact that the subcontinent has a compressed geography, with policy makers having mere minutes to differentiate between a real strike and a false alarm, technical malfunction or a conventional skirmish that's mistaken for a nuclear launch. Advocates say the technology of machine-speed data fusion can eliminate the ambiguity that has led to perilous "close calls" in the past. Critics

say that embedding opaque machine-learning systems into chains of decisions over systems whose outcomes have civilizational consequences could be as much an uncertainty introduced as human error.

A security dilemma lensneeds to ask the question of how is the upgrade of AI-powered NC3 seen by the adversary, and not only if it boosts the security confidence of the one performing the upgrade. If India invests in machine-assisted early warning to lessen its vulnerability to a Pakistani first strike, and if the very same investment is seen as shaking India's reliance on only retaliatory postures, then Pakistan has a reason to interpret the same investment as evidence of India's move toward more usable and faster reacting nuclear options. Jervis recognized that this is the essence of the security dilemma, the "offense-defense indistinguishability" problem, and this is exactly what has been reproduced in the digital world. The Belfer Center's recent definition of a strategic assurance dilemma encapsulates this process exactly: AI advances weaken the credibility of deterrence and undermine the ability of states to give each other the "assurances" necessary to persuade the other side that investments are not forfeited to aggression.

AI-Enabled ISR and Counterforce Temptations in the India-Pakistan Dyad

The second, more closely related, impact of AI on South Asian strategic stability involves intelligence, surveillance and reconnaissance. Machine-learning algorithms, based on satellite imagery, signals intelligence, and open-source data, greatly enhance a state's ability to find an adversary's mobile missile launchers, submarines, and other elements of its nuclear arsenal, which have traditionally been able to survive on the basis of concealment and mobility. In their seminal paper, Christopher Clary and Vipin Narang (2019) have outlined how India has built up a capability package that, in their analysis, allows India to arguably have the technical ability to adopt a doctrine of disarming Pakistan's nuclear arsenal even if there is no political decision to pursue it.

This is where the power of artificial intelligence comes into play. These prior generations of surveillance technology relied on a complex human analysis process to translate raw imagery to actionable targeting data, while machine-learning systems are now capable of merging a variety of intelligence streams within seconds or minutes of a detection, significantly reducing the time between the strike and the detection. Perceived loss of the survivability of Pakistan's nuclear arsenal is doctrinally destabilising because the country's nuclear doctrine has always been built on a mixture of ambiguity and an asymmetric escalation strategy aimed at dissuading conventional aggression by India through a threat of early nuclear use (Narang, 2010). If Pakistani planners decide that Indian AI-based ISR significantly increases the vulnerability of Pakistan's mobile nuclear forces, they may well be tempted to adopt a doctrine that involves pre-delegating launch decision making and/or speeding up the transition from peacetime to alert status, both of which increase the possibility of potential unauthorized or accidental launch.

This is a security dilemma in its truest form: when India sees an investment as an increase in battlefield awareness and deterrence through denial, Pakistan sees it as proof of a first strike capable counterforce posture and vice versa. This interpretation is bolstered by the broader literature on the growing difficulty of assured retaliation, which suggests that technological advances in precision and surveillance have expanded the list of states whose nuclear forces are becoming more vulnerable to disarming strikes, not only in theory but in practice as well, and that AI will only further increase this vulnerability (Lieber & Press, 2017). The asymmetry further exacerbates the issue as India's significantly bigger economy and advanced technology indigenous industry enables it to adapt to AI faster across its ISR framework, whereas Pakistan's lack of such indigenous capability forces it to be more dependent on its external partners,

especially China, for the same, thereby creating its own reliance and delays, which further increase the Pakistani threat perception in the interim.

Understands the concepts of Autonomous Weapons, Automation Bias, and Crisis Instability.

In addition to command systems and intelligence gathering, the increasing number of weapons platforms that can operate autonomously or semi-autonomously creates another dynamic source of instability in a crisis. Lethal autonomous weapons systems, from loitering munitions to autonomous naval and aerial platforms, are being increasingly talked about by military planners as systems that can operate at machine speed, the very thing that military planners value as a way to circumvent the deliberation time of humans. But it is this quickening of the decision-making process that has worried the crisis stability scholars. In the absence of sufficient human oversight mechanisms, lethal autonomous systems may prove to be faster than political leaders can be, thereby allowing them to start an escalating conflict prior to the political leader's willingness to intervene (Horowitz, 2019).

