『広島平和科学』46 (2025) pp. 57-79 ISSN0386-3565

Hiroshima Peace Science 46 (2025)

核兵器システムへの新興技術の導入―抑止・軍備管理への影響と課題

戸﨑　洋史

広島大学　平和センター

Emerging Technologies for Nuclear Weapons Systems: Implications and Challenges for Deterrence and Arms Control

Hirofumi TOSAKI

The Center for Peace, Hiroshima University

Abstract

The introduction of emerging technologies into ISR (Intelligence, Surveillance and Reconnaissance) and NC3 (Nuclear Command, Control and Communications) could help stabilize deterrence relationships by enhancing the operational speed and information-processing capabilities of nuclear weapons systems. At the same time, however, it also has the potential to destabilize those relationships through strengthened damage-limitation capabilities and greater complexity in decision-making processes. Furthermore, introduction of emerging technologies would increase the risk of inadvertent nuclear use due to, inter alia, misinterpretation, miscalculation or malfunction. To prevent these adverse impacts on deterrence and nuclear risks from becoming a reality, it is essential—before such emerging technologies become more widely integrated into ISR and NC3—to first recognize technology as an "intermediary variable" that can work either to the benefit or detriment of security. This calls for a thorough, multifaceted discussion of how nuclear-armed states plan to utilize these technologies, under what objectives and security strategies, and what operational rules will govern their use both in peacetime and during crises. Building upon these discussions, steps such as clarifying the specifics of "human-in-the-loop" arrangements, designing systems with redundancy and robust cyber defenses, and introducing confidence-building measures will be crucial. Moreover, even amidst the stagnation or regression of arms control and intensifying strategic competition, ongoing efforts—including dialogue and discussion on the use of emerging technologies in nuclear weapons systems, drafting codes of conduct, and implementing confidence-building and transparency measures—could not only help avert inadvertent escalation but also serve as a foothold for the broader revitalization of nuclear arms control. In this regard, exploring potential avenues for dialogue and collaboration on arms control and risk reduction in the context of emerging technologies is an urgent task.

はじめに[1]

通常戦力や新興技術（emerging technology）の急速な発展に伴う核・非核の「絡み合い（entanglement）」は抑止や軍備管理を不安定化させる可能性があると2018年にアクトン（James M. Acton）が論じて以降[2]、その論点の1つとして、核兵器システムへの新興技術の導入が抑止および軍備管理に及ぼしうる含意が活発に議論されてきた。この間、フュッター（Andrew Futter）およびザラ（Benjamin Zala）は、核兵器の国際場裡への登場という第一次核時代から、核兵器の（特に地域諸国への）一層の拡散と核関係の多極化の進行という第二次核時代を経て、戦略的非核兵器とこれを可能にする技術の開発が、特に安定性とリスクの問題に関連して核抑止関係を劇的に変化させる第三次核時代（Third Nuclear Age）に入りつつあるとも論じた[3]。

「新興技術」に統一的な定義はないが、安全保障の文脈ではひとまず、「まだ成熟していない、あるいは広く普及していないものの、国際の平和と安全に大きな──そしておそらく破壊的な──影響を及ぼすと予想される技術、科学的発見および技術的応用」[4]という定義が、その特質を的確に踏まえている。具体的には、たとえば米国は2018年11月に、国家安全保障上重要な（また輸出管理を含む経済安全保障の対象となりうる）新興技術として14のカテゴリーを挙げた[5]。核兵器問題との関連で特に近年注視されているのが、核兵器システムの構成要素のなかで情報収集・警戒監視・偵察（ISR）および核指揮・統制・通信（NC3）に人工知能（AI）や量子技術といった新興技術が導入されていく場合の抑止関係に及ぼしうる影響である。

本稿では、第一に、ISRおよびNC3への新興技術の適用可能性を概観する。第二に、それらが抑止関係を安定化させる可能性、また第三に不安定化させる可能性をそれぞれ分析する。第四に、核兵器システムへの新興技術の導入が誤解・誤認や事故などによる意図せざる核兵器使用のリスクを高めかねないのと懸念について検討する。第四に、リスク低減や軍備管理の可能性について考察する。

1. 新興技術とISR／NC3

核兵器不拡散条約（NPT）上の核兵器国をはじめとする核保有国、ならびに少なくとも主要な非核兵器国は、その軍事システムへの新興技術の導入に、極めて高い関心を示してきた[6]。他方で、核兵器

---

[1] 本稿は、トヨタ財団助成事業「先端技術と国際関係」研究会（2020～2021年度）での研究、ならびに拙稿「新興技術と核抑止関係」『研究レポート』（日本国際問題研究所）2021年3月30日、https://www.jiia.or.jp/research-report/post-87.html を骨子としつつ、その後の動向や研究なども踏まえ、改めて書き下ろしたものである。

[2] James M. Acton, “Escalation through Entanglement: How the Vulnerability of Command-and-Control Systems Raises the Risks of an Inadvertent Nuclear War,” *International Security*, Vol. 43, No. 1, Summer 2018, pp. 67-82.

[3] Andrew Futter and Benjamin Zala, “Strategic Non-Nuclear Weapons and the Onset of a Third Nuclear Age,” *European Journal of International Security*, Vol. 6, No. 3, August 2021, pp. 257-277.

[4] Marina Favaro, Neil Renic and Ulrich Kühn, “Negative Multiplicity: Forecasting the Future Impact of Emerging Technologies on International Stability and Human Security,” *IFSH Research Report*, No. 10, June 2022, p. 17.

[5] バイオテクノロジー、AI／機械学習技術、測位技術、マイクロプロセッサ技術、先進コンピューティング技術、データ分析技術、量子情報／センシング技術（量子コンピュータ、量子暗号など）、輸送・補給関連技術（logistics）、不可製造技術（3Dプリントなど）、ロボティクス（マイクロドローン、スワーム技術など）、脳コンピュータインターフェース、極超音速、先端材料、および先端監視技術。

[6] 日本も、2022年12月に閣議決定された「国家防衛戦略」で、「先進的な技術に裏付けられた新しい戦い方が勝敗を決する時代において、先端技術を防衛目的で活用することが死活的に重要となっている。この際、総合的な

システムにおける ISR や NC3 に関しては、核保有国の具体的な目的や構想・計画、あるいはそれらの進捗状況は明らかではない。たとえば、AI については、5 核兵器国は、意思決定支援、信頼できる安全な通信チャネルの維持、膨大なデータ群を迅速に処理することによる状況認識の強化、包括的な脅威評価のための分析ツールとしての活用を構想しているものの[7]、米国、ロシアおよび中国のいずれも、技術的に信頼できる AI モデル／システムを開発できていないことが導入の制約になっているとされる[8]。核保有国はいずれも核兵器システムの近代化を積極的に推進しており、そのなかで弾道ミサイルや極超音速ミサイルといった運搬手段とともに重視されているのが ISR および NC3 であるが、新興技術については技術的な特性を踏まえて、以下のような活用可能性が指摘されてきた。

(1) ISR

核保有国が核攻撃を遂行する際には、核攻撃が必要か否かなど状況を把握し、核攻撃を実施するとの意思決定を行い、これを部隊に伝達するというプロセスを経る。ISR が担うのは状況の把握であり、核兵器の使用やこれに関連しうる情報の収集、核弾頭やその運搬手段といった敵対国の核戦力の探知・追尾、敵対国の行動や（特にミサイル）攻撃に対する早期警戒、ならびに自国の核攻撃による目標達成状況や敵対国の核攻撃による被害結果を把握することなどが含まれる。

たとえば、量子センシング技術などによりリモートセンシング技術・関連技術（センサー、データ通信、分析など）が発展すれば、敵対国の攻撃能力に対する探知・追尾能力が向上する。豪州、英国および米国は豪英米安全保障協力（AUKUS）協定を通じて、AI が水中音響信号やソナーデータの処理を高速化し、より高速かつ正確に中国の潜水艦を追跡する方法を研究している[9]。短・中期的な時間枠での実現は容易ではないとされつつも、磁力計、重力勾配計および量子時計など量子技術の応用は、核戦力のなかで最も残存性が高い潜水艦戦力の探知能力を向上させるとの分析もある[10]。また、クラウドコンピューティング、高速・大容量データ通信、AI などの活用は、早期警戒システムや探知・追尾システムなど各種センサーから収集される膨大な情報の効率的な集約・処理と迅速・高精度な解析・分析を可能にし、敵対国の状況や行動を格段に明確かつ（ニア）リアルタイムで把握することに寄与する。これらにより、核攻撃が進行する状況での適切で効果的な対応の可能性を高めることができる。

また、AI を活用したドローン群による戦略的作戦として、分散した移動式ミサイル発射機とこれに

防衛体制の強化のための府省横断的な仕組みの下、防衛 省・自衛隊のニーズを踏まえ、政府関係機関が行っている先端技術の研究開 発を防衛目的に活用していく。また、防衛産業を活用しつつ、スタートアップ等各種企業、各種研究機関の研究開発の成果を早期の実装化につなげていく取組を実施することとする」とした。

[7] Alice Saltini, “AI and Nuclear Command, Control and Communications: P5 Perspectives,” European Leadership Network, November 2023, p. 20。また、AI の軍事システムへの統合に対する核保有国の関心については、Heiko Borchert, Torben Schütz and Joseph Verbovszky, eds., *The Very Long Game: 25 Case Studies on the Global State of Defense AI*, Springer 2024 なども参照。

[8] Vladislav Chernavskikh, “Nuclear Weapons and Artificial Intelligence: Technological Promises and Practical Realities,” *SIPRI Background Paper*, September 2024, pp. 9-10.

[9] Sydney Freedberg, Jr., “Transparent Sea: AUKUS Looks to AI, Quantum in Hunt for Chinese Submarines,” *Breaking Defense*, January 29, 2024, https://breakingdefense.com/2024/01/transparent-sea-aukus-looks-to-ai-quantum-in-hunt-for-chinese-submarines/.

[10] Katarzyna Kubiak, “Quantum Technology and Submarine Near-Invulnerability,” European Leadership Network, December 2020, pp.5-9.

