
【研究ノート】

人工知能技術が核抑止に及ぼす影響

有江 浩一

<要旨>

近年、人工知能（AI）技術が核抑止に及ぼす影響をめぐって、AI 技術には核抑止の安定化に寄与する面があるとともに、その不安定化をもたらす面もあることが議論されている。この問題の起源は、米ソがコンピュータによる核兵器システムの自動化を進めた冷戦期にさかのぼる。AI 技術は、核指揮統制通信、核攻撃早期警戒システム、核兵器運搬手段に逐次導入されていくと考えられ、すでにロシアは核兵器運搬手段への導入に着手しているとみられる。これに伴い、核兵器システムの自律化による核抑止への影響が懸念されている。今後は、核抑止における AI と人間との関係をいかに律すべきかについて、より丁寧な議論を行っていく必要がある。

はじめに

近年における人工知能（Artificial Intelligence: AI）技術の急速な進歩に伴い、核兵器システムに AI 技術を導入する試みがすでに始まっている。ロシアは、大型の核弾頭を搭載可能な自律型無人水中航走体（autonomous underwater vehicle : AUV）を開発中であり、これに AI 技術を導入しているという¹。中国は、AI を活用した意思決定支援システムを人民解放軍海軍の原子力潜水艦に導入し、艦長の業務負担を軽減させる計画を進めているが、同様の AI 支援システムを弾道ミサイル搭載原子力潜水艦（SSBN）に導入する可能性も否定できない²。また、中露による核兵器システムの近代化に対応するため、アメリカの核指揮統制通信システムに AI を導入すべきだと

1 Matt Field, “As the US, China, and Russia Build New Nuclear Weapons Systems, How Will AI Be Built in?” *Bulletin of the Atomic Scientists*, December 20, 2019, <https://thebulletin.org/2019/12/as-the-us-china-and-russia-build-new-nuclear-weapons-systems-how-will-ai-be-built-in/>.

2 Elsa B. Kania, “Chinese Sub Commanders May Get AI Help for Decision-Making,” *Defense One*, February 12, 2018, <https://www.defenseone.com/ideas/2018/02/chinese-sub-commanders-may-get-ai-help-decision-making/145906/>.

3 Adam Lowther, “The Big and Urgent Task of Revitalizing Nuclear Command, Control, and Communications,” *War on the Rocks*, October 4, 2019, <https://warontherocks.com/2019/10/the-big-and-urgent-task-of-revitalizing-nuclear-command-control-and-communications/>.

の意見もある³。

こうした状況に鑑みて、AI技術は核抑止に及ぼすのかについての議論が安全保障やAI技術の専門家の間で交わされつつある。一方では、AI技術によって相互の核戦力に対する警戒監視能力が向上すれば、相手国に対して早期に核兵器を使用しなければならないという必要性が減少し、あるいは核軍備管理・不拡散のための検証が容易になるといった肯定的な意見がある。また他方では、相手国がAI技術を導入して対兵力攻撃能力を向上させれば、自国の核戦力をより高度の警戒態勢に置かざるを得なくなり、核使用のリスクが高まるといった否定的な見解もある⁴。今後、AI技術の軍事利用が進むにつれて、AIと核抑止との関係をめぐる議論も一段と高まっていくものと思われる。

したがって、AI技術の核兵器システムへの導入が進んだ場合、いかなる問題や影響が生じるのかを考察しておくことは、わが国の安全保障にとっても重要であろう。本稿では、まず、AI技術の核兵器システムへの導入について何が問題になっているのかを過去の経緯も含めて概観する。次に、AI技術の核兵器システムへの導入形態を検討し、最後にAI技術が核抑止に及ぼす影響を考察する。

1. 人工知能 (AI) 技術と核兵器

(1) 問題の所在

AI技術が核抑止に及ぼす影響についての研究は、未だ十分とはいえないながらも、少しずつ行われるようになってきている⁵。例えば、ランド研究所は2017年に、AIの発達が将来の核抑止に及ぼす影響を議論するワークショップを開催し、翌18年にその成果を公表した⁶。また、ストックホルム国際平和研究所 (SIPRI) も2018年から翌19年にかけて一連のワークショップを開催し、AIが戦略的安定性および核リスクに及ぼす影響について米欧とアジアの視点から議論を行った。米欧の視点から行われたワークショップの成果は2019年5月に⁷、東アジアの視点からのものは同年10月

4 Rafael Loss and Joseph Johnson, "Will Artificial Intelligence Imperil Nuclear Deterrence?" *War on the Rocks*, September 19, 2019, <https://warontherocks.com/2019/09/will-artificial-intelligence-imperil-nuclear-deterrence/>.

5 Michael C. Horowitz, Paul Scharre, and Alexander Velez-Green, "A Stable Nuclear Future? The Impact of Autonomous Systems and Artificial Intelligence," *ArXiv.org*, December 2019, p. 2, <https://arxiv.org/ftp/arxiv/papers/1912/1912.05291.pdf>.

6 Edward Geist and Andrew J. Lohn, "How Might Artificial Intelligence Affect the Risk of Nuclear War?" RAND Corporation, 2018, https://www.rand.org/content/dam/rand/pubs/perspectives/PE200/PE296/RAND_PE296.pdf.

7 Stockholm International Peace Research Institute (SIPRI), "The Impact of Artificial Intelligence on Strategic Stability and Nuclear Risk, Volume I: Euro-Atlantic Perspectives," ed. Vincent Boulanin, May 2019, <https://www.sipri.org/sites/default/files/2019-05/sipri1905-ai-strategic-stability-nuclear-risk.pdf>.