An often-overlooked threat is automation bias—the human predisposition to rely on algorithmic decisions despite contrary evidence or personal opinion. Empirical studies conducted in the human-machine decision-making field in national security have found that people prefer to take decisions based on AI just when they are time-constrained, which is exactly what the situation is when India-Pakistan relations go into crisis mode. The deployment of autonomous systems that can operate at machine speed risks to further shrink the narrow margins that have traditionally existed for de-escalatory diplomacy in South Asia, from the Kargil crisis of 1999 to the Balakot crisis of 2019, and from the extreme situation of the two states in 2025.

In Paul Scharre's assessment of how autonomous weapons technology has spread around the world, it is the overall degradation of the human buffer that has historically provided space for crisis management and time to correct misperception that poses the strategic dangers of these systems (Scharre, 2018). Introducing autonomous systems by either India or Pakistan, even if for purely conventional or sub-conventional reasons, like counterterrorism operations along the Line of Control, is hard for the other side to categorise accurately as defensive or otherwise because of the documented entanglement between conventional and nuclear force structures in both states' postures. What occurs is a mutually suspicious relationship that extends beyond the narrow scope of nuclear forces to the whole spectrum of conventional-nuclear deterrence in South Asia.

The new field of cyber-AI convergence and nuclear vulnerability

Another aspect of AI-nuclear entanglement is the integration of AI into the realm of offensive cyber. Cyber tools based on machine learning are now used to better identify vulnerabilities in adversary networks, automate the reconnaissance stage of cyber operations, and to a lesser extent (but more speculatively) produce synthetic communications that could fool human operators. There have been specific cautions concerning the use of generative AI (GAI) tools by an adversary or a third party in a worst case scenario for creating credible but fake messages from military commanders, with potentially catastrophic consequences for a military command structure that takes the word of internal communications for granted (Khan, 2024).

Beyond these extreme examples, the intermingling of nuclear and non-nuclear command systems on shared or adjacent networks can lead to what scholars have termed an inadvertent escalation risk: a cyber attack targeting simply conventional military targets, but due to shared infrastructure or algorithmic misclassification, or both, might appear as an attack on nuclear command systems themselves, which could elicit responses that are far more severe than the threat actually presented (Schwartz & Horowitz, 2025). South Asian analysts who have analyzed these dynamics have found that the South Asian region's militarisation of AI, on top of existing

doctrinal ambiguity and low warning periods, could be a source of destabilising strategic dynamics and reducing the effective nuclear threshold even if there is no malicious escalatory intent on the part of either government (PolicyEast, 2026).

The security dilemma approach that is highlighted here makes cyber-AI convergence perhaps the most extreme example of the indistinguishability of offense and defense, since cyber and AI-enabled weapons are often made two-purpose by design, and are developed for and deployed to legitimate defensive network security applications, but also have implicit offensive applications that cannot be fairly recognized by an external party. Both countries have an incentive to assume the worst of the other's motives, since India is not being allowed to prove to Pakistan that its efforts in this field are defensive, and vice versa; in fact, the same spiral dynamic that Jervis identified a close to half a century ago, around the arms stalemate between conventional and nuclear weapons, applies here too.

Synthesis: AI as an Accelerator of the South Asian Security Dilemma

Combined with the aforementioned analysis, this article argues that artificial intelligence is not the root cause of insecurity in South Asia, but rather a catalyst for a greater security dilemma process that is older and precedes the technology itself. It is important to note that the security dilemma features of South Asia – contiguous borders, historical Kashmir disputes, recurring sub-conventional conflict, and doctrinal postures on both sides that incorporate capabilitarian deterrence with warfighting options – were already well-discussed in the pre-AI literature on the nuclear stability of South Asia (Ganguly & Kapur, 2010). AI brings efficiency and opacity: it shortens the time for the kind of careful signaling and reassurance that Jervis argued was the key remedy to security dilemma dynamics and it also makes it more difficult for both sides to confirm the defensive purpose of the other's technological investments.