付随する通信・指揮・統制・情報（C3I）システムの位置を特定・追跡すること、敵対国の防空システム、ミサイル防衛あるいは対潜水艦戦力（ASW）を大量のドローンによって無効化すること、さらには無人水中艦艇（UUV）、無人水上艦艇（USV）、および AI 対応のスウォーム通信・ISR システムを搭載した無人航空機（UAV）を攻防両面で同時に配備し、敵の防御を飽和させて潜水艦の位置を特定し、無力化・破壊することなども挙げられている[11]。

偵察・監視能力の向上は、敵対国による先制・奇襲攻撃の準備をより早期に把握するのに寄与する。また、AI を用いた異常検知技術などにより早期警戒システムの能力が向上すれば、奇襲攻撃や先制攻撃を含め敵対国による（核）攻撃に対して、その到来を早期に警告することで、攻撃を受ける国は移動可能な核戦力（移動式ミサイル、爆撃機、潜水艦）を移動させる時間を得ることができるかもしれない[12]。さらに、AI 技術による情報処理と状況認識の強化（多様なデータソースを収集・統合）は、核戦力部隊の動き、補給線、その他の情報を先制的に分析し、核の脅威が出現する前にそれを予測するという「予測的予報（predictive forecasting）」を実現する可能性も指摘されている[13]。

(2) NC3

核兵器の使用にかかる意思決定や情報伝達を司る NC3 については、核攻撃が想定される高い緊張状況、あるいはすでに核戦争が勃発するなど高度に複雑で切迫する状況において、そうした状況の判断、核兵器の使用の是非、使用する場合の核攻撃目標、使用する核兵器の割り当てなどに関する選択肢を、先端的な AI などの導入によって迅速、適切、正確かつ効率的に指導者に提供できれば、意思決定に大きく寄与する。上述のような厳しい状況での AI による適切な支援は、実際の脅威と誤報の識別可能性を高めるとともに、認知バイアスやヒューマンエラー、あるいは誤報などによる核兵器の不要・不適切な使用の可能性を低減するものとなろう[14]。また、量子通信の実用化などによる通信能力の向上は、核戦争のような苛烈な環境下でも、センサーから指導者を経て各部隊に至る情報や命令の安全で信頼性が高く、確実な送受信を可能にするとされる。

さらに、新興技術の発展により、核兵器発射の自動化システムの導入を試みる核保有国が出てくる可能性も指摘されている。ソ連が 1985 年に導入したペリメトル（米欧では「デッドハンド」とも称される）半自動報復システムは、ソ連本土における核攻撃に伴う事象（核爆発の振動など）を検知するとともに、軍事司令部からの返答がない場合、核報復を決定できる指導者が核攻撃などで不在になったと判断し（返答がある場合、報復攻撃を命令できるハイレベルの意思決定者が依然として生存し、意思決定を完全にコントロールしていると判断する）、核弾道ミサイルの発射指令を下すというシス

---

[11] James Johnson, “Artificial Intelligence: A Threat to Strategic Stability,” *Strategic Studies Quarterly*, Vol. 15, No. 1, Spring 2020, p. 23.

[12] Benjamin Zala, “Should AI Stay or Should AI Go? First Strike Incentives & Deterrence Stability,” *Australian Journal of International Affairs*, Vol. 78, No. 2 (2024), p. 159.

[13] Alice Saltini, “AI and Nuclear Command, Control and Communications,” p. 12

[14] Michael C. Horowitz, Paul Scharre and Alexander Velez-Green, “A Stable Nuclear Future? The Impact of Autonomous Systems and Artificial Intelligence,” Cornell University, December 2019, p. 14.

テムである[15]。

現在までに、そうした自動化システムを保有する国はロシア以外には確認されていないが、新興技術の発展によって信頼性の高いシステムの構築の実現可能性が高まれば、より自動化を高めたシステムの導入に関心を持つ国が出てくるかもしれない。たとえば北朝鮮は、2022 年 10 月に公表した「核使用法令」で、「国家核戦力の指揮統制システムが敵対勢力の攻撃によって危険に瀕する場合」は、「事前に決定された作戦方案に従って、攻撃地点（starting point）や指揮部をはじめとして、敵対勢力を壊滅させるための核打撃が自動的かつ即時に断行される」[16]とも記載された。北朝鮮は自動化システムについては言及していないものの、朝鮮半島での武力衝突において、米国が指導者に対する斬首攻撃（decapitation）を敢行する可能性があり、また米国は 2022 年の核態勢見直し（NPR）などで、「北朝鮮による米国やその同盟国、パートナーに対するいかなる核攻撃も容認できず、政権の終焉という結末になる。金体制が核兵器を使用して生き残るシナリオはない」[17]とも繰り返しており、北朝鮮がそうした状況でも米国・同盟国に核報復攻撃を可能にする自動化システムに関心を有しているとしても不思議ではない。

(3) 導入の誘因と作用

核兵器システムへの新興技術の導入が抑止関係に与える影響については、核保有国の「技術進歩の速度、新興技術が既存の軍事力と作戦概念に統合される方法、新興技術間の相互作用、国家政策と国際法が新興技術の開発、統合および利用を可能にしたり阻害したりする程度などといった多くの要因の関数になるため、予測は不可能ではないにしても困難である」[18]と指摘されている。ISR や NC3 への新興技術の導入は、主として核兵器システムの「神経系」の能力を高めるものであり、核兵器を最も強力な抑止力たらしめている圧倒的な破壊力という特質に直接的に作用するわけではない。しかしながら、ISR や NC3 は核兵器使用の決定に直結するサブシステムであり、核態勢、抑止能力あるいは抑止関係にその観点から大きな影響を与えうるからこそ、核保有国は下記のような観点から新興技術の導入に関心を有してきた。

たとえば、意思決定にかかる速度における優位性である。自国への（核）攻撃に対して、意思決定の観点からも迅速に報復する能力を構築することで、敵対国に対する抑止力を高めることができる。また、核発射手順の自動化により、敵対国に対して強制的な優位性が得られると考える核保有国もありうる[19]。意思決定にかかる速度の向上は、新興技術の導入による情報の精密性・適時性の向上とも相まって、対兵力打撃（counterforce）による損害限定（damage limitation）の実現可能性を高めうる。

---

[15] ペリメトルについては、David E. Hoffman, *The Dead Hand: The Untold Story of the Cold War Arms Race and Its Dangerous Legacy*, Anchor Books, 2009 などを参照。

[16] "Law on DPRK's Policy on Nuclear Forces Promulgated," September 9, 2022, https://kcnawatch.org/newstream/1662687258-950776986/law-on-dprks-policy-on-nuclear-forces-promulgated/.

[17] United States, *Nuclear Posture Review*, 2022, p. 12.

[18] Kelley M. Sayler, "Emerging Military Technologies: Background and Issues for Congress," *CRS Report*, July 17, 2020, pp. 23-24.

[19] Horowitz, Scharre and Velez-Green, "A Stable Nuclear Future?" pp. 4-5, 13.

核兵器システムへの新興技術導入にかかる誘因の高低は、各国の目的や取り巻く状況にも影響される。たとえば、第二撃能力に一定の自信がある国は、新興技術がもたらし得る利点よりもリスクの両面を重視し、その導入には慎重になる可能性があるのに対して、第二撃能力の脆弱性に懸念を持つ国は、核戦力が無力化される前の使用を確実にするために、新興技術がもたらす効果を重視し、その導入に積極的になる可能性がある[20]。

また、現有の報復能力の残存性や報復攻撃の実施可能性に懸念を有する国（指揮・統制システムや指導者の無能力化への懸念を含む）は、核兵器発射自動化システムの利点を重視するかもしれない[21]。自動化システムの導入に対する関心の程度は、政権のタイプ、政治的安定性と正統性、特定の核保有国の脅威認識にも影響を受けるとされる。核保有国のなかでも権威主義政権は、クーデターや外部からの干渉を懸念し、中央集権的な核の指揮・統制構造を採用するとともに、核兵器使用権限を事前に部隊指揮官に委譲することに消極的であるとの背景から、民主主義政権よりも自動化システムに関心を持つ可能性がある。特に権威主義政権にとっては、指導者以外に核兵器使用決定の権限を委譲する必要がない（指導者が信頼するごく一部の関係者のみに関与を限定できる）ことに加えて、指導者が機能不全の際に自動的に核報復が可能になることで、指導者の殺害を狙った攻撃（断首攻撃）を抑止する可能性を高めること、あるいは自動化による潜在的なリスク（特に、この決定に関連する倫理的、人間の認知的、道徳的な課題）にはあまり関心を持たないであろうことなどから、自動化システムへの親和性が高いとも論じられている[22]。

抑止を構成するのが能力、意図および（被抑止国の）認識であるとすれば、新興技術の導入による抑止国の核兵器システムの能力の変化は、核兵器の使用にかかる意図にも影響を及ぼすだけでなく、能力と意図の変化が被抑止国の認識にも作用する。そうした認識の変化は、被抑止国が自国の核兵器システムにかかる能力、あるいは核兵器使用についての意図を変化させ、このことがさらに抑止国の認識に変化をもたらすという、作用・反作用が生じる可能性もある。その作用・反作用は、導入の途上、さらには導入前であっても関心を有しているとみなされた時点で、（抑止関係に大きな影響を及ぼしうるからこそ）生じかねない[23]。そして、そうした作用・反作用において留意すべきは、核兵器システムへの新興技術の導入が抑止関係に及ぼす影響や変化は、一方向に働くとは限らないということである。以下で見ていくように、抑止関係の安定化・不安定化のいずれにも向かいうるだけでなく、意図せざる核兵器使用の可能性を高めかねないといったリスクをももたらしうる。そうしたことが、ISRおよびNC3への新興技術導入にかかる議論を複雑化させている。

---

[20] Ibid., p. 11; Michael C. Horowitz, “Artificial Intelligence and Nuclear Stability,” Lora Saalman, ed., *The Impact of Artificial Intelligence on Strategic Stability and Nuclear Risk: Volume I—Euro-Atlantic Perspectives*, SIPRI, 2019, p. 79.

[21] Horowitz, Scharre and Velez-Green, “A Stable Nuclear Future?” December 2019, p. 11.

[22] James Johnson, “Delegating Strategic Decision-Making to Machines: Dr. Strangelove Redux?” *Journal of Strategic Studies*, 2020, pp. 7-8, https://doi.org/10.1080/01402390.2020.1759038.

