

国立国会図書館 調査及び立法考査局

Research and Legislative Reference Bureau
National Diet Library

論題 Title	軍事分野における AI 利用とその規制の動向—AI を用いた意思決定支援システムと法的審査の適用可能性を中心に—
他言語論題 Title in other language	The Use of AI in the Military Domain and Trends in Its Regulation: With a Focus on AI-based Decision Support Systems and the Applicability of Legal Review
著者 / 所属 Author(s)	久古 聡美 (KYUKO Satomi) / 国立国会図書館調査及び立法考査局 外交防衛課
雑誌名 Journal	レファレンス (The Reference)
編集 Editor	国立国会図書館 調査及び立法考査局
発行 Publisher	国立国会図書館
通号 Number	898
刊行日 Issue Date	2025-10-20
ページ Pages	49-75
ISSN	0034-2912
本文の言語 Language	日本語 (Japanese)
摘要 Abstract	軍事分野における AI 利用の規制に向けた国際的な取組の現状を概観した後、武力行使に関する意思決定のために用いられる AI-DSS のメリットとリスク、法的審査の適用の可能性等を整理する。

- * この記事は、調査及び立法考査局内において、国政審議に係る有用性、記述の中立性、客観性及び正確性、論旨の明晰（めいせき）性等の観点からの審査を経たものです。
- * 本文中の意見にわたる部分は、筆者の個人的見解です。

軍事分野における AI 利用とその規制の動向

—AI を用いた意思決定支援システムと法的審査の適用可能性を中心に—

国立国会図書館 調査及び立法考査局
外交防衛課 久古 聡美

目 次

はじめに

I 軍事分野における AI 利用の規制に向けた国際的な取組

- 1 国際的な取組の状況
- 2 規制の方向性と展望

II AI を用いた意思決定支援システム（AI-DSS）の概要及び利用事例

- 1 意思決定支援システムとは
- 2 意思決定のプロセスと AI-DSS による支援
- 3 AI-DSS の利用事例

III AI-DSS の利用に伴うメリットとリスク

- 1 メリット
- 2 リスク
- 3 国際人道法の義務を履行する上での有効性
- 4 リスク軽減措置の提案

IV AI-DSS への法的審査の適用に向けた課題と展望

- 1 法的審査の概要
- 2 法的審査の適用に向けた課題

おわりに

キーワード：人工知能、意思決定支援システム、AI-DSS、REAIM、法的審査、兵器審査

要 旨

- ① 近年、軍事分野において、AI を用いた意思決定支援システム（AI-DSS）の利用が広がりつつある。AI-DSS は、武力行使に関する意思決定のために用いられる場合、軍に新たな機会を提供する一方、文民に危害が及ぶような不適切な選択に導いたり、戦闘において意図しない悪影響を及ぼしたりする可能性もあり、リスク軽減のための対応措置や規制が模索されている。
- ② 軍事分野全般にわたる AI 利用に関して、2023 年以降、有志国による「軍事領域における責任ある AI」（REAIM）、米国主導の政治宣言、国連などの枠組みで、国際的な規制に向けた取組が進められている。AI 利用への国際法の適用、人間の適切な関与を維持する必要性など、大きな原則のレベルでは、多国間で見解が一定程度集約されつつある。
- ③ AI-DSS の利用に伴うリスクとして、AI が学習データの質と量に依存し、バイアス、データ汚染や欺瞞（まん）的情報の影響を受けやすいという技術上の限界から、AI-DSS が想定外の挙動をしたり性能が低下したりする可能性がある。また、AI の作用が人間にとってブラックボックス状態であること、さらに、人間の認知的バイアス、軍事作戦のテンポの高速化も受けて、人間が AI システムに批判的に関与できず、受動的な監督者となる可能性も懸念されている。
- ④ AI-DSS は、利用方法によっては、人間による国際人道法の遵守を促進する可能性を有する。一方、武力行使を規制する国際人道法の重要な規則の多くが、定性的で文脈に応じた価値判断の適用を前提とし、現在の技術水準では AI による処理に適していないことも指摘されている。リスク軽減に向けた中心的な課題としては、可能な限りの技術的な対応を行うことに加えて、人間の側が AI-DSS を適切に利用し、システムを効果的に統合できるかが問われる。
- ⑤ AI-DSS の利用に伴うリスク軽減のための措置として、一つには、研究・開発段階から、システムの適切な試験及び評価を行うことが挙げられる。また、その法的側面からの評価に向け、AI-DSS に対し、ジュネーブ諸条約第一追加議定書の第 36 条に基づき、新たな兵器等の導入に当たってその使用の合法性を審査するための「法的審査」を適用する可能性が議論されている。
- ⑥ 「法的審査」は、国際人道法を遵守した利用ができない新たな兵器等の開発や使用に対する効果的な予防策となり得る。一方、AI-DSS への「法的審査」の適用に向けては、審査の対象となり得るか、審査をどのタイミングで、どのような基準で行うべきかといった課題もある。

はじめに

人工知能（Artificial Intelligence: AI）は様々な応用の可能性を持つ汎用技術であり、軍事分野においても、情報収集・警戒監視・偵察（Intelligence, Surveillance and Reconnaissance: ISR）、指揮統制、後方支援、事務処理などの幅広い側面での利用が模索されている。近年では、いわゆる、AI を用いた意思決定支援システム（AI-based Decision Support Systems: AI-DSS）の利用が広がりつつあり、国際的に注目を集めている。2023 年 10 月以降のパレスチナのガザ地区における紛争においても、AI-DSS の実戦での使用例が報告されている。

武力行使に関する意思決定のために用いられる AI-DSS は、軍に新たな機会を提供する一方、文民⁽¹⁾に危害が及ぶような不適切な選択に導いたり、戦闘において意図しない悪影響を及ぼしたりする可能性がある。このような重大なリスクを軽減するためには、適切な対応措置や規制が求められるが、現状では、AI を用いたシステムに対して具体的にどのような措置がとられるべきか、何らかの法的拘束力を持つ規制が必要かなどについて模索されている段階にある。こうした中、AI-DSS のリスク軽減のため、国際人道法⁽²⁾の既存の枠組みを活用して実施することが期待されているものとして、兵器等の「法的審査」（legal review）がある。国際人道法の主要な条約の一つであるジュネーブ諸条約第一追加議定書⁽³⁾（以下「議定書 I」という。）の第 36 条には、新たな兵器、戦闘の手段又は方法の合法性を審査するための義務が定められており、これを AI-DSS に適用する可能性が議論されている。

