

〈3〉 軍事 AI に関する国際的な議論の動向及び 取組みのご紹介

アレシア国際法律事務所 弁護士 有本 真由

1 序論

人工知能（以下「AI」）¹の技術革新は、軍事分野においても急速に進展しており、従来の戦争概念や方法を大きく変容させつつあり、軍事用途で利用される AI（以下「軍事 AI」²）は、国際社会に倫理的・法的な重大課題を突きつけている。それに伴い、AI の軍事利用をめぐるガバナンス形成、国際的ルール作りの必要性、AI による判断の透明性・説明可能性といった論点が浮上しており、現在、複数の国際的な場において議論が進行中である。

本稿では、まず、本稿を読み進めるにあたって必要となる概念を整理した上で（本稿 2）、自律型致死兵器システム（LAWS）をめぐる国連における議論の現状を確認する（本稿 3）。次に、マルチステークホルダー・アプローチの一環として注目される「軍事領域における責任ある AI（Responsible AI in the Military Domain）」（以下「REAIM」）イニシアチブや、米国国務省による「AI と自律性の責任ある軍事

利用に関する政治宣言」（本稿 4）、AI の倫理的な軍事利用にかかる制度的対応の先行事例として米国国防総省（以下「DoD」）の取組みを紹介する（本稿 5）。さらに、AI 兵器に求められる法的審査（legal review）に関する議論状況を整理する（本稿 6）。

なお、本稿に記述する情報は、私見の部分を除き、すべて公になっているものに限られることは付言しておく。

2 軍事 AI に関する概念整理

2.1 OODA ループと人間の関与

OODA ループとは、「Observe（観察）」「Orient（状況把握・分析）」「Decide（意思決定）」「Act（行動）」の 4 段階から構成される意思決定サイクルであり、1980 年代に米空軍のジョン・ボイド大佐によって開発された概念である³。戦闘行動においては、このループをいかに高速かつ正確に回すかが優位性の鍵を握るとされており⁴、AI はこのループの各段階に

¹ 現在、国際的に統一された「AI」の定義は存在しない。それは軍事分野についても同様である。本稿においては、「AI」を「パターンの認識、経験からの学習、結論の導出、予測、行動の実行など、通常は人間の知能を必要とするタスクを機械が実行する能力」（後述の 2022 年 RAI 戦略 & 実施パスウェイにおける定義）と広義にとらえる。また、議論の対象となる文書において AI の定義が明示されている場合には記載するようにした。

² “military AI” の訳語であるが、公式の用語ではないため留意されたい。

³ 本項について、RAND “Military Applications of Artificial Intelligence” 34-35 頁（なお、本稿において、文書の表題で検索できるものについては URL は省略させていただく。）。

⁴ これは、「get inside the enemy’s decision cycle（敵の意思決定サイクルに介入する）」と表現されることがある。敵よりも早く OODA ループを回して Act（攻撃や撤退等）に移ることが重要であることを意味する（前掲 RAND, 34 頁）。

において補完・加速の役割を担う。たとえば、センサー情報のリアルタイム処理（Observe）、敵情勢の自動解析（Orient）、最適行動の提案（Decide）、自律的な実行（Act）というように、同ループの各局面において AI は機能しうる。

OODA ループに AI が組み込まれることで、意思決定のスピードと精度が格段に上がる。他方で、特に、Decide、Act の局面を全面的に AI に委ね、それによって誤爆等が生じた場合、倫理的・法的責任の所在が大きな問題となる。

そこで、ループの各局面、特に Act 前の時点における人間の関与のあり方について、以下の3つの概念が提唱されている。

- ・ Human in the loop: AI による実行の前に必ず人間による積極的な確認・承認を必要とする場合。
- ・ Human on the loop: AI が自律的に判断・実行するが、人間が監視し、必要に応じて AI の実行を停止できる場合。
- ・ Human out of the loop: AI が完全に自律的に実行し、人間の関与が介在しない場合。

これらの区分は、軍事 AI の倫理的評価や法的責任の議論において中心的な要素であり、とりわけ「致死性のある行動」における人間の判断の関与は、国際人道法上の要請と密接に関係している。なお、後述の LAWS として禁止されるべきと議論されている兵器は、主として、Human out of the loop に着目したものであると言える。

2.2 軍事 AI の対象と本稿の対象

軍組織で使用される AI は、その運用目的に応じて、次の3つに分類されうる⁵。

- ・ Enterprise AI: 会計、行政支援など組織業務の効率化に資する AI。
- ・ Mission Support AI: 指揮統制、情報分析、作戦立案支援などの作戦行動の補助のための AI。
- ・ Operational AI: 実際の戦闘・作戦行動における使用を目的とした AI。

本稿で扱う「軍事 AI」は、主として Mission Support AI 及び Operational AI を対象とし、Enterprise AI（業務支援系）は含まないものとする。