[23] Vincent Boulanin, Lora Saalman, Petr Topychkanov, Fei Su and Moa Peldán Carlsson, *Artificial Intelligence, Strategic Stability and Nuclear Risk*, SIPRI, June 2020, p. x を参照。

2. 抑止関係の安定化

核抑止関係を考える際の重要な視点の１つが、核戦争が勃発する公算の低い状況を意味する「戦略的安定（strategic stability）」への影響である。戦略的安定は主として、危機においても先制核攻撃の誘因が生じにくい状態である「危機における安定（crisis stability）」、および戦略核戦力などの量的・質的増強の誘因が抑えられた状態である「軍拡競争にかかる安定（arms race stability）」で構成される。その戦略的安定の維持には、抑止国が、被抑止国あるいは敵対国の先制核攻撃を受けても第二撃を確実に行う能力を保持することが鍵を握るとされてきた。

(1) 適切な判断

新興技術の核兵器システムへの導入が抑止関係の安定化をもたらすという議論は、それが報復能力に依拠する懲罰的抑止力（deterrence by punishment）の向上により戦略的安定の維持に寄与するとの見方によるものである。たとえば、強化されたISR能力によって、抑止国が持つ報復能力たる核戦力に対する敵対国による攻撃を早期に探知できれば、報復能力が無力化されるといった状況を回避するための措置を講じることができるかもしれない。それは、抑止国が核兵器を破壊される前に使用したいという早期使用や先行使用の誘因を低減するものとなる。

たとえば、迅速かつ正確な状況認識は、効果的・効率的な抑止力の行使を可能にするとともに、敵対国の行動に対する誤解や誤認から不要な核攻撃を敢行するといった可能性を低減することで、抑止の信頼性や安定性を高める効果も指摘できよう。

また、大規模なデータ・情報の集積、迅速な解析・分析、これに基づく適切な意思決定、ならびに核兵器使用命令に至るまでの抗堪性・回復力のある通信（通信ネットワークの迅速な切り替えや、敵対国のサイバー攻撃に対する防御など）が、AIを含む新興的な情報通信技術などを導入されたNC3によって可能になる場合、高い緊張状態や武力衝突の生起、さらには核戦争の勃発といった極めて複雑で、情報の過多・煩雑や時間の切迫に直面する状況のなかでも、適切な情報に基づく意思決定を支援し、抑止国が報復攻撃を敢行する確実性を高め、懲罰的抑止力の信頼性は格段に強化される[24]。核兵器使用にかかる決定までの時間──意思決定者が脅威を評価し、標的を検討し、核兵器使用の可否と方法を決定する──を長くすることができれば、報復攻撃を急ぐべきだとの圧力を低減し、それだけ適切な決定がなされる可能性を高めることもできよう[25]。

(2) エラーの防止

AIや機械学習アルゴリズムとセンサー技術の進歩を融合させることで、ISRやNC3がより自律的かつ正確に動作し、情報の解析・分析や意思決定のためのオプションの検討などへの人間の依存度を

---

[24] James Johnson, "The AI-Cyber Nexus: Implications for Military Escalation, Deterrence and Strategic Stability," *Journal of Cyber Policy*, September 2019, pp. 4-5, https://doi.org/10.1080/23738871.2019.1701693; Horowitz, Scharre and Velez-Green, "A Stable Nuclear Future?," p. 14-15.

[25] David C. Gompert and Martin Libicki, "Cyber War and Nuclear Peace," *Survival*, Vol. 61, No. 4, August-September 2019, pp. 48-49.

低減できれば、「戦場の霧」（作戦・戦闘における不確定要素）のなかで人間が陥りやすい戦略的意思決定に内在する欠点やヒューマンエラー――認知バイアス、反復作業、疲労、サンクコストへの投資、偏ったリスク判断、認知ヒューリスティック（経験や先入観からの直感）、集団思考など[26]――の多くを克服し、核兵器が不正・不要に、あるいは抑止関係の不安定化を強く助長するような態様で使用される可能性も低減しうる[27]。さらに、実環境での実施が困難な核兵器の使用にかかる演習が、AI などの導入によって、高度で精密なシミュレーションという形で実施することができれば、より適切な核態勢を検討し、構築することにつながる。機械学習アルゴリズムと自律システムを使用して、物理的な攻撃や非物理的なサイバー攻撃（攻撃的サイバー攻撃、妨害攻撃、高高度核爆発などによって発生する電磁パルスなど）に対する NC3 の防御を強化することも可能になるとされる[28]。

(3) 自動化

敵対国の核攻撃に対して核報復攻撃を自動的に発動するシステムに対しては、後述のような問題から懸念や反対が根強くあるが、安定化という視点で見た場合には、自動化を導入する国は、自国が定める一定の状況において核兵器を自動的に使用するとして自らの手を縛り、選択肢を極限化することになり、このことが敵対国に対する懲罰的抑止の信憑性を高める効果をもたらすかもしれない。そうしたシステムは、技術的・政治的に適切に作動するとの条件付きながら、核兵器使用の権限を持つ意思決定者が敵対国の攻撃目標になったとしても、報復能力を発動できるという安心感をその意思決定者に与えるものとなる。

新興技術が支える ISR や NC3 によって核報復能力の残存性および信頼性が高まれば、敵対国は抑止国に対する先制核攻撃の遂行を強く抑制されることになろう。同時に、抑止国は、懲罰的抑止力が無効化される懸念が低くなることで、先制核攻撃を遂行したいとの誘因も抑制される。また、懲罰的抑止力の非脆弱性や確実性に対する自信が高まれば、報復攻撃に必要な規模以上の核戦力を保持する必要性が低減される。危機における安定および軍拡競争にかかる安定をともに維持・強化できるとすれば、新興技術の核兵器システムへの導入は、抑止関係の安定化に寄与するということになる[29]。

(4) 批判への反論

後述のように、核兵器システムへの新興技術の導入は、抑止関係に大きな不安定化やリスクをもたらす可能性が高いとも指摘されている。これに対しては、新興技術の発展や核システムへの組み込みは、それ自体が直ちにリスクをもたらすわけではないとの反論もある。たとえば、技術は、リスクを増幅させるイネブラー（enabler）として作用しうるが、技術自体が独立してリスクとなるわけではな

---

[26] Johnson, “Delegating Strategic Decision-Making to Machines,” p. 4.

[27] Johnson, “The AI-Cyber Nexus,” pp. 4-5; Johnson, “Delegating Strategic Decision-Making to Machines,” p. 14.

[28] Johnson, “Delegating Strategic Decision-Making to Machines,” p. 14.

[29] Horowitz, Scharre and Velez-Green, “A Stable Nuclear Future?” pp. 4-5; Boulanin, et.al., *Artificial Intelligence, Strategic Stability and Nuclear Risk*, pp. 102-103.

い[30]。抑止関係の不安定化や意図せざる核兵器使用に至るまでには、その経路上に様々な要因があり、技術はそうした要因との相互作用のなかでリスクを高める可能性があるが、同時にリスク抑制要因として機能することもある。技術がいずれに働くかは、安全保障環境や核の脅威の状況、核戦力の運用方法、指導者の選好などに左右され、たとえば意図せざる核兵器使用をもたらしうる従来型の要因（地政学的緊張、コミュニケーションの欠如、意思表示の不十分さなど）は、技術がもたらすリスクよりも重要で影響力が大きいとも指摘されている[31]。さらに、新興技術の導入による戦略的安定の強化や核リスクの低減といった利点を最大限に活用するために、人間と機械の長所と限界を考慮し、自動化の精度とスピードに人間の判断の柔軟性を加味し、自動化の偏りや人間の判断を機械に委ねることを避ける方法で行うといった施策を講じることを検討すべきだとも論じられている[32]。

3. 抑止関係の不安定化？

核兵器システムへの新興技術の導入が抑止関係を不安定化させるとの見方には、以下のようなものが挙げられる。

(1) 損害限定の強化

第一に、NC3 や ISR の強化が抑止態勢の修正をもたらし、これによって戦略的安定が損なわれる可能性である。抑止国が自国の ISR と NC3 を強化することによって、敵対国の核戦力を高いレベルで弱体化させ、あるいは無効化できるような効果的な対兵力打撃能力に基づく損害限定能力を確立する場合、緊張が激化して武力衝突も視野に入る状況におかれれば、その抑止国は自国が被りうる損害を事前に低減すべく、敵対国が核戦力を使用する前に先制攻撃で破壊したいとの誘因を高めるかもしれない。戦略的安定の重要な鍵は第二撃能力の残存性の維持であるが、リモートセンシングの発展により、核戦力の残存性を支える隠匿性を低下させる可能性も指摘されている[33]。また、そうした ISR の発展による標的選定の高精度化・迅速化は、通常戦力のさらなる精密誘導化とも相まって、従来は核兵器に割り当てられていた目標への攻撃が通常戦力で遂行できる可能性を高めるとも考えられる。さらに、高い損害限定能力を獲得するのが力による一方的な現状変更など攻勢的な目標を有する国である場合、自国が被りうる損害を懸念することなく軍事的な行動に踏み切ることも考えられる[34]。逆に、被抑止国も、核兵器が対兵力打撃によって破壊される前に使用したいとの誘因を高めるであろう。抑止国・被抑止国の双方が、危機における安定を損なう行動をとることになりかねない

---

[30] Johnson, “The AI-Cyber Nexus,” p. 4.

[31] Boulanin, et.al., *Artificial Intelligence, Strategic Stability and Nuclear Risk*, p. x.

[32] Michael Depp and Paul Scharre, “Artificial Intelligence and Nuclear Stability,” *War on the Rock*, January 16, 2024, https://warontherocks.com/2024/01/artificial-intelligence-and-nuclear-stability/.

[33] Daryl K. Press, “NC3 and Crisis Instability–Growing Dangers in the 21st Century,” *NAPSNet Special Reports*, October 17, 2019, https://nautilus.org/napsnet/napsnet-special-reports/nc3-and-crisis-instability-growing-dangers-in-the-21st-century/.

[34] Keir A. Lieber and Daryl G. Press, “The New Era of Counterforce: Technological Change and the Future of Nuclear Deterrence,” *International Security*, Vol. 41, No. 4 (Spring 2017), pp. 9-49; Horowitz, Scharre and Velez-Green, “A Stable Nuclear Future,” pp. 4-5.