本稿では、まず、軍事分野における AI 利用の規制に向けた国際的な取組の現状を概観する（I 章）。次に、様々な AI の軍事的応用のうち、国際人道法に違反するリスクを伴う例として、武力行使に関する意思決定のために用いられる AI-DSS に焦点を当て、その概要及び近年の利用事例を確認する（II 章）。続いて、AI-DSS の利用に伴う主なメリット、リスク及び国際人道法の義務を履行する上での有効性に関する議論を整理し、リスク軽減に向けた主な対応措置の提案にも触れる（III 章）。最後に、AI-DSS への適用が期待されている法的審査の概要と適用に向けた課題を検討する（IV 章）。

* 本稿におけるインターネット情報の最終アクセス日は、2025 年 8 月 29 日である。なお、条約の日本語訳は、IV 章を除く部分において、公定訳に従った。ただし、漢数字はアラビア数字に改めた。[] 内は、筆者による補記である。

(1) 文民（civilians）とは、軍隊構成員でなく、かつ群民兵でない全ての者をいう。黒崎将広ほか『防衛実務国際法』弘文堂、2021、p.342。

(2) 国際人道法は、武力紛争の際に紛争当事者に適用される国際法規則の総体をいい、主として、傷病者、捕虜、文民などの武力紛争の犠牲者を保護するための規則、戦闘の手段及び方法に関する規則から構成される。客観的に武力紛争が存在することによって適用される。なお、適用対象となる武力紛争には、国際的武力紛争と非国際的武力紛争（武装集団が責任ある指揮の下で組織化され、一定の烈度を持つという二つの要件を満たす場合）とがあり、適用される国際法は武力紛争当事者の地位に関する規則を中心に異なる。岩沢雄司『国際法 第 2 版』東京大学出版会、2023、pp.728-729, 738-741; 黒崎ほか 同上、pp.267, 313-314。

(3) 正式名称は、「1949 年 8 月 12 日のジュネーブ諸条約の国際的な武力紛争の犠牲者の保護に関する追加議定書（議定書 I）」（Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I), 1125 UNTS 3）。

I 軍事分野における AI 利用の規制に向けた国際的な取組

1 国際的な取組の状況

軍事分野全般にわたる AI 利用に関しては、2023 年以降、国際的なフォーラムにおいて規制⁽⁴⁾に向けた具体的な取組がみられるようになった。軍事分野の中では、AI 利用をめぐり、自律型致死兵器システム (Lethal Autonomous Weapon Systems: LAWS)⁽⁵⁾の問題が先行して注目を集めてきた。特定通常兵器使用禁止制限条約 (Convention on Certain Conventional Weapons: CCW)⁽⁶⁾の枠組みの下では、2014 年以降、LAWS に関する議論が継続して行われている。2019 年 8 月の CCW の LAWS に関する政府専門家会合 (Group of Governmental Experts: GGE)⁽⁷⁾では、国際人道法の適用、人間が使用の決定に責任を負うことなど、LAWS に関する一定の共通理解を示した 11 項目から成る指導原則 (Guiding Principles) が採択された⁽⁷⁾。また、2024 年 12 月には、国連総会で LAWS に関する決議が採択され、国際法の適用、武力行使における人間の役割の重要性、包括的な議論を行うことの重要性などが謳 (うた) われている⁽⁸⁾。一方、AI-DSS を含む軍事分野のその他の AI 利用に関しては、しばらく議論から抜け落ちたような状況であったが⁽⁹⁾、2023 年頃から、AI 技術の飛躍的な進展も背景としてそのリスクが認識されるように

(4) 本稿において、「規制」の語は、法的拘束力の確立した条約などの法文書 (ハード・ロー) に基づく規制のみならず、法的拘束力を持たない行動規範や指針等の非法文書 (ソフト・ロー) に基づく規制も含む概念として用いる。便宜的に、前者をハードな規制、後者をソフトな規制と記す。なお、国際法の分野では、完全な法としてのハード・ローと対照する形で、国際社会の多数によって採択された決議や宣言などの文書を、法的表現を用いた柔らかな法 (ソフト・ロー) であって、規範的意味を持つとする主張がある。例えば、国際会議の宣言、国連総会決議、国際組織の行動綱領・指針などがソフト・ローの例とされる。岩沢 前掲注(2), pp.42-43.

(5) LAWS の国際的な定義は定まっていない。日本政府は、2024 年 5 月に国連に提出した作業文書の中で、LAWS について、「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もの」と説明している。“Working Paper Submitted by Japan to the United Nations on Emerging Technologies in the Area of Lethal Autonomous Weapon Systems (LAWS),” 24 May 2024. 外務省ウェブサイト <<https://www.mofa.go.jp/mofaj/files/100687671.pdf>> なお、LAWS は殺傷力の有無を問わない自律型兵器システム (Autonomous Weapon Systems: AWS) の一種であり、近年では LAWS に限定せず、AWS 全体を対象とする議論も行われている。

(6) 正式名称は、「過度に傷害を与え又は無差別に効果を及ぼすことがあると認められる通常兵器の使用の禁止又は制限に関する条約」(Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects, 1342 UNTS 137)。

(7) “Report of the 2019 Session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems,” UN Doc. CCW/GGE.1/2019/3, 25 September 2019, p.13 (Annex IV: Guiding Principles). <<https://docs.un.org/en/CCW/GGE.1/2019/3>> 指導原則は、2019 年 11 月の CCW 締約国会議で追認された後 (UN Doc. CCW/MSP/2019/9)、2021 年 12 月の CCW 第 6 回運用検討会議で承認された (UN Doc. CCW/CONF.VI/11)。LAWS に関するこれまでの議論をまとめた文献として、例えば、次を参照。岩本誠吾「自律型致死兵器システム (LAWS) 規制の国際法政策論—倫理規範から法規範へ、人道法から軍縮法へ」『防衛法研究』46 号, 2022, pp.7-23; 佐藤丙午「AI と安全保障—LAWS 規制に向けた国際社会の議論について—」『国際開発学研究』22 巻 2 号, 2023.3, pp.67-79. <https://takushoku-u.repo.nii.ac.jp/record/2000177/files/id2202_04.pdf>