This synthesis also explains why the South Asian scenario is different from the Cold War scenario that initially led Jervis to develop his nuclear revolution thesis in crucial ways. Once assured second strike, the security dilemma among nuclear powers was reduced because neither side could be expected to be able to foresee retaliation as a result of a first strike and so could not rationally assume it was not possible. However, as illustrated in the Indian case, the strength of the new targeting paradigm of AI for ISR and counterforce is that it seems to provide exactly the escape from assured retaliation that the nuclear revolution was believed to have closed (Narang, 2014). When, correctly or not, one side believes that AI has given it a significant advantage in terms of spying and striking back before the other side can do the same, the logic of mutual assured destruction becomes more fragile, and the region returns back to the more dangerous, prenuclear-revolution dynamics that Jervis's original 1978 framework sought to explain. This is a warning that can also be drawn from organizational theory: even when political leaders have a healthy sense of the risks of nuclear weapons, the military institutions that operate highly automated systems have long been concerned more with reliability and readiness than with avoiding accidents and unauthorized uses that might lead to failed control (Sagan, 1993), a point that is echoed directly in recent Pakistan-focused scholarship on emerging military technologies. The security dilemma dynamic does NOT need to occur with malicious intent from either side. Indian defense planners can feel confident that their efforts to invest in AI are only for defensive and deterrence purposes, and Pakistani defense planners can feel the same way about their own efforts against Indian investment, without any obstructions, knowing that both are objectively making the regional environment less stable. The key takeaway from Jervis's model is that a lack of communication or diplomatic goodwill between the parties is not the reason for a security dilemma, and that solutions contingent on better information or more generous diplomacy are unlikely to resolve such a dilemma if the competition is genuine. The same point can be

reinforced by command and control considerations: Peter Feaver has written early on about the importance of the institutional architecture that governs use-control, the combination of ensuring weapons can be used when authorized and ensuring they cannot be used when not authorized: weaker controls will tend to be pushed by rising perceived threats, and weaker controls will be more conducive to stability, as India and Pakistan are likely to push for looser controls at the moment when they feel they are in greater danger (Feaver, 1992).

The case illustrates the compressed crises and the logic of automation

The theoretical discussion presented above can be further substantiated or perhaps even better tested, with regard to the pattern of crises between India and Pakistan since 1998, when the two countries became explicit nuclear powers. The Kargil conflict in 1999 proved that limited conventional war between the two nations was not impossible even with the use of nuclear weapons and the military standoff in 2001-2002 after attack on the Indian Parliament proved that even without an exchange of fire, the mobilisation on one side could trigger a reciprocal mobilisation on the other, and create a lengthier and expensive military stand-off. The entire process of provocation, retaliation and de-escalation was compressed in a few days in the Balakot crisis of 2019, which began in the wake of a suicide attack in Pulwama, India's aerial attack inside Pakistani territory, and an aerial engagement that followed.

A first glimpse of what an AI-fueled version of this pattern might look like comes from the events of the last six months in 2025, when the two states fought a far more intense battle in which missile and drone strikes were unleashed on military facilities on both sides before a ceasefire was finally reached. According to early open-source reports, both sides used an array of AI-powered air defense cueing, unmanned aerial systems, and speedy multi-domain surveillance, adding to the already brief timeline of the crisis and leaving diplomats and international mediators with an extremely narrow window to intervene before the situation could have spiraled beyond the realm of conventional warfare. None of this evidences that the crisis was exacerbated as it was through the use of AI; any direct causal statements are beyond the evidence. The episode does suggest, however, as the security dilemma logic outlined in this article would suggest, that once a crisis begins, the speed of the escalation in confrontation is likely to increase, reducing the timeframe available for the political leaders of both sides to engage in the "slow, steady and soothing" signaling for which Jervis credited them with breaking security dilemmas.

This pattern is also a striking example of why South Asia is a challenging test case for the broader literature on AI-nuclear stability, which has been developed for the more geographically and institutionally dispersed relationship between the United States, China, and Russia. Unlike most nuclear relations between the major powers, which offer hours of warning, and decades of legacy infrastructure around crisis communications, the India-Pakistan relationship does not provide either luxury. The arrival of such machine-speed sensing and responding technologies in a relationship that already has a compressed timeline, a lack of territorial resolution, and regular sub-conventional violence, is a tougher case for stability than the great power scenarios that dominate the existing literature, making the case for South Asia as analytically distinct and a pressing policy priority for both scholars and policy makers.

The implications and recommendations of the policy are as follows:

The security dilemma is the best lens through which to understand the potential for instability from AI in South Asia, and the policy response should reflect on the structural uncertainties as they exist, and not just the technology itself. The following are recommendations that can be made based on this analysis.