3 国連における LAWS の議論動向

3.1 経緯

自律型致死兵器システム（lethal autonomous weapons systems）（以下「LAWS」）の国連における議論は、2014 年から特定通常兵器使用禁止制限条約（Convention on Certain Conventional Weapons）（以下「CCW」）の枠組みにおける非公式会合にて始まった。その前年に、Killer Robots 阻止キャンペーンが市民団体によって繰り上げられたことを端緒とする。2017 年からは、議論の場を CCW の政府専門家会合（Group of Governmental Experts）（以下「GGE」）に移し、2019 年には LAWS に関する指針（Guiding Principles）が CCW 締約国会議で承認された。もっとも、コンセンサス方式をとる CCW の枠組みでは一部の国の慎重姿勢によって法的拘束力のある規範形成の進展が阻まれていたため、2023 年、国連総会において、LAWS に関する報告書の作成を国連事務総長に求める決議が採択された⁶。それを受けて、2024 年 7 月、国連事務総長による報告書⁷が公表され、LAWS・GGE のプラットフォームを活かしつつ、2026 年までには法的拘束力のある法文書（instrument）を締結することが強く呼びかけられた。

このように 2014 年から議論されてきている LAWS であるが、その定義は未だ定まっていない。日本としては、「一度起動すれば、操作者の更なる介入なしに標的を識別し、選択し、殺傷力を持って交戦することができる」という特徴を備えている兵器システムとしている⁸。

なお、LAWS・GGE では、LAWS だけでなく、自律型兵器システム（autonomous weapons systems）（以下「AWS」）の禁止又は規制も議論されているため、留意されたい。

⁵ この区分については、RAND “The Department of Defense Posture for Artificial Intelligence Assessment and Recommendations” 25-27 頁。

⁶ 外務省 HP: https://www.mofa.go.jp/mofaj/dns/ca/page24_001191.html

⁷ 原文: <https://docs.un.org/en/A/79/88>

⁸ 前掲外務省 HP

3.2 LAWS に関する指針及び 2023 年 GGE 報告書

LAWS・GGE は、2017 年以降毎年開催されており、多くの場合、その成果としてコンセンサスにより最終報告書がまとめられている。2019 年 9 月、LAWS

に関する指針⁹が 2019 年 GGE 報告書の別紙 IV として確認され、CCW 締約国会議で承認された。その概要は、下表のとおりである。

国際人道法の全面適用	国際人道法は、すべての兵器システム（LAWS の開発と使用を含む。）に引き続き全面的に適用される
人間の責任	兵器システムの使用に関する決定における人間の責任は保持されなければならない。この点は、兵器システムのライフサイクル全体にわたって考慮されるべきである
ヒューマン・マシン・インタラクション	ヒューマン・マシン・インタラクションは、様々な形態をとり、兵器のライフサイクルの様々な段階で行われる可能性があるところ、新興技術に基づく兵器システムの使用が、国際法（特に国際人道法）に準拠することを確保する必要がある。ヒューマン・マシン・インタラクションの質と範囲を決定する際には、運用状況、兵器システム全体の特性と能力を含む、様々な要素を考慮する必要がある
説明責任	兵器システムの開発・配備・使用に関する説明責任は、国際法に従って確保されなければならない
法的審査	国際法に基づく国家の義務に従い、新たな兵器、又は戦闘手段・方法の研究・開発・調達・採用においては、その使用が一部又はすべての状況において国際法によって禁止されるかどうかを判断しなければならない
セキュリティその他のリスク	兵器システムを開発・調達にあたっては、物理的セキュリティ・サイバーセキュリティ・テロ集団による取得のリスク・拡散リスク等を考慮しなければならない
リスク評価と緩和策	リスク評価と緩和策は、兵器システムにおける新興技術の設計・開発・試験・配備サイクルの一環として実施されなければならない
新興技術の使用にあたっての国際法の検討	LAWS 分野における新興技術の使用にあたっては、国際人道法及びその他の適用される国際法上の義務を遵守しているかを検討しなければならない
擬人化禁止	政策措置を策定する際には、LAWS 分野における新興技術を擬人化すべきではない
平和利用の促進	CCW の文脈で行われる議論や政策措置は、高度な自律技術の平和利用の進歩やアクセスを妨げるべきではない
CCW の枠組みの確認	CCW は、軍事上の必要性和人道的配慮のバランスをとることを目的とする条約の趣旨と文脈において、LAWS 分野における新興技術の問題に対処するための適切な枠組みを提供している

その後、2023 年に採択された GGE 報告書¹⁰では、改めて上記指針が確認された。その他、同報告書では、各国は、LAWS 分野の新興技術に基づく兵器システムのライフサイクル全体を通じて、国際法、特に国際人道法に基づく義務の遵守を確保する必要があるとして、必要に応じて、とりわけ、①システムが攻撃できる標的の種類を制限し、②兵器システム運用の期間・地理的範囲・規模を制限し、③人間のオペレータに適切な訓練と指示を提供すべきである、とされた。

なお、この GGE 報告書には、日米英独仏のほか、

中国やロシアを含めて 88 か国（オブザーバ国を除く。）が参加した。

3.3 日本の対応・立場

日本は、LAWS に関し、2019 年 3 月にその立場を示す作業文書を提出し、2023 年 3 月には、米英加豪韓（後にポーランドも参加）と共に、作業文書「国際人道法を基礎とした禁止と制限の方法にかかる自律型兵器システムに関する条項案（Draft articles on autonomous weapon systems – prohibitions and other regulatory measures on the basis of international

⁹ CCW/GGE.1/2019/3, Annex IV

¹⁰ CCW/GGE.1/2023/2, 4 頁