(8) “Resolution Adopted by the General Assembly on 2 December 2024,” UN Doc. A/RES/79/62, 10 December 2024. <<https://docs.un.org/en/A/RES/79/62>> 賛成 166、反対 3、棄権 15。なお、国連総会における LAWS に関する決議の採択はこれで 2 回目となる。2023 年 12 月に採択された初の決議 (UN Doc. A/RES/78/241) では、LAWS がもたらす課題と懸念に対処することの緊急性の確認のほか、国連事務総長に対して加盟国等に意見を求めた上で LAWS に関する報告書を提出するよう求めることが定められた。

(9) Giacomo Persi Paoli et al., *Modernizing Arms Control: Exploring Responses to the Use of AI in Military Decision-making*, Geneva: United Nations Institute for Disarmament Research (UNIDIR), 2020, pp.9-10. <<https://unidir.org/files/2020-08/Modernizing%20Arms%20Control%20Final.pdf>> AI-DSS の問題への注目が比較的遅くなった背景には、概念として分かりやすい有形の兵器システムである LAWS の問題が早くから注目され CCW 等の場で先行して取り上げられたこと、2022 年末頃からの AI 技術の急速な進展 (ChatGPT の一般公開等) を受けて軍事分野における AI 利用

なり、規制に向けた取組が活発化することとなった。

(1) 「軍事領域における責任ある AI」(REAIM)

軍事分野全般にわたる AI 利用の規制に向けた最初の主要な取組は、オランダ政府の主導により設置された国際的なフォーラム「軍事領域における責任ある AI」(Responsible Artificial Intelligence in the Military Domain: REAIM)である。2023 年 2 月、オランダで初の REAIM サミットが開催され、「REAIM 行動要請」(REAIM Call to Action)が採択された(2025 年 4 月時点で、58 か国・地域が賛同)⁽¹⁰⁾。同文書では、軍事分野における AI 利用について、国際的な法的義務に従った利用の重要性(前文 5 項)、適切なセーフガード(安全装置)と人間による監視を確保する重要性(本文 12 項)などが確認された。また、意思決定の形成等のための AI 利用に関しては、人間が意思決定に対する責任と説明責任を維持できるよう努めていく(本文 6 項)とする方針が示されている。

2024 年 9 月の韓国における第 2 回 REAIM サミットでは、「REAIM 行動のための青写真」(REAIM Blueprint for Action)が採択された(2025 年 8 月時点で、63 か国・地域が賛同)⁽¹¹⁾。同文書は、全般的に、前回の行動要請の方向性を基本とし、更に具体的内容に踏み込んだものとなっている。まず、その前文で、AI が軍事作戦から後方支援に至るまで、軍事のあらゆる側面を変革する可能性を持つこと(前文 1 項)、AI のライフサイクル全体を通じて責任を持って利用され、国際法、特に国際人道法を遵守した利用がなされる必要があること(前文 3 項)などが謳われている。

続いて、本文には、三つの観点から、各ステークホルダー(各国、産業界、学界、市民社会、地域・国際組織等)への呼び掛けが記されている。最初の「AI が国際の平和と安全に与える影響」の観点からは、まず、AI が国際の平和、安全、安定を維持し、損なわない方法で開発され、配備され、使用されるべきことが確認されている(本文 1 項)。また、AI が、国際人道法の履行を強化し、武力紛争における文民や民用物⁽¹²⁾の保護の取組を支援する可能性がある一方で、設計上の欠陥、データ、アルゴリズム及びその他のバイアスによる意図しない結果等に起因する予見可能・予見不可能なリスクをもたらし得ることが認識として示されている(本文 2 項、3 項)。政策上の特段の配慮に値する AI の利用としては、AI 搭載兵器(AI-enabled weapons)、AI 搭載の戦闘作戦のための意思決定支援システム(AI-enabled decision-support systems for combat operations)、サイバー作戦、電子戦及び情報作戦での利用などが挙げられて

の具体的な可能性とリスクへの認識が高まったこと、AI-DSS が人間の意思決定者とともに使用されることから責任の所在といった問題が認識され難かったことなどがあるとされる。*ibid.* 及び赤十字国際委員会(International Committee of the Red Cross: ICRC)のリチャード・レナーン(Richard Lennane)政策アドバイザーからの聞き取り調査(2025 年 3 月)等による。

(10) “REAIM Call to Action,” 2023 February 15-16. Government of the Netherlands Website <<https://www.government.nl/binaries/government/documenten/publications/2023/02/16/ream-2023-call-to-action/REAIM+2023+Call+to+Action.pdf>>; “REAIM [2023 Endorsing Countries and Territories],” [April 4, 2025]. *ibid.* <<https://www.government.nl/binaries/government/documenten/publications/2023/02/16/ream-2023-endorsing-countries/Endorsing+Countries+and+Territories+REAIM+Call+to+Action.pdf>>

(11) “Blueprint for Action,” 2024.9.11. REAIM Summit 2024 Website <<https://ream2024.kr/home/reameng/board/bbsDetail.do?shareBbsNo=163&shareBbsMngNo=10264&shareMenuId=11613&shareTabMenuId=>>>

(12) 民用物(civilian objects)とは、議定書 I の第 52 条 2 節に定める物的軍事目標以外の全ての物をいう(同第 52 条 1 節)。物的軍事目標について、第 52 条 2 節は、「軍事目標は、物については、その性質、位置、用途又は使用が軍事活動に効果的に資する物であってその全面的又は部分的な破壊、奪取又は無効化がその時点における状況において明確な軍事的利益をもたらすものに限る。」と規定している。