もちろん、核兵器システムへの新興技術の導入だけで、強力な損害限定能力の構築が可能になるわけではない。しかしながら、実際には抑止国の核態勢に大きな変容をもたらさない（つまり、損害限定ではなく懲罰的抑止態勢が維持される）場合であったとしも、抑止国による NC3／ISR の強化の可能性を把握した被抑止国は、抑止国の先制攻撃によって弱体化・無力化される前に核兵器を使用できようにすべく、核戦力を高い警戒態勢においたり、核兵器使用の権限を事前に部隊指揮官などに委譲したりするなど、一般に不安定化を招きやすいとされる核態勢の採用に踏み切るかもしれない。しかも、状況認識、通信、意思決定の高速化に伴い、抑止国と被抑止国による損害限定攻撃や報復攻撃の迅速な遂行をめぐる競争を惹起することとも相まって、危機時の、あるいは武力紛争におけるエスカレーションが加速化する可能性もある[35]。

(2) NC3／ISR への攻撃

第二に、核兵器システムにおける ISR や NC3 の価値が従前以上に高まる──特に自国の核抑止力の非脆弱性を高め、あるいは敵対国の核抑止力に対する中核的機能になる──なかで、そうした ISR や NC3 が攻撃対象になることによる不安定化の惹起である。ISR および NC3 に対しては、電子攻撃（ジャミングとスプーフィングの両方を含む）、物理的損壊・破壊（衛星に対する攻撃、電磁パルスを含む核兵器による破壊、海底ケーブル切断、特殊作戦など）、サイバー攻撃をはじめとして多様な攻撃手段が想定される[36]。他方で、敵対的な攻撃に対する防御を強化すると、機械学習システムなどの新たな脅威を検知する能力が損なわれることが多いとされる[37]。

この点で、ISR や NC3 に用いられる重要なアセットが、核戦力専用ではなく、予算の活用や作戦の遂行における効率性といった観点から、核・通常両用のものも一定程度存在すると考えられることは、危機における安定を損ないかねない。通常戦力による攻撃の阻害を目的として核・通常両用の ISR や (N)C3 に、攻撃するとしても、攻撃を受けた国は核関連アセットへの重大な攻撃で（も）あると認識する公算が高い。ISR や NC3 が機能しなくなれば、核戦力が機能不全となりかねないからである。そうした事態は、核兵器使用へのエスカレーションを招きかねない。そして、ISR や NC3 への攻撃能力を紛争当事国双方が保有する場合、紛争の初期段階でこれを実行したいという誘因が高まることもあり、そのエスカレーションの可能性と速度は不可避的に高まることになる[38]。

NC3 や ISR は、情報通信技術に多くを依存するという特質からサイバー攻撃の対象となりやすく、このことが危機における安定に影響を与える可能性も懸念されてきた。ISR や NC3 には、サイバー攻撃に対する多くの脆弱性があり[39]、それらを完全に防御するのは容易ではない。核兵器システムに対

---

[35] Michael C. Horowitz, "Artificial Intelligence and Nuclear Stability," Lora Saalman, ed., *The Impact of Artificial Intelligence on Strategic Stability and Nuclear Risk: Volume I—Euro-Atlantic Perspectives*, SIPRI, 2019, p. 82-83.

[36] Carol Ann Jones, "Counter Nuclear Command, Control, and Communications," *NAPSNet Special Reports*, November 7, 2019, https://nautilus.org/napsnet/napsnet-special-reports/counter-nuclear-command-control-and-communications/.

[37] Vladislav Chernavskikh, "Nuclear Weapons and Artificial Intelligence: Technological Promises and Practical Realities," *SIPRI Background Paper*, September 2024, pp. 8-10.

[38] Acton, "Escalation through Entanglement," pp. 58-59.

[39] 脆弱性のポイントとして、たとえばルイス（Patricia Lewis）とウナル（Beyza Unal）は、指揮所と管制センタ

するサイバー攻撃のシナリオについても、早期警戒システムに対して核攻撃に関する誤情報を提供すること、当局者、オペレーター、核兵器システム、敵対国などとの間の通信を混乱化すること、サプライチェーンなどからの核兵器の有効性を損なう可能性のある方法で核兵器に欠陥や悪意のあるコードを導入すること、あるいはサイバー攻撃による盗用やセキュリティ装置の破壊を通じて核兵器を不正に制御することなど多岐にわたる[40]。

核兵器システムへのサイバー攻撃は、高い費用対効果での奇襲を可能にすること、即時の対応を要するという時間的プレッシャーを増大させること、通信チャネルの混乱や破壊によって指揮統制が困難になること、軍事的対応に代わる実行可能な代替策を見出すのは容易ではないことなど、攻撃側にとって利点が大きい[41]。さらに、軍事攻撃は抑止失敗時に発生するのに対して、「サイバー紛争は武力紛争の閾値以下で起こるため、攻撃的サイバー作戦と NC3 の危険な組み合わせは、実質的に核の閾値を下げることになる」[42]。こうしたことも、敵対国によるサイバー攻撃によって無力化される前に核兵器を使用したいとの誘因を高めかねない。

(3) 不明瞭な能力

第三に、ISR や NC3 は、その重要性に対して、いかなる能力を有しているのかが敵対国からは明確ではないという問題である。新興技術を組み込んだ ISR や NC3 の能力について、その中核が非物理的なプログラムや通信であることから、核弾頭や運搬手段であるミサイルや航空機などとは異なり、実験や演習の形で能力を敵対国に示すこと、あるいは敵対国が抑止国の能力を測ることは容易ではなく、そうした能力の発展によって抑止国の核態勢にいかなる変化が生じるかについても、敵対国は推測によらざるを得ない。

このため、抑止国の ISR／NC3 能力の向上を察知した敵対国は、自国に対する最も厳しい想定に基づいて抑止国の核態勢を推測し、これに対応すべく核態勢や作戦計画の修正を図るかもしれない[43]。このことは、抑止関係の安定化にかかる明示的、暗黙的、あるいは疑似的な合意に関係国が収斂するのを一層難しくする可能性もある。

ー間の通信、司令部からミサイルプラットフォーム（潜水艦など）やミサイルへの通信、ミサイルから地上・宇宙をベースとした指揮統制アセットへの遠隔測定データ、長期・リアルタイムの情報を収集・解釈するための分析センター、輸送におけるサイバー技術、研究所や組立施設におけるサイバー技術、アップロードのための発射前標的情報、全地球航法システムからの位置データ・航法データ・タイミングデータを含む宇宙ベースのシステムからのリアルタイムターゲッティング情報、宇宙・航空・地上ベースのセンサからのリアルタイム気象情報、打ち上げプラットフォーム（潜水艦など）の位置情報、地上局からのリアルタイムのターゲッティング情報、同盟軍司令部間の通信、戦略的インフラストラクチャ内のロボット自律システムを挙げている。Patricia Lewis and Beyza Unal, “Cyber Threats and Nuclear Weapons Systems,” John Borrie, Tim Caughley and Wilfred Wan, eds., *Understanding Nuclear Weapon Risks*, UNIDIR, 2017, p. 62.

[40] Page O. Stoutland and Samantha Pitts-Kiefer, “Nuclear Weapons in the New Cyber Age,” Nuclear Threat Initiative, September 2018, pp. 13-20.

[41] Andrew Futter, “Cyber Threats and Nuclear Weapons: New Questions for Command and Control, Security and Strategy,” *Occasional Paper*, Royal United Services Institute, July 2016, pp. 28-29.

[42] Jon Lindsay, “Digital Strangelove: The Cyber Dangers of Nuclear Weapons,” *Lawfare*, March 12, 2020, https://www.lawfareblog.com/digital-strangelove-cyber-dangers-nuclear-weapons.

[43] Horowitz, Scharre and Velez-Green, “A Stable Nuclear Future,” p. 13.

さらに、ドゥカレク（Jacek Durkalec）らは、以下のようにも指摘している。

> 新興・破壊的技術（EDT）で武装した核保有国は、独自の技術で武装した他の核保有国と対峙することになる。それぞれの側がどのように行動するかは、自国の能力と目的だけでなく、相手の能力にも左右される。この相互作用のプロセスには、意思決定者が自らの技術システムと敵の技術システムとの相互作用と闘いながら戦略的に行動しようとするとき、人間と人間、人間と機械、機械と機械がさまざまな程度で相互作用することになるだろう。これらの相互作用はすべて、「戦争の霧」に覆われた不確実な環境の中で行われる[44]。

核兵器が使用された場合に生じうる事態の不確実性や不可測性は、核兵器の使用に対する慎重さを高め、このことが核抑止を機能たらしめる要因の１つにもなる。ISR や NC3 への新興技術の導入は、その不確実性や不可測性の低減によって（少なくとも自国から見た）核抑止の安定化を企図したものであっても、またそうした役割を担う可能性があるとしても、同時に従来とは異なる不確実性や「戦争の霧」をもたらすことで危機における安定に否定的な影響を及ぼす可能性は排除できない。

(4) 軍拡競争の可能性

第四に、軍拡競争の安定性への影響である。抑止国が効果的な損害限定の遂行を可能にするような ISR や NC3 を構築する場合、これに伴って損害限定を可能にする核・通常戦力を保持したいという誘因が高まりうる。他方で、それは第二撃能力の弱体化・無効化を懸念する被抑止国に、残存性を高めるべく核戦力の量的・質的拡充を迫るものともなりうる。抑止国がこれに対抗して損害限定能力の一層の拡充を図るといったスパイラルに陥れば、軍拡競争に係る安定性は損なわれる。

また、新興技術の導入による ISR や NC3 の強化が核抑止力の強化に不可欠で、核戦争の帰趨を大きく左右しうるとすれば、米国、中国およびロシアだけでなく、他の核保有国もその獲得に強い関心を持つと考えられる。さらに、ISR や(N)C3 は核戦力に特有のものではなく、通常戦力においても重要な構成要素であり、核・非核両用での導入が可能であることも、核保有国による推進を後押しし、競争が加速化する要因となりえよう[45]。こうしたことが抑止関係の複雑性を一段と高め得ることにも留意しなければならない。

4. 意図せざる核兵器使用のリスク

新興技術の NC3 や ISR への導入が抑止関係の安定性・不安定性のいずれに働くかという論点以上に注視されているのが、誤解、誤認、誤算あるいは事故などによって意図せざる核兵器の使用がもたらされるというリスクが高まることへの懸念である。

(1) 核・非核両用

第一に、特に ISR や NC3 が核・非核両用として構築される場合のリスクである。武力紛争では、

---

[44] Jacek Durkalec, Anna Péczeli and Brian Radzinsky, “Nuclear Decision-Making, Complexity and Emerging and Disruptive Technologies: A Comprehensive Assessment,” European Leadership Network, February 2022, p. 14.