に公表されている⁸。

これらの研究によれば、AI 技術が核兵器システムに導入されるに伴い、従来の技術では限界のあったものが飛躍的に改善される可能性があるという。この場合、AI 技術が核抑止の安定化に寄与する一方で、その不安定化をもたらす側面もあることが指摘されている。例えば、AI 技術の導入によって、相手国の核戦力に関する情報・監視・偵察（ISR）データをより高速で処理することが可能になるであろう。これにより、危機の際に意思決定者が核使用の要否を判断するための時間の余裕が得られ、核抑止の安定化につながると考えられる⁹。その反面、AI による ISR 能力の向上が進むと、核保有国の間で相互に自国の第二撃能力の信頼性に対する疑念が生じるために、危機の際に偶発的なエスカレーションが起こりやすくなるとの見方もある¹⁰。

核保有国間の抑止を安定化させるには、両国ともに非脆弱な第二撃能力を保有することが必要とされている。非脆弱な第二撃能力とは、相手国の第一撃（先制核攻撃）を受けた後でも生存し、相手国に壊滅的な損害を与え得る核報復能力のことである。これにより、どちらも第一撃の誘因を持たなくなり、相互抑止の関係が成立し、核抑止が安定化すると想定されている¹¹。しかし、軍事技術の進展によって第一撃までの時間が短縮され、第二撃の判断や指令が困難になると、第二撃能力が脆弱化する可能性が出てくる。このため、少しでも時間の余裕を得るために ISR に AI 技術を導入する必要性、あるいは第二撃の判断や指令のプロセスに AI 技術を導入する必要性が生まれると考えられる。

実際、近年における極超音速ミサイルや核弾頭搭載ステルス巡航ミサイルなどの兵器技術の進歩により、第二撃の判断のための時間はますます短くなると予想されている。核兵器システムに AI 技術を導入した場合、敵の核攻撃をより正確に探知できるとともに、意思決定者が核報復すべきか否かを判断するための時間の余裕をより多く確保できるようになることが期待されるのである¹²。

同時に、核兵器システムへの AI 技術の導入は核抑止の安定化に寄与するのか、それとも不安定化をもたらすのかが問題とされ、様々に議論されている現状にある。

8 Stockholm International Peace Research Institute (SIPRI), “The Impact of Artificial Intelligence on Strategic Stability and Nuclear Risk, Volume II: East Asian Perspectives,” ed. Lora Saalman, October 2019, https://www.sipri.org/sites/default/files/2019-10/the_impact_of_artificial_intelligence_on_strategic_stability_and_nuclear_risk_volume_ii.pdf.

9 SIPRI, “The Impact of Artificial Intelligence, Volume I,” p. 80.

10 Geist and Lohn, “How Might Artificial Intelligence Affect the Risk of Nuclear War?” p. 8.

11 Robert Legvold, “Contemplating Strategic Stability in a New Multipolar Nuclear World,” Workshop Report, American Academy of Arts & Sciences, August 2019, pp. 4-5, <https://www.amacad.org/sites/default/files/publication/downloads/Contemplating-Strategic-Stability.pdf>.

12 Lowther, “The Big and Urgent Task of Revitalizing Nuclear Command, Control, and Communications”.

この問題の起源は、次項に述べるように、米ソがコンピュータによる核兵器システムの自動化を進めた冷戦期にさかのぼることができる。

(2) 過去の事例

冷戦期における核兵器システム自動化の最初の実践として、アメリカの半自動防空管制システム (Semi-Automatic Ground Environment: SAGE) が挙げられる。SAGEは、北極圏から北米大陸に向かって侵入してくるソ連の核搭載爆撃機をレーダー網で探知・追跡・要撃するための大型のコンピュータシステムで、1950年代に米空軍が開発・配備した。これにより、遠距離早期警戒線 (Distant Early Warning Line) と呼ばれる、北米大陸北岸の北極圏に設置されたレーダー網からの情報をもとにソ連爆撃機の進路を予測し、要撃指示を出すまでの膨大な計算を短時間で行えるようになった。しかし、SAGEの初号機が運用開始された1958年には、米本土に対するソ連の脅威は爆撃機から大陸間弾道ミサイル (ICBM) に移りつつあり、SAGEではICBMに対応できないことから、米空軍は別のシステムを新たに開発・配備する必要に迫られた¹³。

このため、SAGEと並行して、ソ連のICBMを探知し得る弾道ミサイル早期警戒システム (Ballistic Missile Early Warning System: BMEWS) の開発が進められることになった。BMEWSは、グリーンランド・アラスカ州・英本土の3カ所に設置されたICBM探知用のレーダー網からの情報をコロラド州の北米防空司令部 (NORAD、現在の北米航空宇宙防衛司令部) に自動転送して警報を発するシステムで、情報の検証はNORADが人手で行うことになっていた¹⁴。

1960年9月に、グリーンランドにあるBMEWSのレーダーサイトから誤警報が発出される事案が生起している。これは、BMEWSレーダーの出力や開口部が大きかったために、384,400km離れた月からの反射波を感知してしまい、それがミサイル攻撃の警報メッセージとしてNORADに自動転送されたために起こった事案であった。NORADの担当者がレーダーサイト勤務の空軍士官に電話で確認したところ、誤警報であることが判明した¹⁵。

13 David F. Winkler, "Searching the Skies: The Legacy of the United States Cold War Defense Radar Program," Headquarters, Air Combat Command, United States Air Force, June 1997, https://fas.org/nuke/guide/usa/airdef/searching_the_skies.htm; Lincoln Laboratory, Massachusetts Institute of Technology, "SAGE: Semi-Automatic Ground Environment Air Defense," <https://www.ll.mit.edu/about/history/sage-semiautomatic-ground-environment-air-defense-system>; SIPRI, "The Impact of Artificial Intelligence, Volume I," p. 44.