いる（本文 4 項）。

二つ目の「軍事領域における責任ある AI の導入」の観点からは、AI アプリケーションは、国連憲章、国際人道法、国際人権法などの国際法と、必要に応じてその他の関連する法的枠組みに従って開発され、配備され、使用される必要があることを確認している（本文 7 項）。また、責任ある AI 利用（responsible AI applications）の確保に向けて、国家戦略、原則、基準、規範、法律等を適切に確立することの重要性も強調された（本文 8 項）。責任ある AI の確保に向けて認識すべき事項としては、AI 利用の責任と説明責任は機械に移譲できず、人間が負い続けるべきこと、AI アプリケーションの信頼性を確保するためにデータ、アルゴリズム及びその他のバイアスによる誤動作や意図しない結果のリスクを軽減するための適切なセーフガードを確立すべきこと、武力行使に関する人間の判断や制御に関連する措置を含め、AI の開発、配備、使用において人間の適切な関与を維持する必要があること等が示されている（本文 9 項（c）、（d）、（e））。また、効果的な法的審査の手続、信頼醸成措置及び適切なリスク軽減措置の構築と、グッド・プラクティス（良い実践）や教訓に関する各国間の情報交換・協議をそれぞれ推奨することも謳われている（本文 11 項）。

最後の「軍事領域における AI の将来のガバナンス構想」の観点からは、AI 利用においてデータが果たす極めて重要な役割から、適切なデータガバナンスメカニズムについて更なる議論を行う必要があること（本文 17 項）、技術の急速な発展と進歩に対応するために、柔軟でバランスのとれた現実的なガバナンスのアプローチが必要なこと（本文 18 項）などが記されている。

第 3 回 REAIM サミットは、2025 年 9 月にスペインにおいて開催される予定である。

（2）米国主導の政治宣言

米国は、2023 年 2 月の第 1 回 REAIM サミットの中で、別途、軍事分野における AI の責任ある開発、配備及び使用に関する国際的なコンセンサスを構築し、各国が実施すべき措置の在り方を示すことを目的とした「AI と自律性の責任ある軍事利用に関する政治宣言」（Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy）の枠組みを設置することを発表し、同年 11 月に同政治宣言を正式に公表した（2024 年 11 月時点で、58 か国・地域が賛同）⁽¹³⁾。同宣言では、軍事 AI 能力（military AI capabilities）に関して、国際法の遵守、人間が説明責任を果たすこと、その責任ある開発、配備、使用を確保するためにライフサイクル全体を通じて適切な措置を実施すべきことが確認され、国家が実施すべき A から J までの 10 項目の措置が示されている。例えば、AI が国際法、特に国際人道法の下での各義務に合致して利用されることを確保するため、法的審査などの適切な措置を実施すること（B 項）、AI（AI を用いた兵器システムを含む。）の開発、配備、使用において、関係者が適切な注意を払うことを確保すること（E 項）、AI を利用し、又は利用を承認する要員が、文脈に応じて適切な判断を下し、自動化バイアス（Ⅲ章 2 節（2）参照）のリスクを軽減するため、システムの能力

(13) “Building Consensus on the U.S. Framework for a Political Declaration on the Responsible Military Use of Artificial Intelligence and Autonomy,” February 16, 2023. U.S. Department of State (2021-2025 Archived Content) Website <<https://2021-2025.state.gov/building-consensus-on-the-u-s-framework-for-a-political-declaration-on-the-responsible-military-use-of-artificial-intelligence-and-autonomy/>>; “Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy,” November 9, 2023. U.S. Department of State Website <<https://www.state.gov/political-declaration-on-responsible-military-use-of-artificial-intelligence-and-autonomy-2/>>; “Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy, Endorsing States.” *ibid.* <<https://www.state.gov/bureau-of-arms-control-deterrence-and-stability/political-declaration-on-responsible-military-use-of-artificial-intelligence-and-autonomy>>

と限界を十分に理解するよう訓練されていること（G 項）が挙げられている。さらに、AI の安全性、セキュリティ、有効性が、明確に定義された用途の範囲内で及びライフサイクル全体にわたって、適切かつ厳格な試験と保証の対象となることを確保すること、また、自己学習又は継続的な更新が行われる AI に関して、モニタリングなどを通じて重要な安全機能が低下していないことを確保すること（I 項）、AI による失敗のリスクを軽減するため、適切なセーフガード（予期しない結果の検知・回避やシステムの停止・非活性化のための能力）を導入すること（J 項）なども盛り込まれている。

2024 年 3 月には、同政治宣言を支持する国々による初の全体会合が開催され、各国が実施すべき具体的な措置を検討するための三つの作業部会の設置が確認された⁽¹⁴⁾。

(3) 国連での取組

軍事分野全般にわたる AI 利用の規制に向けた取組の場は、有志国によるフォーラムから、より普遍性の高い国連総会へと広がりつつある。REAIM を主導してきたオランダと韓国のほか米国を含む 22 の有志国は、2024 年 10 月、国連総会第一委員会に「軍事分野における AI 及びその国際の平和と安全への影響」と題する決議案を提出した⁽¹⁵⁾。同決議案は、同年 11 月に同委員会で採択され⁽¹⁶⁾、続いて、同年 12 月の国連総会で総会決議として採択された（賛成 159、反対 2、棄権 5）⁽¹⁷⁾。軍事分野全般にわたる AI 利用に関する国連総会決議が採択されたのは、これが初めてとなる。

同決議では、まず、その前文において、国際法を遵守して利用され、人間中心で、説明可能で、安全で、安定し、信頼可能なことを含む、AI の責任ある利用（responsible application）を確保することの重要性（前文 2 項）、決議が AI のライフサイクル全体（プレ設計、設計、開発、評価、試験、配備、使用、販売、調達、運用、廃棄の各段階を含む。）に焦点を当てていること（前文 3 項）が確認されている。また、AI が武力紛争における文民及び民用物の保護などの点を含めて軍事分野において潜在的な機会と利益をもたらす一方で、人道、法、安全保障、技術、倫理の観点から課題と懸念をもたらすこと（前文 7 項、8 項）も記されている。本文には、9 の事項が掲げられており、まず、国連憲章、国際人道法及び国際人権法を含む国際法が、軍事分野における AI のライフサイクルの全ての段階を通して発生する事項に適用されることを

(14) “Inaugural Plenary Meeting for the Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy,” March 29, 2024. U.S. Department of State Website <<https://2021-2025.state.gov/inaugural-plenary-meeting-for-the-political-declaration-on-responsible-military-use-of-artificial-intelligence-and-autonomy/>> 三つの作業部会のテーマは、それぞれ、AI の信頼性保証（AI assurance）、説明責任（accountability）、監督（oversight）である。Jules Palayer and Laura Bruun, “Artificial Intelligence and International Peace and Security,” Stockholm International Peace Research Institute (SIPRI), *SIPRI Yearbook 2025: Armaments, Disarmament and International Security*, Oxford: Oxford University Press, 2025, pp.338-339.