Transparency Measures

India and Pakistan, preferably with the involvement of China, as a nuclear armed neighbor and a major technology provider to Pakistan, should discuss a structured dialogue to establish what is acceptable when it comes to the use of AI in nuclear border systems. Some degree of transparency about the aims of investments in AI-powered ISR and early warning could help alleviate the issue of indistinguishability between the offense and the defense identified by Jervis as an issue in the security dilemma cycles, without the need of either side to reveal technical information. It is important that there is a level of restraint between the human and the machine command authority.

There should be clear doctrinal statements in both states that will ensure an effective human in the loop when it comes to any decision involving the nuclear release, and that this is not a matter of operational pressure to commit to automation for the sake of getting a decision done quickly. These pledges would serve as "expensive signals of restraint" that would correspond to the reassurance mechanisms Jervis identifies as being essential for breaking cycles of security dilemmas – and which are not completely verifiable.

Crisis Communication Hotlines in the Age of AI.

Current bilateral crisis communication protocols between the Indian and Pakistani Armed Forces must be brought to current standards to ensure that protocols and procedures are in place to deal with the issues pertaining to AI introduced in the field of crisis communication, such as establishing a robust system to verify the authenticity of communications in a timely fashion in case of suspected synthetic media manipulation.

Engage in a regional and Multilateral approach

The international discourse on the subject of AI and nuclear risk has largely been limited to the great powers, leaving South Asian countries entirely out of the picture. The gap could be bridged if multilateral institutions and research organisations deliberately start to integrate South Asian views into global norm-setting on military artificial intelligence, to ensure that any new international norms for this technology are adjusted to the specific characteristics of South Asian nuclear dyads, rather than great powers.

Limitations and future directions

There are certain caveats in this analysis that leave space for some fruitful future studies. First, a lot of the empirical evidence on the role of AI in Indian and Pakistani military systems is classified or simply unavailable for outsiders to review, so this article's conclusions about the speed and extent of AI-nuclear involvement are based on open-source data, official statements of intent, and inference instead of firsthand observation of operational systems. Second, the framework of the security dilemma used here is satisfactory for explaining why an investment using AI would lead to mutual insecurity, but not as good as other tools from crisis bargaining and formal deterrence modeling would be for providing a more exact picture of the sequence or timing of the subsequent crisis. Third, the article has focussed on India and Pakistan as the main units of analysis, but the involvement of China as a nuclear armed neighbor of India and a major technology supplier to Pakistan, creates a true trilateral dynamic, which is not treated as an exogenous variable in future studies. Lastly, the technology is constantly evolving, and any assessment of AI's impact on South Asian strategic stability should be considered provisional and subject to periodic reevaluation as the technology keeps developing, as do the doctrines that apply to its use.

Conclusion

The application of AI in South Asia military forces is not only destabilising, but also mediating the extent of its impact due to the pre-existing structural conditions of the India-Pak nuclear relationship. This article has suggested that the security dilemma theory of Robert Jervis is the

most plausible explanation for why AI, even when proponents of both sides envisioned it as a secure and efficient instrument, on a practical level, has contributed to exacerbation of the security dilemma in the region. In each of the areas of command and control, intelligence collection and counterforce targeting, autonomous weapons, and cyber-AI convergence, the same pattern emerges: the investments in technology that seem to be defensive are seen as offensive by the other side, prompting them to respond in ways that make them more vulnerable on the objective level than before the other side made the defensive investment.

The implications of correct analysis go beyond the scholarly literature on international relations theory. Two nuclear stockpiles with hundreds of warheads apiece, both based on a common land border with a long history of conventional warfare and sustained sub-conventional hostility over Kashmir inhabit South Asia. In contrast to the arms control regimes established between NATO and the Warsaw Pact over the decades of the Cold War between the U.S. and the Soviet Union, there is no such bilateral arms control architecture, via arms control treaties, no such verification regimes and no such crisis communication channels between India and Pakistan. The institutional context in which this new set of AI-powered military systems is being deployed is thin, and if the technology has a destabilising tilt, this is far more likely to be exacerbated by the absence of pre-existing systems for managing risks and vulnerabilities compared to a more institutionalised nuclear relationship.

This is not the end of the story of the destabilising potential of AI in South Asia. Jervis has outlined a set of paths by which the security dilemma spirals can be broken, most importantly by the high cost of signals of restraint, by the development of measures that make attacks easier to distinguish from defenses, and by the continuous diplomatic process that can replace the worst case scenario of reassurance. The decision of India, Pakistan and their external partners to pursue these routes, or to instead let AI be used to drive military competition towards the Jervis model of a relatively stable, although tense, “Great Powers” equilibrium, or towards the earlier model of the classical security dilemma, will greatly shape the nuclear future of South Asia. In this environment, with more than 1.5 billion people living within range of the potential nuclear exchange between India and Pakistan, the academic and policy communities have an urgent need to further develop theoretical and empirical insights into this fast-moving field. A clear idea of theory, of why AI is likely to exacerbate the security dilemma, rather than solve it, is a first step in designing measures that can disrupt it before it leads to a catastrophic miscalculation – and theory in itself is not enough to make the region's current lack of an architecture to reduce risks.