14 Melvin L. Stone and Gerald P. Banner, "Radars for the Detection and Tracking Ballistic Missiles, Satellites and Planets," *Lincoln Laboratory Journal*, vol. 12, no. 2 (2000), pp. 217-244, <https://pdfs.semanticscholar.org/477d/7311053bd57e12d74355548c83e770420ad6.pdf>.

15 Ibid., p. 221.

1967年5月には、異常な太陽活動に起因した磁気嵐が発生し、BMEWSの全レーダーに大規模な電波障害を引き起こした。これをソ連軍による電波妨害と捉えた米戦略空軍は警戒態勢を引き上げ、核搭載爆撃機を緊急発進態勢に置いた。この時、NORADの太陽予報官が、BMEWSレーダーの障害は巨大な太陽フレアの発生に伴う磁気嵐が原因であると指摘したため、軍は警戒態勢を解除している¹⁶。

NORADのコンピュータシステムに関わる事案も起こっている。1979年11月、NORADのシステムがソ連による大規模なミサイル攻撃の警報を発した。しかし、早期警戒衛星やレーダーなどからはミサイル攻撃の兆候を示す何らの情報も入ってこなかったため、誤警報と判断された。その後の調査で、NORADの担当者が訓練用のテープを誤ってコンピュータの入力装置にかけてしまい、誤警報が発せられたことが判明した¹⁷。

類似の事案はソ連でも生じたが、最も知られているものは1983年9月の事案であろう。同月26日の夜、ソ連の早期警戒衛星システム「オコ (Okko)」がアメリカから弾道ミサイル5発が続けざまに発射され、ソ連に向かってしていると警報を発した。当直将校として勤務していたソ連防空軍のスタニスラフ・ペトロフ (Stanislav Petrov) 中佐は、飛来してくるはずのミサイルを地上レーダーの操作員が探知していないことを電話で確認した上で、その警報はシステムの誤作動により発せられたものと独自に判断した。ペトロフが上官にその旨を報告したために、報復攻撃などの対応行動は取られなかった。その後の調査で、「オコ」システムの衛星が雲からの太陽光の反射によって誤作動を起こしたことが判明した。ペトロフの回想によれば、この時の判断について、アメリカがたった5発の核ミサイルで先制攻撃を行うのは合理的でないと考えたという¹⁸。

冷戦後の1995年1月には、ロシアの早期警戒レーダーがノルウェー沖から打ち上げられた米航空宇宙局 (NASA) のオーロラ観測用ロケットをアメリカの潜水艦発射弾道ミサイル (SLBM) と誤認し、ロシア戦略ロケット軍が警戒態勢に置かれる事案が起こっている。この際、核報復攻撃の大統領命令を伝達するための「核のブリーフケース」がボリス・エリツィン (Boris Yeltsin) 大統領に渡され、史上初めてケースが起

16 D. J. Knipp, et al., "The May 1967 Great Storm and Radio Disruption Event: Extreme Space Weather and Extraordinary Responses," *Space Weather*, vol. 14, issue 9 (September 2016), pp. 614-633, <https://agupubs.onlinelibrary.wiley.com/doi/full/10.1002/2016SW001423>.

17 Daniel Uria, "False Alarm: 1979 NORAD Scare Was One of Several Nuclear Close Calls," UPI, November 8, 2019, https://www.upi.com/Top_News/US/2019/11/08/False-alarm-1979-NORAD-scare-was-one-of-several-nuclearclose-calls/7491573181627/.

18 Megan Garber, "The Man Who Saved the World by Doing Absolutely Nothing," *The Atlantic*, September 26, 2013, <https://www.theatlantic.com/technology/archive/2013/09/the-man-who-saved-the-world-by-doingabsolutely-nothing/280050/>.

動され使用可能な状態となった。幸い、ロケットがモスクワでなく海に向かって落下していることが伝えられたため、ケースは使用されなかった¹⁹。

これらの事例は、核兵器システムの自動化の陥穽を示すものであるが、将来のAI技術による核兵器システムの自律化の課題に通じるものと思われる。特に、いずれの事例でも、人間の介在によって危機を回避できたことが示されている点を考慮すべきであろう。

2. 核指揮統制通信へのAI技術の導入

（1）導入の形態

ここでは、核指揮統制通信（nuclear command, control and communications: NC3）にAI技術を導入する形態を考えてみたい。NC3をコンピュータにより自動化した例としては、ソ連が1985年に運用を開始したとされる自動核報復システム「死者の手（Dead Hand）」が挙げられる。

「死者の手」は、アメリカの先制核攻撃によってソ連指導部が壊滅し、核報復の命令を出せなくなった事態に備えて開発・配備されたもので、核攻撃の早期警戒情報を受けてシステムが起動されるとほぼ自動的に報復核戦力を発動する仕組みになっていた。具体的には、「死者の手」が起動されると、まず地震波や放射能などのセンサーネットワークを監視して核爆発の兆候を感知し、ソ連がアメリカの核攻撃を受けたことを確認する。次に、ソ連指導部との通信回線をチェックし、回線がつながっていれば指導部がまだ生存していると判断して所定の時間後にシステムは自動停止する。回線が途絶して応答が無かった場合は、「死者の手」は核報復の発動に係る通常の指揮命令系統を迂回して、直ちにシステムの当直将校に核報復を発動する権限を委譲する。最後に、当直将校が発射ボタンを押し、複数のサイロから指令用のミサイル（command missile）が発射されると、その後に人間が介在する余地は無い。飛び立った指令用ミサイルはソ連の各地に残存している核戦力に対して空中から無線で暗号化された発射

19 Robert Johnson, "The World Was Never Closer to Nuclear War than on Jan. 25, 1995," *Business Insider*, August 7, 2012, <https://www.businessinsider.com/the-world-has-never-been-closer-to-nuclear-war-than-it-was-duringthis-1995-event-2012-8>.