(15) “Artificial Intelligence in the Military Domain and Its Implications for International Peace and Security,” UN Doc. A/C.1/79/L.43, 16 October 2024. <<https://docs.un.org/en/A/C.1/79/L.43>> なお、第一委員会は、国連総会に置かれた六つの主要委員会のうちの一つで、軍縮と国際安全保障に関する問題を取り扱っている。

(16) “Seventy-ninth Session, Agenda Item 98: General and Complete Disarmament, Report of the First Committee,” UN Doc. A/79/408, 15 November 2024, pp.53-54. <<https://docs.un.org/en/A/79/408>>

(17) “Resolution Adopted by the General Assembly on 24 December 2024,” UN Doc. A/RES/79/239, 31 December 2024. <<https://docs.un.org/en/a/res/79/239>>; “United Nations General Assembly, Seventy-ninth Session: 55th Plenary Meeting: Official Records,” UN Doc. A/79/PV.55 (Resumption 1), 24 December 2024, p.19. <[https://docs.un.org/en/a/79/pv.55\(Resumption1\)](https://docs.un.org/en/a/79/pv.55(Resumption1))> 反対国は北朝鮮及びロシア、棄権国はベラルーシ、エチオピア、イラン、ニカラグア及びサウジアラビア。

確認している（本文 1 項）。各国に対しては、AI の利用が国際の平和と安全に与える影響の評価を継続することを促し（本文 3 項）、国家、学界、市民社会、国際・地域機関、民間部門と協力し、国連内を含めた、AI の責任ある利用に関する意見交換を奨励している（本文 5 項）。また、国連事務総長に対して、AI の利用が国際の平和と安全にもたらす機会と課題について、特に LAWS 以外の分野に焦点を当て、加盟国及びオブザーバー国に見解を求め、各国による更なる議論のために、それらの見解と既存の及び新たな規範に関する提案（normative proposals）の一覧をまとめた報告書を第 80 会期（2025 年 9 月開幕）の国連総会に提出するよう要請している（本文 7 項）。

2025 年 8 月には、決議本文 7 項の要請に基づく、国連事務総長による報告書（同年 6 月 5 日付け）が公表されている⁽¹⁸⁾。

2 規制の方向性と展望

本章 1 節のとおり、軍事分野における AI 利用の規制の方向性に関しては、国際法の適用、人間の適切な関与を維持する必要性、人間が説明責任を負うべきこと、データのバイアス等による影響や意図しない結果のリスクに対応する必要性、各ステークホルダーが協力する必要性など、大きな原則のレベルでは、多国間で見解が一定程度集約されつつある。同時に、こうした原則に沿った上で、具体的にどのような段階でどのような対応措置が必要であるかは、依然模索されている状況となっている。

規制のレベルに関しては、これまでのところ、法的拘束力を持つ条約等によるハードな規制ではなく、多国間での行動規範や政治文書などを通じたソフトな規制が進んでいる⁽¹⁹⁾。AI が利用方法によってはメリットをもたらすこと、また、軍事分野に限定した規制の実現性の乏しさが認識されていることなどから、現時点で、軍事分野における AI 利用を全面的又は広範に禁じるような形でのハードな規制が目指されているわけではない⁽²⁰⁾。なお、将来には、各国、

(18) 同報告書は、AI のもたらす機会と課題、既存の及び新たな規範の提案、今後の議論の在り方等の主要論点に関する加盟国等の意見を列挙した上で、各国に対し、同報告書の内容を検討し、軍事領域における AI と国際の平和及び安全への影響の論点に取り組むプロセスを確立する措置を講じるよう勧告すること等を含む、事務総長の見解と結論を提示している。“Artificial Intelligence in the Military Domain and Its Implications for International Peace and Security: Report of the Secretary-General,” UN Doc. A/80/78, 5 June 2025. <<https://docs.un.org/en/A/80/78>>; 「「軍事領域における AI と国際の平和及び安全への影響」に関する国連事務総長報告書」2025.8.20. 外務省ウェブサイト <https://www.mofa.go.jp/mofaj/gaiko/arms/pagew_000001_01890.html>

(19) 岩本誠吾「AI 民生利用、AI 軍事利用および AI 自律兵器における法規制の関係性—2024 年度—」『京都産業大学世界問題研究所紀要』40 号, 2025.3, p.15. 今後におけるソフトな規制の推進を提案しているものとして、例えば、軍事分野における AI 利用に伴う課題に対応するための政策のロードマップを提案した、国連軍縮研究所（United Nations Institute for Disarmament Research: UNIDIR）による 2025 年 7 月の報告書がある。同報告書は、多国間レベルでの取組として、まずは、国連の傘下に設置する多国間による議論のプラットフォームで軍事分野における責任ある AI に関する包括的な基本原則を策定した上で、将来、この基本原則を、国際的な自主規範又は指針に発展させることを提案している。それらは、行動規範や政治宣言の形をとる可能性があり、必要に応じて、より技術的手段（例えば、AI の信頼性保証、試験と評価のための強固なプロトコルなど）によって補完され得るとしている。UNIDIR’s Security and Technology Programme, *Artificial Intelligence in the Military Domain and Its Implications for International Peace and Security: An Evidence-Based Road Map for Future Policy Action*, Geneva: UNIDIR, 2025, pp.7, 24-25. <https://unidir.org/wp-content/uploads/2025/07/UNIDIR_AI_military_domain_implications_international_peace_security.pdf> なお、UNIDIR は、国連の枠内にありながら、軍縮と国際安全保障の問題に関して独立した研究を行う機関である。

(20) 軍事分野における AI 利用を一律に禁じることの難しさに言及した資料として、例えば、次を参照。Jürgen Altmann, “AI and the Future of Arms Control,” Robin Geiß and Henning Lahmann, eds., *Research Handbook on Warfare and Artificial Intelligence*, Cheltenham: Edward Elgar Publishing, 2024, p.347; Paul Scharre, *Four Battlegrounds: Power in the Age of Artificial Intelligence*, New York: W. W. Norton, 2023, pp.289-290; 佐藤 前掲注(7), pp.76-77.