[45] Tristan A. Volpe, “Dual-Use Distinguishability: How 3D-Printing Shapes the Security Dilemma for Nuclear Programs,” *Journal of Strategic Studies*, Vol. 42, No. 6, 2019, pp. 816-817, 826-827.

敵対国の軍事行動を阻害すべく、核兵器用、通常戦力用、あるいは核・非核両用のいかんにかかわらず、ISR や(N)C3 を攻撃する強い誘因が生じる。しかしながら、敵対国は、抑止国の ISR や(N)C3 が核兵器専用か核・非核両用かを正確に識別するのは容易ではない。そして、敵対国が通常攻撃に対する損害限定を企図して抑止国の核専用あるいは核･非核両用のシステムに非核攻撃を敢行した場合に、抑止国は核兵器システムへの攻撃だと判断して核兵器を用いた報復を決断するといった、核戦争への意図せざるエスカレーションがもたらされることになりかねない。プレス（Daryl K. Press）は、潜在的・顕在的敵対国に通常戦力で劣勢な核保有国を取り上げつつ、以下のようにも論じている。

> いくつかの核保有国（パキスタン、北朝鮮、ロシアなど）は、敵対国の優勢な通常戦力を抑止あるいは阻止するために核兵器に依存している。問題は、この任務、すなわち制御された強制的な戦時のエスカレーションが NC3 に大きな要求を課していることである。その結果、通常戦争時にこれらの国の NC3 がわずかに劣化しただけでも、その重要な機能を実行する能力を危うくし、危機における不安定という、研究者が恐れる危険な行動（警告、分散、事前移譲、運用など）を誘発する可能性がある[46]。

(2) 意思決定・自動化バイアス

第二に、意思決定の自動化が進む場合、自動化バイアス（automation bias）──AI が下した判断や提案を、仮にそれが不完全、不正確あるいは不確実であったとしても、まさに AI が示したものであるとの理由で「最適解」だと信頼し、選択する──のリスクも指摘されている。武力衝突、さらには核戦争（の可能性）という極めて複雑かつ高ストレスの状況では、人間の注意力を低下させ、矛盾する情報を却下する傾向が強くなり、そうしたなかで意思決定に AI など自動化システムを使用する場合、情報探索のためのヒューリスティックな代替手段（またはショートカット）として自動化システムを使用する傾向が強まる可能性がある[47]。新興技術の導入に伴う意思決定の加速化によって人間の意思決定に許される時間が短縮されれば、意思決定者の柔軟性を低下させたり、さらなる自動化を促すことで人間による制御・判断を排除したりするなど[48]、指導者の判断に与える影響も無視しえない。さらに、2010 年代半ば以降の力の移行（power transition）に伴い戦略的競争が激化し、核抑止関係の多極化も進行する状況は、AI の不安定化効果をさらに悪化させ、将来のエスカレーションリスクを増大させるとも指摘されている[49]。

もとより、AI が常に「正答」を示す保証はない。AI の信頼性はデータの質、トレーニングの有効性、操作のパラメータに大きく依存するが、特に 1945 年 8 月の広島・長崎への原爆投下以来、80 年にわたって実戦で使用されず、危機や使用の実例に関するデータが不足する核兵器について、必要な

---

[46] Daryl K. Press, “NC3 and Crisis Instability.”

[47] Johnson, “Delegating Strategic Decision-Making to Machines,” p. 7. 他方で、逆に AI による出力に過度に懐疑的になる場合、実際には適切な出力が行われているにもかかわらず、それとは異なる判断や決定を下すという、いわゆる信頼バイアス（trust bias）の問題もありうる。

[48] Boulanin, et.al., *Artificial Intelligence, Strategic Stability and Nuclear Risk*, pp. 113-114.

[49] James Johnson, “The Impact of Artificial Intelligence on Strategic Stability, Escalation and Nuclear Security,” Sam Dudin and Chelsey Wiley, eds., *The 2019 UK PONI Papers*, Royal United Services Institute, December 2019, p. 44.

データやトレーニングをシステムに十分に信頼できるレベルで組み込むことができるのか、どの程度のレベルであれば信頼に足るものとなるのか。

> アルゴリズムが訓練データセットに『過剰適合』し、訓練データに特有で、実世界シナリオの一般化には関連しないパターンに基づいて教訓を学習してしまう可能性がある。戦略的早期警戒アルゴリズムは、…全面的な奇襲攻撃に関する何千もの合成シミュレーションを与えられ、実際には取るに足らない警告の兆候を推定するかもしれない。アルゴリズムの特性が不透明なため、通常攻撃やデモンストレーションの核爆発を核攻撃の前兆と誤認し、不当な核エスカレーションのリスクが高まる可能性がある[50]。

敵対国の核兵器使用を含む武力行使の意図、あるいは敵対国の抑止国に対する認識について、抑止国の AI を導入した NC3 がどこまで高い確度で分析できるかという問題もある。「やや直感に反する技術的課題は、AI システムが人間的なバイアスで構築されていることである。『客観的』とされる自律システムの基礎となるアルゴリズムとコードは、人間が提供したデータを使って訓練される。人間のプログラマーがいかに善意であっても、訓練プロセスに人間のバイアスを持ち込むことは避けられない」[51]という指摘も無視し得ない。そうした AI によって提示される分析、判断、選択肢への過度の依存は、意図せざるエスカレーションの重大な原因となりうる。

さらに、AI がいかにして結論を出したのかを人間が完全に理解することは難しいという、いわゆる「ブラックボックス」問題がある。事故、故障あるいは攻撃の可能性を含め、AI による出力の適切性を検証することができなければ、透明性が極めて重要である核の意思決定における AI の有用性は低下する[52]。

(3) 機械の不十分性

第三に、新興技術を導入した核兵器システムの複雑性・不確実性や、アルゴリズムのミス、予測できない誤作動・誤検知、予期せぬ重大な事故やリスクの顕在化、サイバー攻撃などの影響、さらにはシステムのブラックボックス化により人間がその健全性や出力された情報の（不）正確性を適切に認識できない可能性によって、核兵器の使用がもたらされる可能性もある[53]。特に、技術的優位性の維持や獲得のために新興技術導入の競争が生じる場合に、十分なリスク評価、保障措置あるいは冗長性が

---

[50] Michael Klare and Xiaodon Liang, "Beyond a Human 'In the Loop': Strategic Stability and Artificial Intelligence," *Issue Brief*, Arms Control Association, November 12, 2024, https://www.armscontrol.org/issue-briefs/2024-011/beyond-the-loop.

[51] Peter Rautenbach, "Keeping Humans in the Loop Is Not Enough to Make AI Safe for Nuclear Weapons," *Bulletin of the Atomic Scientists*, February 16, 2023, https://thebulletin.org/2023/02/keeping-humans-in-the-loop-is-not-enough-to-make-ai-safe-for-nuclear-weapons/.

[52] Alice Saltini and Yanliang Pan, "Beyond Human-In-The-Loop: Managing AI Risks in Nuclear Command-And-Control," *War on the Rocks*, December 6, 2024, https://warontherocks.com/2024/12/beyond-human-in-the-loop-managing-ai-risks-in-nuclear-command-and-control/.

[53] Elsa B. Kania, "Emerging Technologies, Emerging Challenges—The Potential Employment of New Technologies in Future PLA NC3," *NAPSNet Special Reports*, September 5, 2019, https://nautilus.org/napsnet/napsnet-special-reports/emerging-technologies-in-future-pla-nuclear-command-control-and-communications /.

なされない状態で信頼性の低い技術が早期採用されたり[54]、潜在的・顕在的なリスク要因を軽視・過小評価した導入がなされたりすれば、核リスクを一段と高めることになる[55]。核兵器発射の自動システムを導入する核保有国に関しては、機械的な問題で NC3 への信号の接続ができなくなる場合、自動システムが引き継いで核発射の司令を発報するリスクもある。

また、ジョンソン（James Johnson）は、「人間の共感性（またはマインドに関する他の理論的な感情属性）を持たない機械が引き起こす可能性のある危険性、すなわち事前に設定された目標を容赦なく最適化すること、あるいは自己動機に基づく未来の反復を追求することで、予期せぬ意図しない結果をもたらすこと」[56]といったリスクも指摘する。ジョンソンは別の論考で、人々の判断は、推論、想像力、検討、内省、社会的・歴史的背景、経験、そして危機にとって重要な共感といった能力に依存しているとしたうえで、以下のようにも論じている。

> AI を活用した戦争では、速度、情報過多、複雑で緊密に結合したシステム、多極化が重なり、複雑で動的な状況を発見的に管理し続けるために、危機の際に人々がニュアンスやバランスを避けるという既存の傾向が増幅される可能性が高い。そのため、敵対国に対する誤った信念やイメージ（既存の信念に由来する）は、危機の際に修正されるよりも、むしろ悪化する可能性がある。さらに、不屈の機会速度で行われる危機管理は意思決定の時間枠を圧縮し、非人間的なエージェントが意思決定プロセスに巻き込まれるため、敵の意図について明確な情報が出てきたとしても、時間的な圧力によって外交の微妙なシグナリングや慎重な熟慮がフィルタリングされてしまう（あるいは完全に制限されてしまう）可能性が高い。このように、ある問題に対する決意を示すと同時に、自制の意思を示す、つまり今のところは発砲を控えるという意思を示すことは、行為者が直面する困難であり、基本的に人間の努力に非人間的なエージェントを関与させる（あるいはそれに取って代わる）ことの認知的・技術的な障害によって、飛躍的に複雑化することになる[57]。

秋山信将は、「核危機やエスカレーションゲームにおいて、（狭い）AI の性格や強みと、意思決定に必要なスキルとの間に親和性が欠けているように思われる。戦略目標が常に変化する状況では、狭い AI が『戦争の霧』を晴らす役割を果たすとは限らない。むしろ、…AI における意思決定のブラックボックス化の問題が、AI 特有の『戦争の霧』を生み出す」[58]とも論じている。

(4) 抑止における認識

第四に、抑止をめぐる認識の問題である。抑止は、抑止側が自らにとって望ましくない行動を被抑止側に実行させないために十分な力を持ち、必要な場合にはその力を行使する意思を有し、そのことを被抑止側が適切に認識する場合に機能する。抑止が機能するか、抑止関係がどのように推移するか

---

54 Saltini and Pan, “Beyond Human-In-The-Loop.”

55 Johnson, “The Impact of Artificial Intelligence on Strategic Stability, Escalation and Nuclear Security,” p. 44; Johnson, “Delegating Strategic Decision-Making to Machines,” pp. 13-14.