20 「死者の手」は別名「ペリメーター（Perimeter）」とも呼称され、その存在が知られるようになったのは冷戦後になってからであった。Nicholas Thompson, "Inside the Apocalyptic Soviet Doomsday Machine," *Wired*, September 21, 2009, <https://www.wired.com/2009/09/mf-deadhand/>; Bruce G. Blair, "Opinion: Russia's Doomsday Machine," *New York Times*, October 8, 1993, <https://www.nytimes.com/1993/10/08/opinion/russias-doomsday-machine.html>. なお、「死者の手」は「ペリメーター」とは別のシステムだとする説もある。"Dead Hand / Myortvaya Ruka," *GlobalSecurity.org*, June 2, 2018, <https://www.globalsecurity.org/wmd/world/russia/dead-hand.htm>.

命令を送信し、アメリカに対する報復攻撃が自動的に発動されるのである²⁰。「死者の手」の運用は1995年にいったん中止されたが、ウラジーミル・プーチン（Vladimir Putin）政権下で2011年に運用が再開され、現在も稼働中とみられている²¹。

「死者の手」では、人間（当直将校）が最後に指令用ミサイルの発射ボタンを押すことになっているが、このプロセスまでもAIによって自律化した場合は、核使用の判断や指令はすべて機械に委ねられることになる。このプロセスの発動は、最高意思決定者の不在を前提条件にしているとはいえ、前述のペトロフ中佐のような人間の常識的な判断や、結果に対する慎重さを完全に排除する形で行われるであろう。このため、後述のような否定的意見が多く出されている。

NC3へのAI技術の導入形態として、「死者の手」のような形態以外に、核使用の決定に関してAIに助言を行わせる「AI アドバイザー」の導入も考えられる。すでに、車の運転やスケジュール管理といった日常生活のレベルではAIによる助言機能が広く普及しており、今後のAI技術の進展に伴って軍のウォーゲームや演習などにもAIが重要な役割を果たすようになって見込まれることから、将来的には核抑止の分野においても人間と同等あるいはそれ以上の助言をAIが行えるようになるかもしれない。しかし、核抑止の分野に「AI アドバイザー」を導入するのは信頼性の上で問題があるとともに、核抑止において新たなリスクを生みかねない²²。

他方、中国が原子力潜水艦にAI支援システムを導入し、艦長の業務を補佐させる計画を進めているのは先述の通りであるが、この計画がSSBNにも拡張されていけばNC3への「AI アドバイザー」導入の先例となるかもしれない。テキサス大学のジャガナート・サンカラ（Jaganath Sankaran）は、NC3へのAI技術の導入により、核兵器システムに係る機械的なエラーや人為的な過誤をある程度軽減し得ると期待されることから、人間の意思決定者が核抑止における慎重な判断を行えるようにAIを活用する方向を模索すべきだと指摘している²³。

（2）導入をめぐる議論

米陸軍高等軍事研究院のアダム・ローザー（Adam Lowther）は、アメリカの核兵器システム、特にNC3へのAI技術の導入を進めて、最高意思決定者が十分な時間の

21 “Video: Russia’s ‘Dead Hand’ Command Missiles, Deterrent in the Event of a Crippling Nuclear Attack,” *Global Research*, January 23, 2020, <https://www.globalresearch.ca/video-russian-real-life-doomsday-machine/5701454>.

22 Geist and Lohn, “How Might Artificial Intelligence Affect the Risk of Nuclear War?” p. 18.

23 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/>.

余裕を得て核報復を確実に発動できるようにすべきだと提言している。さもなければ、中露の核攻撃能力が向上していくにつれて、アメリカが攻撃される前に最高意思決定者がNC3を通じて核報復を判断するための時間がさらに短縮されてしまう。これは、戦略的安定性およびアメリカの核戦力の生存性に影響を及ぼすものだという²⁴。

これに対して、戦略国際問題研究所のブライス・ファラボー（Bryce Farabaugh）は、NC3へのAI技術の導入は核使用の決定プロセスから人間を排除することにつながりかねず、特にAI技術が未成熟な段階にある現状を踏まえれば、核戦争を回避し得る機会をさらに減じることになると警鐘を鳴らす²⁵。

また、米経済開発委員会のロリ・エスポージト・マレー（Lori Esposito Murray）は、NC3へのAI技術の導入にあたっては、核使用に係る人間の意思決定をAIに代替させるようなことがあってはならないと指摘する。マレーは、NC3へのAI技術の導入によって抑止の安定性が損なわれるようなことがないように、アメリカが率先して中露両国と協力し、戦略的安定性に係る交渉を始めるべきだと提言している²⁶。

ローザーの提言は、NC3をAIによって完全に自律化すべきだというものではなく、核使用を判断するのはあくまでも最高意思決定者であるとの前提を覆すものではない。にもかかわらず、その前提を覆しているかのような反論が出されるのは、核抑止におけるAIと人間との関係をいかに律するべきかについて、より丁寧な議論が必要であることを示すものであろう。

NC3への「AIアドバイザー」導入をめぐるっては、米科学雑誌「ブレティン・オブ・ジ・アトミック・サイエンティスト」編集委員のマット・フィールド（Matt Field）が、AIの導入に係るオートメーション・バイアスのリスクを指摘している²⁷。これは、人間が自動化システムを過度に信用してしまう傾向があることを示すもので²⁸、AIによって自律化されたNC3システムの発する警報や「AIアドバイザー」の助言を人間が盲目的に信用し、核使用の決定を行ってしまうリスクである。このようなリスクは、核

24 Lowther, “The Big and Urgent Task of Revitalizing Nuclear Command, Control, and Communications”.

25 Bryce Farabaugh, “Bad Idea: Integrating Artificial Intelligence with Nuclear Command, Control and Communications,” *Defense 360*, December 3, 2019, <https://defense360.csis.org/bad-idea-integrating-artificial-intelligence-with-nuclear-command-control-and-communications/>.