地域、多国間のフォーラム、国連などの様々な枠組みで、必要に応じて、AI の特定の用途や利用方法に対する法的拘束力を持つ規制が検討されていく可能性も考えられる⁽²¹⁾。

Ⅱ AI を用いた意思決定支援システム (AI-DSS) の概要及び利用事例

1 意思決定支援システムとは

意思決定支援システム (Decision Support Systems: DSS) は、人間の意思決定を支援するために設計されたコンピュータ化されたツールを意味する。通常、関連する情報を表示し、合成し、分析し、又は目標を達成するための選択肢を提示することによって、人間の意思決定を支援する⁽²²⁾。軍事分野においては、DSS は、軍の意思決定者による効果的かつ迅速で合法的な反応や行動につながる意思決定を可能とすることを目的に利用される⁽²³⁾。なお、軍の意思決定は、平時における人員や物資の管理から、装備の維持整備や前方輸送などの後方支援、そして武力行使に至るまで、様々な局面で行われることが想定されるが、本稿では、特に国際人道法に違反するリスクを伴い得るものとして、武力行使に関する意思決定に焦点を当てる。

武力行使に関する意思決定に AI-DSS が用いられるようになった背景には、2000 年代半ば以降の機械学習技術やアルゴリズムの実装の進歩に加え、それを支えるコンピューティング性能の向上及びデジタルデータの爆発的増加といった要因によって、AI を用いた自律化技術の実用化が急速に進んだことがある。とりわけ、機械学習の一種である深層学習の実装の進歩によって、高度な画像認識、音声認識、自然言語処理が可能となった⁽²⁴⁾。このような中、AI-DSS においても、人物や物体を特定したり、人物の行動パターンを分析したり、過去のデータの分析を通じて対応の選択肢を導き出したりするなど、意思決定に係る様々なタスクの実行が一定程度可能となりつつある。

2 意思決定のプロセスと AI-DSS による支援

武力行使に関する意思決定プロセスは、上位の政治レベルから、戦略、作戦、戦術までの各レベルに分けて考えられることが多い⁽²⁵⁾。通常、一つの武力行使のプロセスは、これらの各

(21) 民生分野における AI に関しては、ソフトな規制とともに、次の段階として、2024 年 5 月に成立（同年 8 月に施行）した EU の AI 規制法（Artificial Intelligence Act）のように、ハードな規制が行われる動きもみられる。こうした動きも踏まえ、京都産業大学法学部教授の岩本誠吾氏は、「元来、軍事領域での AI 利用は、法文書の作成を直接的な目標とするものではなく、指針原則や行動準則（ソフト・ロー）の明確化が求められている。その後で、それを法文書化する必要があるか否かが検討されるだろう」としている。岩本 前掲注(19) また、UNIDIR は、2025 年 7 月の報告書で、軍事分野における AI 利用に関する国際的な基本原則と将来における自主規範や指針を策定するという同機関の提案に関して、LAWS における AI といった特定の AI の利用をより制限的に規制する法的拘束力を持つ文書を国家が交渉する可能性を排除することを意図したものではないと説明している。UNIDIR's Security and Technology Programme, *op.cit.*(19), p.25.

(22) ICRC and Geneva Academy, *Expert Consultation Report: Artificial Intelligence and Related Technologies in Military Decision-making on the Use of Force in Armed Conflicts*, Geneva: ICRC, March 2024, p.25. <<https://shop.icrc.org/download/ebook?sku=4758/002-ebook>>

(23) *ibid.*, p.8.

(24) James Johnson, "Artificial Intelligence & Future Warfare: Implications for International Security," *Defense & Security Analysis*, vol.35 no.2, 2019, p.147. <<https://doi.org/10.1080/14751798.2019.1600800>>; UNIDIR, "The Weaponization of Increasingly Autonomous Technologies: Artificial Intelligence, a Primer for CCW Delegates," *UNIDIR Resources*, no.8, 2018, pp.2-3. <<https://unidir.org/wp-content/uploads/2023/05/the-weaponization-of-increasingly-autonomous-technologies-artificial-intelligence-en-700.pdf>>

(25) 本段落では、武力行使に関する典型的な意思決定のプロセスを説明した例として、次の ICRC 刊行の報告書及び UNIDIR 刊行の資料における説明を参照した。ICRC and Geneva Academy, *op.cit.*(22), p.7; Merel Ekelhof and

レベルの意思決定が連なった結果であり、プロセスを通じて「様々なタイプのアクターが関与する」ことになる⁽²⁶⁾。典型的なプロセスとして、政治レベルでは、政治指導者は、軍事行動の必要性を決定し、他の全てのレベルの意思決定に適用される目的や指針を設定する。戦略レベルでは、政治的目的等は、武力行使の条件や程度といった重要な決定を含む、より具体的な軍事的目的と指針に変換される。作戦レベルでは、軍において指揮を行う機構は、設定された目的等に従い、標的の分析、選択、優先順位付けを含む、作戦や任務を実行するための条件やルールを決定する。それによって、戦術レベルでは、部隊が具体的な標的の属性、兵器の使用と影響等を評価し、攻撃を実行する⁽²⁷⁾。専門的な助言者（指揮官に法的助言を行う法律顧問など）は、全ての決定が政治的な指針、法的義務、その他のルールに従って行われるよう、あらゆるレベルの意思決定プロセスに関与できる⁽²⁸⁾。