56 Johnson, “Delegating Strategic Decision-Making to Machines,” p. 5.

57 James Johnson, “Nuclear Brinkmanship in Ai-Enabled Warfare: A Dangerous Algorithmic Game of Chicken,” September 28, 2023, https://warontherocks.com/2023/09/nuclear-brinkmanship-in-ai-enabled-warfare-a-dangerous-algorithmic-game-of-chicken/.

58 Nobumasa Akiyama, “AI Nuclear Winter or AI That Saves Humanity? AI and Nuclear Deterrence,” Joachim von Braun, Margaret S. Archer, Gregory M. Reichberg and Marcelo Sánchez Sorondo, eds., *Robotics, AI, and Humanity: Science, Ethics, and Policy* (Springer, 2021), p. 164.

を予見し難いのは、抑止の成否が相手側の認識に多分に依拠するためであるが、NC3 への AI の導入が進めば、自国・相手国の人と人との間だけでなく、人と AI の間、あるいは AI と AI の間の認識、さらにはそれらの相互作用が抑止（関係）に及ぼす影響も計算に入れなければならない。多数国間の抑止関係と同様に、AI という新たな要素が加わることで、認識の相互作用がより複雑化し、それだけ誤解、誤認あるいは誤算の可能性は高まりうる。AI 化された NC3 の段階的な移行・発展によって新旧システムが自国・敵対国に混在する状況は、人間と機械の認識の相互作用を一層不安定化させかねないとも指摘されている[59]。

(5) サイバー攻撃への脆弱性

第五に、NC3 や ISR のデジタル化が一段と進むことで、サイバー攻撃に対する脆弱性が高まりかねない。そうした攻撃として、たとえば偽情報の提供、通信の混乱化、コミュニケーション・チャネルの妨害・破壊などは、状況認識の正確性を低下させ、意思決定の不確実性を増加させる。また、サプライチェーンを通じて、あるいはその他の方法で、核兵器にその有効性を損なう可能性のある方法で欠陥や悪意のあるコードを導入し、核兵器システムの誤作動や不正な制御をもたらすことも考えられる[60]。しかも、そうしたサイバー攻撃が武力紛争時だけでなく平時からなされ得るとすれば、平時から核のエスカレーションの可能性を考えなければならず、結果として核兵器使用の閾値を低下させかねない。武力紛争に向けて緊張が極度に高まる状況では、核兵器システムへのサイバー攻撃がエスカレートし、その結果として核兵器が意図的あるいは偶発的に使用される可能性が高めうる。

さらに、核兵器システムへのサイバー攻撃が、紛争当事国間だけでなく、第三国、あるいは非国家主体や個人によって引き起こされ、紛争当事国へのなりすまし、あるいは（可能性は低いながらも）システムにハッキングして核兵器使用の不正な命令の送信するなどによって、核戦争が誘発されるといったリスクも指摘されている[61]。意図的なサイバー攻撃だけでなく、マルウェアが非核作戦を支援するシステムから核関連システムに誤って拡散する可能性、あるいは第三者が行った作戦が無関係の二国間対立における一方の当事国のシステムにも作用し、他方からの攻撃だと誤認される可能性も皆無ではない[62]。

5. リスク低減・軍備管理の可能性

(1) 問題の所在

新興技術の核兵器システムの導入が核抑止関係にいかなる含意をもたらすか、少なくとも現状では明確な方向性を示すことは難しい。ただ、抑止関係の不安定化、あるいは意図せざる核兵器使用のリスクといった重大な事態を招く可能性が想定されるとすれば、これをいかにして抑制・防止するかが

---

[59] Yuna Huh Wong, et.al., *Deterrence in the Age of Thinking Machines* (RAND, 2020), p. xiii.

[60] 核兵器システムに対するサイバー攻撃のシナリオについては、Page O. Stoutland and Samantha Pitts-Kiefer, "Nuclear Weapons in the New Cyber Age," Nuclear Threat Initiative, September 2018, pp. 13-20 などを参照。

[61] Franz-Stefan Gady, 'Could Cyber Attacks Lead to Nuclear War?', The Diplomat, 4 May 2015.

[62] James M. Acton, "Cyber Warfare & Inadvertent Escalation," *Dædalus*, Vol. 149, No.2, Spring 2020, pp. 133-151.

核兵器問題における重要な課題の１つであることは明らかである。

他方で、核兵器の意図的・偶発的な使用に至るまでには、新興技術の導入がもたらす問題だけでなく、それ以外に不安定化を高める多様な要因──核兵器の規模と洗練度、技術導入の速度、地理的・地政学的緊張、技術的対称性・非対称性、戦略的関係の状況と成熟度など──も複合的に作用し、むしろ技術よりも重要な要因になる指摘されている[63]。タルマッジ（Caitlin Talmadge）も、「新しい技術が採用される戦略的・政治的文脈が最も重要」であり、そうした文脈、あるいは導入を決定する国の意図によって、同じ技術でももたらしうる含意は異なり、「技術を管理あるいは制限することが、必ずしもエスカレーションを管理あるいは制限するわけではない」[64]と論じている。また、新興技術に対する誇張、過大な期待、あるいは過剰な懸念が、新興技術の適切な導入を阻害する可能性、あるいは新興技術に関する非生産的な議論や対立をもたらす可能性にも留意する必要がある[65]。

（新興）技術は、単独で核抑止関係の動向を決定する独立変数ではなく、他の要因との組み合わせによって抑止関係に作用する媒介変数である。技術は功罪いずれにも働き得るし、これを決定するのは人間である。だからこそ、新興技術の核兵器システムの導入が核態勢や抑止関係に及ぼし得る影響を功罪両面から熟考すること、国・専門家が多角的な見地からこの問題を議論することが、第一に重要である。今後の核戦力近代化計画に新興技術がどのように適用され得るか、どのような脅威やリスクを認識しているか、核兵器使用に至るまでにどこで、またどのようにリスクを低減・遮断するかについての情報の共有も有益である。

(2) 現状

本稿執筆時点で、新興技術の核兵器システムへの導入がもたらしうる核抑止関係の不安定性、あるいは核リスクへの対応を主眼としたリスク低減措置や軍備管理措置は極めて限定的に実施されているに過ぎない。

米国は核兵器使用にかかる意思決定における「ヒューマン・イン・ザ・ループ」を重視し、他の核保有国にも同様の措置を講じるよう求めている。2022 年の NPR で、米国は、「いかなる場合においても、核兵器使用の開始および終了に関する大統領の決定に情報を提供し、それを実行するために重要なすべての行動について、『ヒューマン・イン・ザ・ループ』を維持する」[66]と明記した。英国も、同年の「国防 AI 戦略」で、「戦略システムにおける AI のいかなる使用にかかわらず、核兵器に対する人間の政治的コントロールが常に維持されるようにする。他の核保有国にも同様のコミットメントを行うよう強く奨励する」[67]とした。

---

[63] Boulanin, et.al., *Artificial Intelligence, Strategic Stability and Nuclear Risk* などを参照。また、Johnson, “The AI-Cyber Nexus,” p. 4; Caitlin Talmadge, “Emerging Technology and Intra-War Escalation Risks: Evidence from the Cold War, Implications for Today,” *Journal of Strategic Studies*, Vol. 46, No. 6, 2019, p. 865 なども参照。

[64] Talmadge, “Emerging Technology and Intra-War Escalation Risks,” p. 883.

[65] Lindsay Rand, “The Risk of Bringing AI Discussions into High-Level Nuclear Dialogues,” Carnegie Endowment for International Peace, August 19, 2024, https://carnegieendowment.org/posts/2024/08/ai-nuclear-dialogue-risks-npt?lang=en.

[66] United States, *2022 NPR*, p. 13.

[67] U.K. Ministry of Defense, “Defence Artificial Intelligence Strategy,” June 15, 2022, https://www.gov.uk/

米国は、他の核兵器国に対する議論や措置の実施も積極的に提案してきた。たとえば、サリバン（Jake Sullivan）米国家安全保障担当大統領補佐官は 2023 年 6 月の講演で、戦略的リスク削減に関する多国間フォーラム、特に 5 核兵器国間の対話の重要性を指摘しつつ、その具体的措置の第一に「核兵器の指揮・統制・使用に関する『ヒューマン・イン・ザ・ループ』の維持」を挙げた[68]。また、2024 年 4～5 月の NPT 運用検討会議第 2 回準備委員会で、米国は、5 核兵器国がとるべき核リスク低減措置の 1 つにもそうしたコミットメントを挙げ[69]、この点を含めて中国およびロシアが核リスク低減の取り組みに実質的に関与していないと批判した。これに先立ち、ディーン（Paul Dean）米国務省軍備管理担当官は同年 5 月、米国、フランスおよび英国は核兵器使用に関する決定を AI に決して委ねないとの「明確で強いコミットメント」を行ったとし、中露にも同様のコミットメントをとるよう求めるとともに、「それは責任ある行動規範として非常に重要なものであり、（5 核兵器国の）文脈において非常に歓迎されるものだと考えている」[70]（括弧内引用者）と発言した。

米中間では、中国が米国からの軍備管理対話の提案に消極的で、ほぼ受け入れを拒否してきたが、2024 年 5 月 14 日に、AI に関連する戦略的リスクについて初の二国間対話をジュネーブで開催し、リスク管理や規制の必要性が議論された[71]。また、両国は、2024 年 11 月 16 日の首脳会談で AI の軍事利用についても議論し、米中首脳共同記者会見では、「両首脳は、核兵器使用の決定について、人間による管理を維持する必要性を確認した。両首脳はまた、潜在的なリスクを慎重に検討し、慎重かつ責任ある方法で軍事分野の AI 技術を開発する必要性を強調した」[72]ことが明らかにされた。

他方、サルティニ（Alice Saltini）は、5 核兵器国[73]は核の意思決定を完全に自動化することは容認できないという点では一致しているが、ヒューマン・イン・ザ・ループに対するアプローチや解釈は異なっていると分析している。米国が核の決定において人間の主体性を確保することを明確にしているのに対して、中国は兵器システムに対する人間の管理を維持する重要性を強調するものの、公式には核兵器に関する AI の役割は明記されておらず、ロシアのアプローチは中国よりもさらに透明性が低いとされる[74]。上述のように、ロシアは核兵器発射の自動化システムを導入しているが、これに新興

---

government/publications/defence-artificial-intelligence-strategy/defence-artificial-intelligence-strategy.