26 Lori Esposito Murray, “Commentary: Pairing AI and Nukes Will Lead to Our Autonomous Doomsday,” *Defense News*, November 13, 2018, <https://www.defensenews.com/opinion/commentary/2018/11/13/pairing-ai-and-nukes-will-lead-to-our-autonomous-doomsday/>.

27 Matt Field, “Strangelove Redux: US Experts Propose Having AI Control Nuclear Weapons,” *Bulletin of the Atomic Scientists*, August 30, 2019, <https://thebulletin.org/2019/08/strangelove-redux-us-experts-propose-having-ai-control-nuclear-weapons/>.

28 他方、場合によっては人間が自動化システムを信用しない傾向があることも指摘されている。前東晃礼他「自動化システム利用における Misuse/Disuse の実験的検討」2011年度日本認知科学会第28回大会プロシーディングス、103-111頁、https://www.jcss.gr.jp/meetings/JCSS2011/proceedings/pdf/JCSS2011_O5-2.pdf.

抑止における AI と人間との関係を考える上で議論を深めていくべき論点であろう。

3. 核攻撃早期警戒システムへの AI 技術の導入

(1) 導入の形態

核攻撃早期警戒システムにコンピュータを導入して自動化する試みは、先述のように冷戦期から続けられてきた。しかし、従来のシステムは、飛来してくる核ミサイルや戦略爆撃機をレーダーで探知・追跡し、あるいはミサイルの発射を早期警戒衛星の赤外線センサーで感知して警報を自動的に伝えるものであり、それらを発射前に搜索・発見・識別することはできなかった。特に、移動式のミサイルを発射前に見つけ出すことは、1991 年の湾岸戦争で米軍を中心とする多国籍軍が相当規模の航空戦力と特殊部隊を投入したにもかかわらず、イラク軍の移動式スカッドミサイルを効果的に発見・破壊することができなかった例にみられるように²⁹、長年の課題とされていた。

核攻撃早期警戒システムに最新の AI 技術を導入すれば、ISR データの分析能力が大きく向上することが見込まれる³⁰。これにより、敵の核兵器の位置を発射前に探知し、追跡することができるようになると考えられる³¹。ミズーリ大学の研究者が 2017 年に発表した研究論文によれば、深層畳み込みニューラルネットワーク (deep convolutional neural networks) 技術を用いて中国南東部の膨大な広域衛星画像データを分析させた結果、データに含まれた地対空ミサイル陣地を約 98% の精度で探知できたという³²。なお、ニューラルネットワークとは、人間の脳内にある神経細胞 (ニューロン) とそのつながりである神経回路網を人工ニューロンという数式モデルで表現し

29 Mark E. Kipphut, "Crossbow and Gulf War Counter-Scud Efforts: Lessons from History," Air University, February 2003, pp. 17-18, <https://media.defense.gov/2019/Apr/11/2002115481/-1/-1/0/15CROSSBOW.PDF>.

30 Vincent Boulanin, "AI & Global Governance: AI and Nuclear Weapons – Promise and Perils of AI for Nuclear Stability," Center for Policy Research, United Nations University, December 7, 2018, <https://cpr.unu.edu/ai-global-governance-ai-and-nuclear-weapons-promise-and-perils-of-ai-for-nuclear-stability.html>.

31 Zachary S. Davis, "Artificial Intelligence on the Battlefield: An Initial Survey of Potential Implications for Deterrence, Stability, and Strategic Surprise," Center for Global Security Research, Lawrence Livermore National Laboratory, March 2019, p. 7, https://cgsr.llnl.gov/content/assets/docs/CGSR-AI_BattlefieldWEB.pdf.

32 Richard A. Marcum, et al., "Rapid Broad Area Search and Detection of Chinese Surface-to-Air Missile Sites Using Deep Convolutional Neural Networks," *Journal of Applied Remote Sensing*, vol. 11, no. 4 (October-December 2017), pp. 1-31, <https://www.spiedigitallibrary.org/journalArticle/Download?fullDOI=10.1117%2F1.JRS.11.042614>. なお、SSBN などの潜水艦の搜索・探知については、AI 技術の導入によって必ずしも容易になるとは限らないとの指摘がある。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. 36-37.

33 Lili Tcheang, "Are Artificial Neural Networks like the Human Brain? And Does It Matter?" *Medium*, November 7, 2018, <https://medium.com/digital-catapult/are-artificial-neural-networks-like-the-human-brain-and-does-it-matter-3add0f029273>.

たものであり、人工ニューロンを多数組み合わせた人工ニューラルネットワークを多層的に構成することで、より複雑な処理が可能になるとされている³³。

ISR能力の向上にAI技術を導入する試みとしては、2017年に開始された米国防総省の「プロジェクト・メイヴン (Project Maven)」が挙げられる。これは、米軍のドローンが撮影・記録した膨大な量の映像データをAI技術と画像認識技術によって迅速に処理・分析し、それらに映っている対象物を識別しようというプロジェクトで、「イスラム国 (ISIS)」を対象とする映像の分析を目的としたものであった³⁴。これと同様の技術が、核兵器関連の早期警戒およびISR能力の向上にも活用できると考えられる³⁵。また、「プロジェクト・メイヴン」との関連は不明であるが、アメリカは偵察衛星やドローンなどが撮影した大量のデータの中から北朝鮮などの核ミサイル運搬用輸送起立発射機 (transporter-erector-launcher: TEL) の情報を自動的に抽出するためのアルゴリズムの開発に着手したとも報じられている³⁶。