AI-DSS は、こうした武力行使に関する意思決定の連鎖の中で、戦略から戦術までのいずれかの段階において、人間の意思決定者を支援するために用いることが想定されている。AI-DSS の規制に関する検討や提案を行っている赤十字国際委員会（ICRC）は、現在開発・導入が進められている AI-DSS が担うタスクについて、幾つかの具体例を挙げている。これによれば、AI-DSS は、戦略レベルでは、他国の行動分析や、戦争ゲームのようなシミュレーション、作戦レベルでは、潜在的な標的の特定、選択、優先順位付けや、巻き添え被害の推定、さらに、戦術レベルでは、リアルタイムで戦場のインテリジェンス情報を処理して、特定の標的に最適な兵器を推奨するといったタスクを担っている⁽²⁹⁾。

AI-DSS は、人間の意思決定者の関与を前提としたシステムである。AI システムが従来人間に委ねられてきた意思決定に係るタスクの一部を担うとしても、人間が武力行使に関する最終的な責任を持つことが想定されている⁽³⁰⁾。一方、後述するとおり（Ⅲ章 2 節（2）参照）、意思決定者が無批判に AI-DSS の推奨情報に従うなど、標的の選定や攻撃の決定が実質的な人間の関与なしに実行されるような場合には、法的、倫理的な問題を生じる可能性が懸念されている⁽³¹⁾。そのため、AI-DSS の利用に当たっては、「必要な人間の判断をどのように維持するのか」が問われることとなり、人間の意思決定者による適切な関与を維持することが極めて重要となる⁽³²⁾。

3 AI-DSS の利用事例

本節では、近年報告されている、武力行使に関する意思決定のための AI-DSS の代表的な利用事例を確認する。なお、利用国や開発企業は、AI が具体的にどのようなもので、どのようなタスクで用いられているのかなど、技術や運用に関する詳細を明らかにしていない場合が多く、本節における記述の大部分は、メディアやジャーナリストの調査記事といった情報ソースに基づく。

Giacomo Persi Paoli, “The Human Element in Decisions about the Use of Force,” Geneva: UNIDIR, 2020. <https://unidir.org/wp-content/uploads/2023/05/UNIDIR_Iceberg_SinglePages_web.pdf>

⁽²⁶⁾ Ekelhof and Paoli, *ibid.*, p.2.

⁽²⁷⁾ ICRC and Geneva Academy, *op.cit.*(22); *ibid.*

⁽²⁸⁾ Ekelhof and Paoli, *ibid.* 議定書 I の第 82 条は、「締約国はいつでも、また、紛争当事者は武力紛争の際に、諸条約及びこの議定書の適用並びにその適用について軍隊に与えられる適当な指示に関して軍隊の適当な地位の指揮官に助言する法律顧問（legal advisers）を必要な場合に利用することができるようにする。」と規定している。

⁽²⁹⁾ ICRC and Geneva Academy, *op.cit.*(22), pp.13-14.

⁽³⁰⁾ *ibid.*, p.9; Matthias Klaus, “Transcending Weapon Systems: The Ethical Challenges of AI in Military Decision Support Systems,” September 24, 2024. Humanitarian Law & Policy Blog <<https://blogs.icrc.org/law-and-policy/2024/09/24/transcending-weapon-systems-the-ethical-challenges-of-ai-in-military-decision-support-systems/>>

⁽³¹⁾ ICRC and Geneva Academy, *ibid.*; Klaus, *ibid.*

⁽³²⁾ ICRC and Geneva Academy, *ibid.*, pp.9, 21.

(1) プロジェクト・メイブン（米国）

米国では、軍と民間企業の双方において、AI や関連する新興技術の軍事分野における利用が進められている。プロジェクト・メイブン（Project Maven）は、米国防省が主導する、コンピューター・ビジョンと機械学習ベースのアルゴリズムを用いて、ドローンが事前に収集した映像データ等に基づいてリアルタイムで標的を特定することを目的としたプロジェクトである。2017 年 4 月に設立された後、米国内でプロジェクトに対する抗議行動⁽³³⁾に直面しつつも、迅速にシステムの開発が進められ、同年末までに運用が開始された⁽³⁴⁾。当初のプロジェクトの目的は、中東においてドローンで収集した膨大な映像データの処理であり、それまでは、敵の軍事資産を特定するため、情報分析チームが多く時間を費やして人手で映像データを精査していたとされる⁽³⁵⁾。ブルームバーグニュースの記事によれば、今では、ドローンの映像のみならず、衛星、レーダー、赤外線センサーの収集データや通信傍受による位置情報データなど、数多くのデータセットからの情報が、機械学習を通して分析処理される⁽³⁶⁾。そして、処理された情報は、メイブン・スマート・システムという単一のインターフェース上に統合した形で表示され、「司令官が戦場全体を一目で見ることができる」とされる⁽³⁷⁾。

メイブンのシステムは、運用開始後も配備環境に応じて改良が重ねられ、実際に米軍の作戦の一環として、イエメン内のロケットランチャーや紅海における水上艦艇の位置の特定、イラクとシリアでの標的の絞り込みなどに役立てられたとされる⁽³⁸⁾。米中央軍⁽³⁹⁾の最高技術責任者であるシュイラー・ムーア（Schuyler Moore）氏は、2024 年 2 月のイラクとシリアの軍事目標に対する 85 回以上の空爆⁽⁴⁰⁾の際にメイブンを利用したことをメディアに対して明らかにした⁽⁴¹⁾。

(33) プロジェクト・メイブンに当初参加していたグーグル（Google）は、社内外からの「戦争用の技術を開発すべきでない」などとする抗議行動を受け、2018 年 6 月、翌年に期限を迎える国防省との契約更新を申請しない方針を表明し、プロジェクトから撤退した。“Google Vows Its A.I. Will Not Be Used for Weapons,” *New York Times*, June 8, 2018. グーグルのプロジェクト撤退後におけるその他の企業や技術者の動きについて、次を参照。葛岡成樹「人工知能を用いた新しい GEOINT プロジェクト Project Maven — 偵察画像からの目標物自動抽出 —」『防衛技術ジャーナル』516 号, 2024.3, pp.19-25. <<https://www.jiss.or.jp/wp-content/uploads/2024/03/journal20240318.pdf>> グーグルのその後の対応については、次も参照。“Google Drops AI Pledge Related to Weapons, Surveillance,” *Washington Post*, February 6, 2025.