[68] "Remarks by National Security Advisor Jake Sullivan for the Arms Control Association (ACA) Annual Forum," June 2, 2023, https://www.armscontrol.org/events/2023-06/remarks-national-security-advisor-jake-sullivan-arms-control-association-aca-annual.

[69] "Statement of the United States," General Debate, Second PrepCom for the 11th NPT RevCon, July 22, 2024.

[70] Greg Torode, "US Official Urges China, Russia to Declare Only Humans, Not AI, Control Nuclear Weapons," *Reuters*, May 2, 2024, https://www.reuters.com/world/us-official-urges-china-russia-declare-only-humans-not-ai-control-nuclear-2024-05-02/.

[71] Xiaodon Liang and Shizuka Kuramitsu, "China Silent on U.S. Risk Reduction Proposals," *Arms Control Today*, June 2024, https://www.armscontrol.org/act/2024-06/news/china-silent-us-risk-reduction-proposals.

[72] "Readout of President Joe Biden's Meeting with President Xi Jinping of the People's Republic of China," The White House, November 16, 2024, https://www.whitehouse.gov/briefing-room/statements-releases/2024/11/16/readout-of-president-joe-bidens-meeting-with-president-xi-jinping-of-the-peoples-republic-of-china-3/

[73] ウィリアムスは、ヒューマン・イン・ザ・ループについて、4 核兵器国がそうしたコミットメントを行っているとしている（Heather Williams, "The Nuclear Order and Emerging Technologies," Centre for Science & Security Studies, King's College London, February 2024, p. 9）。国名は明記されていないが、行っていない国はロシアだと考えられる。

[74] Alice Saltini, "AI and Nuclear Command, Control and Communications: P5 Perspectives," European Leadership

技術が（いかに）使用されるかも明らかではない。

国際社会でも、核兵器システムへの新興技術の導入に関する問題に関心が高まりつつある。2022 年に開催された NPT 運用検討会議の最終文書はロシアの反対により採択できなかったが、その最終ドラフトでは、核リスク低減の文脈で核兵器国に対して、「核兵器国間、および非核兵器国との間で、…新興技術の潜在的影響に関する定期的な対話を強化すること」[75]が盛り込まれていた。また、この会議では、スウェーデンが主導して核リスク低減の推進などを提唱する非核兵器国のグループ「ストックホルム・イニシアティブ」が作業文書を提出し、5 核兵器国が核リスク低減でさらなる作業を行うことが奨励される分野の 1 つに、「特にデジタル分野（サイバー、人工知能、機械学習）や運搬システムの分野における新しい技術が、新たな核リスクにつながり、既存の核リスクを悪化させる可能性を低減するための措置」を含めた。また、核リスク低減に関してさらなる研究および対話が求められる課題の 1 つとして、「サイバー攻撃能力などのデジタル領域や機械学習を含む人工知能など、核リスクに関する新興技術の影響」を挙げた[76]。

(3) 技術的アプローチ

上述のように、核兵器システムへの新興技術の導入にかかるリスク低減および軍備管理は、ようやく議論が始まったばかりだが、専門家などからは様々な観点から多くの提案がなされてきた。このうち、技術的観点からのリスク低減措置としては、（事故や故障による核兵器使用可能性を抑制するためにも）過度に複雑なシステムを構築しないこと、拙速な技術導入を回避すること、技術の限界を認識して導入すること、サイバー攻撃への適切な対応能力（抗堪性、残存性、冗長性の確保など）を組み込むことなどが挙げられる。

冗長性：たとえば、核兵器の使用にかかる判断や決定について、単一の情報源ではなく、複数の種類の ISR システム（異なるタイプのセンサー、異なるデータセットでトレーニングされ、また互いの結果を相互に照合できる異なるパターン認識アルゴリズムなど）に基づくこととすることで、誤った情報による核兵器使用のリスクを低減できる[77]。

拙速な技術導入の回避：機械学習システムの試験・評価の信頼性の高い方法が見つかるまでは、核兵器システムの最も重要な部分、特に NC3 に機械学習システムを拙速に導入するのを避けるべきだと論じられている[78]。また、システムに問題が生じた場合に、安易に（とりわけ十分な信頼を提供することが不可能な AI のような）新たな技術を導入して解決しようとしないことも指摘されている[79]。

Network, November 2023, pp. 20-21.

[75] NPT/CONF.2020/CRP.1/Rev.2, August 25, 2022.

[76] NPT/CONF.2020/WP.9, May 14, 2021. 作業文書の共同提出国は、アルゼンチン、ベルギー、カナダ、デンマーク、エチオピア、フィンランド、ドイツ、アイスランド、インドネシア、日本、ヨルダン、カザフスタン、ルクセンブルク、オランダ、ニュージーランド、ノルウェー、韓国、スペイン、スウェーデン、スイス。

[77] Boulanin, et.al., *Artificial Intelligence, Strategic Stability and Nuclear Risk*, p. 129.

[78] Ibid., pp. 129-130.

[79] Nancy Leveson, "An Engineering Perspective on Avoiding Inadvertent Nuclear War," *NAPSNet Special Reports*, July 25, 2019, https://nautilus.org/napsnet/napsnet-special-reports/an-engineering-perspective-on-avoiding-inadvertent-nuclear-war/.

技術の限界の認識：核兵器システム、特に早期警戒システム、ISR、核指揮統制システムへの機械学習技術の未熟な導入を防ぐために、機械学習システムの試験と評価に関する研究、説明可能な AI に関する研究、AI システムのサイバーセキュリティに関する研究、ディープフェイクや他の AI が生成した誤報を発見できるシステムに関する研究を推進すべきだとの提案もある[80]。また、新興技術が民間における開発によって大きく進展してきたことも踏まえ、システムを設計する技術者と軍の運用者との間の対話を促進し、最も安全な技術的解決策の開発に向けて共同で取り組むことも重要である[81]。

ヒューマン・イン・ザ・ループ：ヒューマン・イン・ザ・ループに関しては、まずはこれを宣言することは重要だが、それだけではリスク低減は保証されない。自動化バイアスや信頼バイアスなどによって適切に機能しない可能性があるだけでなく、逆に「誤った安心感を与え、…リスク低減の義務を果たせるといった幻想を助長する」[82]とも指摘されている。こうした課題に対して、デップ（Michael Depp）らは、「人間と機械のシステムを設計する際には、人間と機械の適切な役割を意識的に決定することが不可欠である。機械は正確さやスピードに優れていることが多いが、人間はより広い文脈を理解し、判断を下すことに長けていることが多い。…軍は、システム設計、オペレーターの訓練、ドクトリン、運用手順に情報を提供するためのガイダンスを確立し、ヒューマン・イン・ザ・ループがたんに機械の中の何も考えない歯車ではなく、実際に人間的判断を行使することを確実にする必要がある」[83]と提案した。また、失敗や誤用のリスクを予測し、またシステムの能力と限界の両方を判断するために、AI ベースのシステムや自律型システムの試験と評価を十分に行うべきである。ヒューマン・マシン・インターフェースは、すべての人間のオペレータに十分な状況認識を与え、自動化のバイアス、信頼不足、ループ外の制御問題のリスクを低減するような方法で設計されることが求められる[84]。

サイバー攻撃への対応能力の導入：最後に、サイバー攻撃に対しては、ISR や NC3 をはじめとする核兵器（関連）システムの強靭性の確保（多様性や冗長性のあるシステム、代替手段の導入など）[85]、信頼性の高い干渉検出機能、厳格なリスク評価（脅威、脆弱性、結果の組み合わせの分析など）に基づく信頼できるシステムの構築[86]などといった対策を、核保有国として果たすべき責任の一つとして重視すべきである[87]。その際に、NC3 に関連するあらゆるセグメントのサイバー・セキュリティを評

---

[80] Boulanin, et.al., *Artificial Intelligence, Strategic Stability and Nuclear Risk*, pp. 129-130.

[81] Ibid., pp. 129-130.

[82] Saltini and Pan, “Beyond Human-In-The-Loop.”

[83] Michael Depp and Paul Scharre, “Artificial Intelligence and Nuclear Stability,” *War on the Rock*, January 16, 2024, https://warontherocks.com/2024/01/artificial-intelligence-and-nuclear-stability/.

[84] Boulanin, et.al., *Artificial Intelligence, Strategic Stability and Nuclear Risk*, pp. 128-129.

[85] Beyza Unal and Yasmin Afina, “How to Deter Cyberattacks on Nuclear Weapons Systems,” Chatham House, December 18, 2020, https://www.chathamhouse.org/2020/12/how-deter-cyberattacks-nuclear-weapons-systems.

[86] Beyza Unal and Patricia Lewis, “Cybersecurity of Nuclear Weapons Systems: Threats, Vulnerabilities and Consequences,” Chathum House, January 2018. サルティに（Alice Saltini）らは、NC3 への AI の活用に伴う重大なリスクを低減するために、たんなる「ヒューマン・イン・ザ・ループ」ではなく、量的に測定可能な最大許容リスクの設定が不可欠とだとし、各国が包括的な安全目標（たとえば偶発的核発射の確率を年間 1,000 万分の 1 以下に抑えるなど）に合意して、確率論的リスク評価を用いて AI・非 AI 双方のシステム構成を定量的に評価する必要があると論じている。Saltini and Pan, “Beyond Human-In-The-Loop.”