(2) 導入をめぐる議論

核攻撃早期警戒システムにAI技術を導入して情報分析能力を向上させ、相手国の核兵器を正確に探知・追跡できるようになれば、核態勢を相互に監視することによって透明性が高まり、核戦争の可能性が減少するとの見方がある。ランド研究所のエドワード・ガイスト (Edward Geist) らは、自国が早期警戒システムにAI技術を導入することによって相手国が秘密裏に核攻撃の準備を行うことが難しくなれば、相手国による核使用の脅しの信憑性が低下し、核戦争のリスクは減少するであろうと述べている。ただし、このためには相手国も自国と同等の情報分析能力を獲得し得ることが必要であるという³⁷。

他方、危機においては早期警戒システムへのAI技術の導入が核保有国相互の緊張を高め、紛争のエスカレーションを招く恐れがあるとも指摘されている³⁸。レスター大学のジェームズ・ジョンソン (James Johnson) は、核攻撃に対する早期警戒システ

34 Kelsey D. Atherton, "Targeting the Future of the DoD's Controversial Project Maven Initiative," C4ISRNET, July 27, 2018, <https://www.c4isrnet.com/it-networks/2018/07/27/targeting-the-future-of-the-dods-controversial-project-maven-initiative/>.

35 SIPRI, "The Impact of Artificial Intelligence, Volume I," p. 54.

36 Marcus Weisgerber, "The Increasingly Automated Hunt for Mobile Missile Launchers," *Defense One*, April 28, 2016, <https://www.defenseone.com/technology/2016/04/increasingly-automated-hunt-mobile-missile-launchers/127864/>.

37 Geist and Lohn, "How Might Artificial Intelligence Affect the Risk of Nuclear War?" p. 21.

38 SIPRI, "The Impact of Artificial Intelligence, Volume II," pp. 64-65.

39 James Johnson, "Artificial Intelligence & Future Warfare: Implications for International Security," *Defense & Security Analysis*, vol. 35, no. 2 (2019), p. 152, <https://www.tandfonline.com/doi/pdf/10.1080/14751798.2019.1600800?needAccess=true>.

ムに AI 技術を導入した場合、核使用の決定に係る時間が短縮され、核レベルにおける危機の安定性に悪影響を及ぼすと分析している³⁹。

先述した TEL の探知については、AI が TEL の情報を正確に抽出するためには、TEL に関する正確な画像データを十分に含む大量のデータを AI に学習させておく必要があるが、実際に収集された画像データは移動中の民間車両や TEL 以外の軍用車両が圧倒的多数を占めており、TEL の情報は極端に少ないため、AI による高精度の分析は困難だとされている。これは「データの不均衡」の問題として知られており、TEL に関する AI の学習データにある程度の拡張性を持たせるなどの技術的な対策はあるものの、根本的な解決は難しいという。これに加えて、撮影された 2 次元の画像から 3 次元の TEL の形を AI に高精度で類推させることの困難性、あるいは対象国が TEL に偽装や隠蔽を施す蓋然性などを考え合わせれば、AI による分析の精度はさらに落ちる。このため、ローレンス・リバモア国立研究所のジョセフ・ジョンソン (Joseph Johnson) は、発見した TEL を破壊するための対兵力攻撃を検討する場合、目標情報の確度は 100% に近いことが求められるが、その精度を達成することは AI 技術を導入しても困難であろうと指摘する⁴⁰。

早期警戒システムへの AI 導入と偶発的核戦争のリスクとの関係については様々な議論が展開されている。肯定的なものとしては、AI 技術を導入して信頼性の高い早期警戒システムを構築することができれば、先制攻撃に係る安定性 (first-strike stability) が高まり、危機の際に偶発的なエスカレーションの危険を減少させ得るとの見方がある⁴¹。

逆に、次のような否定的な議論もある。早期警戒システムに AI 技術を導入して情報分析能力を向上させようとする場合、核使用に係る人間の判断は AI が処理・分析したデータに基づいて行われることになる。この際、AI がなぜそのような分析に至ったのかを論理的に解明することが不可能なために、人間は AI の出した分析結果を信用すべきか否かの深刻なディレンマに直面する。また、訓練データに含まれたバイアスによって AI のアルゴリズムが実際の状況を誤って解釈し、不正確な分析データを提示する恐れもある。これらの問題が、偶発的なエスカレーションのリスクを生む可能性は排除できないと指摘されている⁴²。

40 Joseph Johnson, "MAD in an AI Future?" Center for Global Security Research, Lawrence Livermore National Laboratory, June 2019, pp. 2-6, <https://www.osti.gov/servlets/purl/1527284>.

41 Geist and Lohn, "How Might Artificial Intelligence Affect the Risk of Nuclear War?" p. 21.

42 SIPRI, "The Impact of Artificial Intelligence, Volume I," p. 92.

4. AI技術が核抑止に及ぼす影響

(1) 核兵器システムの自律化による影響

ペンシルヴァニア大学のマイケル・ホロヴィッツ (Michael C. Horowitz) によれば、ある核保有国の指導者が自国の第二撃能力に確信を持てなくなるにつれて、当該国はAIによる核兵器システムの自律化を進める可能性が高くなるという。なぜなら、AIによって自律化された核兵器システムは人間が運用するよりも速く反応でき、また人間に特有の感情や錯誤などの影響を受けないために精度も高くなると考えられるからである。しかし、自律化によるリスクも高まる。例えば、核兵器の早期警戒システムをAIによって自律化すれば、敵の核兵器をより速く確実に識別できるようになると考えられる⁴³。ただし、現在のAIの画像認識技術では、画像撮影時の天候の違いや対象物の画像に別の物が写り込むなどの僅かな環境やタスクの差異で、対象物を正しく認識できなくなることが報告されている⁴⁴。このため、AIが全く別の物を敵の核兵器と誤認識し、偶発的な核使用を引き起こす可能性は否定できない。