References

Altmann, J., & Sauer, F. (2017). Autonomous weapon systems and strategic stability. Survival, 59(5), 117–142.

Aurangzeb, M., Uddin, S. S., & Shafiq-un-Nisa. (2026). Emerging military technologies, artificial intelligence and strategic stability in South Asia (2010–2025): Challenges and opportunities for Pakistan in the post-nuclear deterrence era. Journal of Social Sciences and Humanities, 65(1), 75–102.

Boulanin, V., Saalman, L., Topychkanov, P., Su, F., & Carlsson, M. P. (2020). Artificial intelligence, strategic stability and nuclear risk. Stockholm International Peace Research Institute.

Clary, C., & Narang, V. (2019). India's counterforce temptations: Strategic dilemmas, doctrine, and capabilities. International Security, 43(3), 7–52.

Depp, M., & Scharre, P. (2024, January 16). Artificial intelligence and nuclear stability. War on the Rocks.

Feaver, P. D. (1992). Command and control in emerging nuclear nations. International Security, 17(3), 160–187.
Ganguly, S., & Kapur, S. P. (2010). India, Pakistan, and the bomb: Debating nuclear stability in South Asia. Columbia University Press.
Geist, E., & Lohn, A. J. (2018). How might artificial intelligence affect the risk of nuclear war? RAND Corporation.
Horowitz, M. C. (2019). When speed kills: Lethal autonomous weapon systems, deterrence and stability. Journal of Strategic Studies, 42(6), 764–788.
Horowitz, M. C., & Kahn, L. (2024). Bending the automation bias curve: A study of human and AI-based decision making in national security contexts. International Studies Quarterly, 68(2), 1–15.
Horowitz, M. C., & Scharre, P. (2021). AI and international stability: Risks and confidence-building measures. Center for a New American Security.
Horowitz, M. C., Scharre, P., & Velez-Green, A. (2019). A stable nuclear future? The impact of autonomous systems and artificial intelligence (arXiv:1912.05291). arXiv.
Jervis, R. (1978). Cooperation under the security dilemma. World Politics, 30(2), 167–214.
Jervis, R. (1989). The meaning of the nuclear revolution: Statecraft and the prospect of Armageddon. Cornell University Press.
Johnson, J. (2023). AI and the bomb: Nuclear strategy and risk in the digital age. Oxford University Press.
Kania, E. (2018, October 29). The AI titans' security dilemma. Hoover Institution.
Khan, A. A. (2024, September 23). Generative artificial intelligence and deterrence stability between India and Pakistan. Global Security Review.
Lieber, K. A., & Press, D. G. (2017). The new era of counterforce: Technological change and the future of nuclear deterrence. International Security, 41(4), 9–49.
Narang, V. (2010). Posturing for peace? Pakistan's nuclear postures and South Asian stability. International Security, 34(3), 38–78.
Narang, V. (2014). Nuclear strategy in the modern era: Regional powers and international conflict. Princeton University Press.
Payne, K. (2018). Artificial intelligence: A revolution in strategic affairs? Survival, 60(5), 7–32.
PolicyEast. (2026, February 9). AI in South Asian strategic stability: Opportunity or destabiliser?
Sagan, S. D. (1993). The limits of safety: Organizations, accidents, and nuclear weapons. Princeton University Press.
Scharre, P. (2018). Army of none: Autonomous weapons and the future of war. W. W. Norton.
Schwartz, J., & Horowitz, M. C. (2025). Out of the loop: How dangerous is weaponizing automated nuclear systems? arXiv.
Stokes, J., Kahl, C., Kendall-Taylor, A., & Lokker, N. (2025, February). Averting AI Armageddon: U.S.-China-Russia rivalry at the nexus of nuclear weapons and artificial intelligence. Center for a New American Security.
Warren, S., & Ganguly, Š. (2026, July 10). Strategic assurance dilemma: AI, strategic stability, and nuclear crises in South Asia. Belfer Center for Science and International Affairs.