[87] Page O. Stoutland and Samantha Pitts-Kiefer, “Nuclear Weapons in the New Cyber Age,” Nuclear Threat

価し、より広範なサイバースペース環境と NC3 の相互作用も含めて評価すること、ならびに分析を技術的な侵入に限定しないことが重要だとされる[88]。

(4) 制度的／政策的アプローチ

技術的アプローチに加えて、リスク低減や軍備管理に対する制度的／政策的アプローチも必要である。もちろん、その推進は容易ではない。第一に、ISR や NC3 への新興技術の導入は緒についたばかりであり、その導入が国家安全保障にいかなる含意をもたらすか、国家安全保障や核態勢に悪影響を及ぼす公算が高いとの認識を共有できるか、リスク低減や軍備管理の実施が相対的な利益をもたらすか（逆に言えば新興技術の導入にかかるリスクが利益を上回るか）といって点について必ずしも（あるいは実際に）明らかにならない限り、予防的な軍備管理（anticipatory arms control）に合意する可能性は高くはない。第二に、新興技術は非核戦力に対する導入が進んでいるが、ISR や(N)C3 について、核・非核を明確に切り離して運用することに抵抗感を持つ核保有国がある場合、核兵器システムに関する措置であったとしても消極的になりうる[89]。第三に、新興技術の導入がもたらしうる核戦力の（特に運用面での）強化の可能性を考えると、少なくともすべての核兵器国が（さらに言えば NPT 外の核保有国も含む形で）参加しなければ、参加による不利益や抑止関係における劣勢の可能性という観点から、いずれの国も参加や実施に消極的になりうる。

核保有国間では、NC3 に直接関係する意思決定は、機械学習アルゴリズムに関連する説明可能性、透明性あるいは予測不可能性といった問題を理由に、機械に委ねるべきではないという点で一般的な合意があるとも指摘されてきた[90]。これが今後も続く保証はないが、そうした自制が働いている間に、あるいは抑止関係の不安定化や核リスクが顕在化する前に、核保有国はそれぞれ新興技術の導入にかかる核態勢や抑止関係への（潜在的な）影響やリスクを評価するとともに、認識の共有を図るべく対話や議論を開始すべきであろう。そうした対話や議論は、戦略的競争の激化に伴い核軍備管理が停滞、さらには逆行するなかで、核兵器問題についても対立する核保有国間の（限定的ながら）一定の協力を促進することに寄与する可能性があり、その意味でも曖昧な問題より、AI の具体的な使用事例に焦点を当てた議論などを取り上げることが有益だとも論じられている[91]。新興技術の発展を主導するのが民間であるとすれば、政府間のトラック 1 だけでなく、民間を含めたトラック 1.5 やトラック 2 といったプロセスでの議論や対話も重要である。

そのリスク低減や軍備管理の態様に関しては、以下のような議論を踏まえれば、能力の制限や条約

Initiative, September 2018.

[88] Jon Lindsay, “Digital Strangelove: The Cyber Dangers of Nuclear Weapons,” *Lawfare*, March 12, 2020, https://www.lawfareblog.com/digital-strangelove-cyber-dangers-nuclear-weapons

[89] Acton, “Escalation through Entanglement,” pp. 92, 98.

[90] James Johnson, “Delegating Strategic Decision-Making to Machines,” pp. 5-6.

[91] Rand, “The Risk of Bringing AI Discussions Into High-Level Nuclear Dialogues.” また、議論や対話の重要性に関しては、Michael Klare and Xiaodon Liang, “Beyond a Human ‘In the Loop’: Strategic Stability and Artificial Intelligence,” *Issue Brief,* Arms Control Association, November 12, 2024, https://www.armscontrol.org/issue-briefs/2024-011/beyond-the-loop も参照。

化といった伝統的な軍備管理よりはむしろ、信頼醸成措置、透明性措置、リスク低減措置、危機管理メカニズム、あるいは規範の構築といった取り組みを推進することが現実的である[92]。

> ミサイルのような有形の核ハードウェアを数えるのと同じように、無形のサイバー活動を定量化あるいは追跡することはできない。こうした理由から、我々は「軍備管理」措置についての考え方を大きく変える必要があるだろう。特に、新興技術を NPT の議論に持ち込もうとするのであれば、それが必要になるだろう。将来の軍備管理措置は、能力を規制するものではなく、ある種の行動を制限するものでなければならない。正式な軍備管理措置は、新しい技術（極超音速滑空飛翔体など）を含むように適応させることができるが、非公式な軍備管理措置は、サイバーや AI といったデュアルユースを可能にする技術…に対処するのに最適だと思われる[93]。

たとえば、行動規範としては、NC3 への AI の導入にあたって、最終的な判断は人間が責任をもって行うこと（そのために、機械が提示する情報、分析、選択肢、判断の誤りを見極めることができる人材を備えること）、意図せざる核エスカレーションのリスクを高めるような行動を抑制すべく NC3 へのサイバー攻撃は行わないこと、危機・緊張状況において敵対国に誤解が生じるような作戦・可能性があることを十分に認識した上で作戦計画を策定すること[94]、AI が支援する警戒データの使用目的を回避行動の開始のみに限定すること（発射命令の開始には使用しない）[95]などが挙げられる。こうしたことをはじめとして、核兵器国（および他の核保有国）が一致して行動規範として合意・実施することが求められる。

また、CBM や透明性措置に関しては、ボウラニン（Vincent Boulanin）らは、国家 AI 戦略の起草や公開、AI の軍事利用に関連した CBM（AI 関連の戦略、政策、軍事的な状況下での AI 関連技術の使用・不使用の意図を概説した軍事的なドクトリンの開示など）、ならびに核兵器に関連した AI の利用に関連した CBM（将来の核近代化計画に AI がどのように適合するかについての情報を共有することや、人的制御手段の形での核戦力関連システムにおける AI の利用に制限を設けることなど）を提案した[96]。

また、上述のように技術だけが問題なのではなく、より幅広い文脈で問題を捉える必要があるとすれば、NC3 や ISR に直接関連しないものの、抑止関係の不安定化や核リスクに関係する課題についての軍備管理を推進する必要もある。新興技術問題は核軍備管理・不拡散に大きな含意を与えつつあり、

---

[92] 核兵器システムへの先端技術の導入がもたらし得る課題への対応策については、Unal and Lewis, "Cybersecurity of Nuclear Weapons Systems"; Acton, "Escalation through Entanglement," pp. 98-99; Edward Geist and Andrew J. John, "How Might Artificial Intelligence Affect the Risk of Nuclear War?" RAND Corporation, 2018; James Johnson and Eleanor Krabill, "AI, Cyberspace, and Nuclear Weapons," *War on the Rocks*, January 31, 2020, https://warontherocks.com/2020/01/ai-cyberspace-and-nuclear-weapons/; Johnson, "Delegating Strategic Decision-Making to Machines," pp. 22-23 などを参照。

[93] Madeline Zutt and Michael Onderco, "How Emerging Technologies Impact the Future of Nuclear Risk and Arms Control," *European Leadership Network*, September 1, 2020, https://www.europeanleadershipnetwork.org/commentary/how-emerging-technologies-impact-the-future-of-nuclear-risk-and-arms-control/.

[94] James A. Acton, *Is It a Nuke? Pre-Launch Ambiguity and Inadvertent Escalation*, Carnegie Endowment for International Peace, 2020.

[95] Zala, "Should AI Stay or Should AI Go," p. 159.

[96] Boulanin, et.al., *Artificial Intelligence, Strategic Stability and Nuclear Risk*, p. 141.

その観点では、たとえばNPTの文脈でも議論されるべきである。ISRやNC3への攻撃（特にサイバー攻撃）は、核保有国のみならず非核兵器国でも実施可能である。ウィリアムス（Heather Williams）は、「核兵器国は新興技術が核態勢に与える影響について透明性をより高め、核兵器国と非核兵器国は協力して新興技術技術に伴うリスクを低減し、NPTはより機動的に運用検討プロセスを更新しなければならない」[97]とも論じている。こうした取り組みは、戦略的競争が激化し、また核軍備管理の停滞・逆行が深刻化するなかで、核兵器の使用可能性を低減するとともに、核軍備管理の再活性化に向けた小さいながらも重要なステップになると考えられる。

おわりに

ISRおよびNC3への新興技術の導入は、核兵器システムの運用速度や情報処理能力の向上を通じて抑止関係を安定化させうる一方で、損害限定能力の強化や意思決定プロセスの複雑化などを通じて不安定化をもたらす可能性もある。さらに、誤解・誤算や誤作動などによる意図せざる核兵器使用のリスクも増大させかねない。

抑止関係や核リスクへの好ましくない影響が現実化するのを防止するために、ISRやNC3への新興技術の導入が進む前に、まずは技術を「功罪どちらにも作用し得る媒介変数」と捉え、核保有国がどのような目的や安全保障戦略のもとで利用するのか、平時・危機時の運用ルールをどう定めるのかなどを多角的に議論することが必要である。そうした議論を踏まえつつ、たとえばヒューマン・イン・ザ・ループの具体的内容を明確化したり、冗長性やサイバー防御力などを備えたシステム設計や信頼醸成措置を導入したりすることが求められる。また、核軍備管理の停滞・逆行や戦略的競争の激化を踏まえても、新興技術の核兵器システムへの利用にかかる対話や議論、行動規範や信頼醸成・透明性措置の策定といった取り組みは、意図せざるエスカレーションを防ぐだけでなく、より広く核軍備管理の再活性化に向けた協力の手がかりとなりうる。新興技術に関連する核軍備管理やリスク低減に向けた対話と協調の可能性を探る作業は、喫緊の課題である。

---

[97] Heather Williams, “Remaining Relevant: Why the NPT Must Address Emerging Technologies,” Center for Science and Security Studies, King’s College, August 2020, p. 6. 他方、NPT運用検討プロセスではすでに多くの課題が議論されており、AIの複雑性を持ち込むことで、他の重要な核問題への関心がそがれる可能性、あるいは外交上の障害をさらに固定化させる形で未解決の対立分野と結びつく可能性も指摘されている（Rand, “The Risk of Bringing AI Discussions Into High-Level Nuclear Dialogues”）。