これに関連して、核兵器システムへのAI技術の導入は偶発的核戦争のリスクを高めるといった否定的な意見もある。例えば、ある核保有国が一方的に新たなAI技術の導入を進めた場合、他の核保有国は自国の核戦力の近代化や核ドクトリンの変更、あるいは警戒態勢の強化といった措置をとって対抗しようとするであろう。こうした状況自体、偶発的核戦争へのエスカレーションを起こしやすくするものであり、その意味ではAI技術は偶発的核戦争の原因になり得る⁴⁵。

第三者によるデータの不正操作が意図しない核使用につながるリスクについても議論されている。国際戦略研究所アメリカ支部のマーク・フィッツパトリック (Mark Fitzpatrick) は、核兵器システムにAI技術を導入しつつある核保有国同士の国際危機において、第三者がそれぞれのAIの訓練データに不正なデータを紛れ込ませて分析結果を操作しようとするデータ・ポイズニング攻撃を行う可能性を指摘している。フィッツパトリックによれば、このような攻撃を受けた核保有国は軍の警戒態勢を引き上げるとともに、他の情報ソースを用いてAIの分析結果が実際の状況を正しく反映しているのかを確認するために貴重な時間を使うことになる。この間、AIのアルゴ

43 Taylor Sullivan, "NTI Seminar: A Stable Nuclear Future? Autonomous Systems, Artificial Intelligence and Strategic Stability with UPenn's Michael C. Horowitz," Nuclear Threat Initiative, November 15, 2018, <https://www.nti.org/analysis/atomic-pulse/nti-seminar-stable-nuclear-future-autonomous-systems-artificial-intelligence-and-strategic-stability-upenns-michael-c-horowitz/>.

44 Dan Hendrycks, et al., "Natural Adversarial Examples," Cornell University, July 16, 2019, pp. 1-14, <https://arxiv.org/pdf/1907.07174.pdf>.

45 Boulanin, "AI & Global Governance".

リズムは直ちに対応行動をとるよう人間に督促し続けることから、偶発的なエスカレーションのリスクが高まるという⁴⁶。

AI 技術は、NC3 および早期警戒システムのみならず、核兵器運搬手段にも導入されていくと考えられる。AI を装備した核兵器運搬手段は、発射されれば外部からの指令によることなく自律的に目標に向かって進んでいくことができるので、電波妨害による通信の途絶や通信システムがハッキングされるといったリスクを克服し得ると考えられる⁴⁷。その反面、一度発射されれば指令によって呼び戻す手段が無いため、攻撃を中止させることは不可能となり、エスカレーション・コントロールが失われる恐れもある⁴⁸。

先述したように、ロシアは AI 技術を導入した核搭載可能 AUV を開発しており、2018 年 3 月の年次教書演説でプーチン大統領がその概要を初めて公表した。それによると、開発中の AUV は海中深く潜航しつつ大陸間を高速で移動可能であり、敵のいかなる対抗策にも影響されることなく、相手国の沿岸部などの様々な目標を攻撃できる戦略兵器だという⁴⁹。この AUV は「ポセイドン」と名付けられ、潜水艦に搭載して発射される模様であり、その運用試験を行うための特殊用途原子力潜水艦の建造も進められている⁵⁰。

「ポセイドン」をめぐるのは、偶発的核戦争のリスクを生む可能性が示唆されている。ホロヴィッツらは、「ポセイドン」は AI 技術を導入した新たな核兵器とされているが、ロシアが企図するような限定核使用のオプションにはさほど裨益しないと分析している。このため、仮にロシアが西側諸国との紛争の際に「ポセイドン」を使用するとの脅しを伝え、紛争をロシア側に有利なように収束させようと企図したとしても、アメリカはロシアがもし核の敷居を越えるならば同種の対応をとると明言するであろう。この場合、ロシアによる「ポセイドン」の使用あるいはその脅しはアメリカとの核の相互エスカレーションを引き起こす可能性が高くなるという⁵¹。

核兵器運搬手段への AI 技術の導入について、現時点で具体的に開発計画が進めら

46 Mark Fitzpatrick, "Artificial Intelligence and Nuclear Command and Control," IISS, April 26, 2019, <https://www.iiss.org/blogs/survival-blog/2019/04/artificial-intelligence-nuclear-strategic-stability>.

47 SIPRI, "The Impact of Artificial Intelligence, Volume I," p. 81.

48 Paul Scharre, *Army of None: Autonomous Weapons and the Future of War* (New York and London: W. W. Norton & Company, 2018), p. 304.

49 President of Russia, "Presidential Address to the Federal Assembly," March 1, 2018, <http://en.kremlin.ru/events/president/news/56957>.

50 2019 年 4 月、「ポセイドン」搭載試験艦として最初の原子力潜水艦「ベルゴロド」がロシア北部のセヴマシュ造船所で進水し、さらに 2020 年春には 2 番目の試験艦「ハバロフスク」の進水が予定されている。Franz-Stefan Gady, "Russia's Second Poseidon Underwater Drone-Carrying Submarine to be Launched in 2020," *The Diplomat*, May 15, 2019, <https://thediplomat.com/2019/05/russias-second-poseidon-underwater-drone-carrying-submarine-to-be-launched-in-2020/>.

51 Horowitz, Scharre, and Velez-Green, "A Stable Nuclear Future?" pp. 21-22.

れているものは「ポセイドン」以外には無いと思われるが、理論的には核ミサイルや戦略爆撃機などにAI技術を導入することも考えられよう。

ミサイルについては、極超音速滑空飛翔体（hypersonic glide vehicle: HGV）にAI技術を導入するための基礎研究が中国で盛んに行われているとの指摘がある。東西研究所のローラ・サールマン（Lora Saalman）によれば、HGVはアメリカのミサイル防衛を突破し得る次世代の核弾頭としての可能性を有する一方で、極超音速滑空の飛行制御や機動性の確保などが課題とされている。それらの課題を解決すべく、多くの中国の研究者がAIのニューラルネットワーク技術をHGVに組み込むための研究に取り組んでいる⁵²。

（2）非核兵器システムにAI技術が導入された場合の影響

近年、非核兵器へのAI技術の導入が国際的に懸念されている。特に、自律型致死性兵器システム（LAWS）の規制をめぐることは、2014年から特定通常兵器使用禁止制限条約（CCW）の枠組みで専門家による議論が行われている⁵³。

他方、LAWSの議論とは別に、非核兵器へのAI技術の導入に伴い、核抑止に影響が及ぶ可能性を指摘する声もある。例えば、前出のジェームズ・ジョンソンは、小規模な核戦力の保有国がAIを搭載した小型無人機の集団によるスウォーム攻撃を受けた場合、当該国は核戦力を使うか失うかの状況に陥り、核使用を迫られる恐れがあると分析している⁵⁴。また、ローレンス・リバモア国立研究所のザカリー・デイヴィス（Zachary Davis）は、AI技術がミサイル防衛システムに導入された場合、弾道ミサイル目標の捕捉・追跡・識別能力が飛躍的に向上するため、新たな攻撃力と組み合わせることによって奇襲攻撃の誘因が高まり、核保有国間の戦略的安定性を揺るがしかねないと指摘する⁵⁵。

ミサイル防衛と組み合わせ得る新たな攻撃力として、米中露3カ国が鎬を削って開発中の極超音速兵器にAI技術を導入することも考えられる⁵⁶。

極超音速兵器をめぐることは、中露が核・非核両用兵器として開発を進めているのに

52 Lora Saalman, "China's Integration of Neural Networks into Hypersonic Glide Vehicles," in *AI, China, Russia, and the Global Order: Technological, Political, Global, and Creative Perspectives*, ed. Nicholas D. Wright (December 2018), pp. 153-160, https://nsiteam.com/social/wp-content/uploads/2019/03/AI-China-Russia-Global-WP_FINAL2-compressed.pdf.

53 佐藤丙午「AI（人工知能）と安全保障」『海外事情』第66巻3・4号（2018年3月）21頁。

54 James Johnson, "Artificial intelligence: A Threat to Strategic Stability," *Strategic Studies Quarterly*, vol. 14, issue 1 (Spring 2019), p. 19, https://www.airuniversity.af.edu/Portals/10/SSQ/documents/Volume-14_Issue-1/Johnson.pdf.

55 Davis, "Artificial Intelligence on the Battlefield," p. 7.

56 Johnson, "Artificial intelligence: A Threat to Strategic Stability," p. 25.

対し、アメリカは核弾頭を搭載する予定はないとされている⁵⁷。中露は、アメリカが極超音速兵器技術を活用した新たな非核攻撃力により両国の核抑止力を弱体化させようとするのではないかと恐れている。カーネギー国際平和財団のジェームズ・アクトン（James Acton）は、アメリカによる非核対兵力攻撃（conventional counterforce）の可能性がエスカレーションのリスクを高め、中露との核抑止関係を不安定化させる恐れがあると指摘する⁵⁸。今後、アメリカが非核極超音速兵器およびミサイル防衛にAI技術を導入するようになった場合、米中露の戦略的安定性はますます危殆に瀕することになりかねない。このような事態を回避するために、米中露がAIなどの新興技術の影響を含めた軍備管理に取り組んでいくことが期待される。

おわりに

これまで述べてきたように、近年におけるAI技術の急速な進歩は、将来の核抑止に少なからず影響を及ぼしていくものと考えられる。AI技術は、核抑止の安定化に寄与する可能性を有する一方で、その不安定化をもたらす恐れもある。現状においては、AI技術は未成熟な段階にあり、核兵器システムへの導入についてもロシアの「ポセイドン」以外に具体的な計画は無い。ただし、AI技術が本格的に核兵器システムに導入されていくのはそれほど遠い将来ではないかもしれない。特に、アメリカの核兵器システムにAI技術が導入されていけば、わが国の安全保障にも影響が及ぶことになると思われる。

本稿で紹介したランド研究所やSIPRIのワークショップのように、AI技術が核抑止に及ぼす影響について先を見据えた議論を行う必要がある。現在、CCWの枠組みで行われているLAWSの議論は非核兵器を焦点としたものであり、核兵器システムにAI技術が導入された場合の影響や問題点については語られていない⁵⁹。今後は、LAWSの議論と並行して、AIと核抑止との関係についての議論も高まっていくと思われる。その際、1983年のペトロフ事案をはじめとする過去の核兵器システム自動化

57 Kelley M. Saylor, “Statement before United States Commission on Security and Cooperation in Europe Hearing on ‘AI, UAVS, Hypersonics, and Autonomous Systems: Emerging Technologies and Euro-Atlantic Security,’” Congressional Research Service, 7-5700, January 22, 2020, pp. 5-7, https://www.csce.gov/sites/helsinkicommission.house.gov/files/Saylor%20-%2001222020%20Helsinki%20Commission%20Testimony%20-%20Cleared_0.pdf

58 James M. Acton, ed., “Entanglement: Russian and Chinese Perspectives on Non-Nuclear Weapons and Nuclear Risks,” Carnegie Endowment for International Peace, 2017, pp. 77-83, https://carnegieendowment.org/files/Entanglement_interior_FNL.pdf.

59 Boulanin, “AI & Global Governance”.

の事例は、将来の核抑止において AI と人間との関係をいかに律すべきかを考える上で有益な示唆を与えるであろう。

（防衛研究所）