(34) Scharre, *op.cit.*(20), p.58; Lauren A. Kahn, “Risky Incrementalism: Defense AI in the United States,” Heiko Borchert et al., eds., *The Very Long Game: 25 Case Studies on the Global State of Defense AI*, Cham: Springer, 2024, pp.46-47. なお、2022 年以降、プロジェクト・メイブンの地理空間情報に係るサービスと能力の運用管理権は、国防次官室から、国防省の下で地理空間情報に基づくインテリジェンスを担当する国家地理空間情報局（National Geospatial-intelligence Agency: NGA）に移管されている。Theresa Hitchens, “Pentagon’s Flagship AI Effort, Project Maven, Moves to NGA,” April 27, 2022. *Breaking Defense Website* <<https://breakingdefense.com/2022/04/pentagons-flagship-ai-effort-project-maven-moves-to-nga/>>

(35) Scharre, *ibid.*, p.55; Kahn, *ibid.*, p.46; Gregory C. Allen, “Project Maven Brings AI to the Fight against ISIS,” December 21, 2017. *Bulletin of the Atomic Scientists Website* <<https://thebulletin.org/2017/12/project-maven-brings-ai-to-the-fight-against-isis/>>

(36) Katrina Manson, “AI warfare Is Already Here,” *Bloomberg News*, February 28, 2024.

(37) *ibid.*

(38) Scharre, *op.cit.*(20), p.58; *ibid.*

(39) 米軍は、地域別及び機能別に編成された、軍種横断的な 11 の統合軍を置いている（2025 年 8 月現在）。米中央軍（U.S. Central Command）は、地域別統合軍の一つで、中東と中央アジアを中心に、北東アフリカと南アジアのそれぞれ一部にまたがる地域を管轄している。“Area of Responsibility,” U.S. Central Command Website <<https://www.centcom.mil/AREA-OF-RESPONSIBILITY/>>

(40) 米軍は、2024 年 1 月 28 日にヨルダン北東部の米軍基地において親イラン武装勢力からのドローン攻撃を受けて米兵 3 名が死亡したことへの対応として、同年 2 月 2 日から翌日未明（現地時間）にかけて、イラクとシリアにおける同勢力の 7 施設、85 以上の標的を対象とした一連の空爆を行った。Joseph Clark, “U.S. Strikes Targets in Iraq and Syria in Response to Deadly Drone Attack,” Feb. 2, 2024. U.S. Department of Defense Website <<https://www.defense.gov/News/News-Stories/Article/article/3665734/>>

(41) Katrina Manson, “US Used AI to Help Find Middle East Target for Airstrikes,” *Bloomberg News*, February 26, 2024.

その際、メイブンの出力結果は潜在的な標的を見つけるために利用され、空爆の決定を下すために利用されたのではないこと、メイブンからの推奨情報は人間によってダブルチェックされたことを強調している。

なお、メイブンのアルゴリズムは中東の砂漠の多い環境に合わせて改良が図られてきており、別の環境では認識の精度が劣る可能性があることが指摘されている。例えば、ブルームバーグニュースの記事によれば、総合的にみて、米陸軍総軍の第 18 空挺軍団の情報分析官が物体を正確に識別できる確率が 84% であるのに対し、メイブンが物体を正確に識別できる確率は 60% 程度であり、雪などの条件で画像の解析が困難な状況下では 30% を下回ることもあったという⁽⁴²⁾。

(2) ゴスペル、ラベンダー（イスラエル）

イスラエルでは、防空システムのアイアンドーム（Iron Dome）を始め、AI 技術を統合した様々な兵器システムの開発が進められている⁽⁴³⁾。イスラエル国防軍の幹部は、2021 年 5 月におけるパレスチナのガザ地区での 11 日間にわたる軍事作戦（壁の守護者作戦（Operation Guardian of the Walls））⁽⁴⁴⁾において、初めて作戦全般で AI を用いたシステムを取り入れたことを、メディアに対して明らかにした⁽⁴⁵⁾。

イスラエル国防軍は、ガザ地区の武装勢力に関する全てのデータを一つのシステムに集約し、情報の分析と抽出を可能にする高度な AI 技術プラットフォームを確立する目的で、複数の AI-DSS を開発してきた⁽⁴⁶⁾。軍諜報部門の 8200 部隊（Unit 8200）が開発したゴスペル（Gospel）（ヘブライ語で、ハブソラ（Habsora））は、AI を用いて信号、視覚、地理空間などの複数のインテリジェンスデータからの膨大な情報を分析し、攻撃の標的となり得る建物や構造物を特定して、攻撃対象となる標的を推奨するシステムである⁽⁴⁷⁾。8200 部隊においてデータサイエンス及び AI の責任者を務める指揮官は、2021 年 5 月の壁の守護者作戦で AI システムが 200 の潜在的な標的の情報を生成したと述べている⁽⁴⁸⁾。イスラエルの国際法分野の研究者は、これは従来であれば「複数の情報分析官が数週間かけて行う必要があった作業」であり、ゴスペルはそれを「数秒以内に実行した」としている⁽⁴⁹⁾。効率性の向上の一方で、その利用方法には、懸念の声も上がっている。イスラエルのネットメディアである +972 マガジンの報告によれば、

<<https://www.bloomberg.com/news/articles/2024-02-26/us-says-it-used-ai-to-help-find-targets-it-hit-in-iraq-syria-and-yemen>>

(42) Manson, *op.cit.*(36)

(43) Paul D. Williams, “Emerging Technologies,” Paul D. Williams and Matt McDonald, eds., *Security Studies: An Introduction*, Fourth edition, New York: Routledge, 2023, p.640; Inbar Dolinko and Liran Antebi, “Embracing the Organized Mess: Defense AI in Israel,” Borchert et al., eds., *op.cit.*(34), pp.397-398.

(44) イスラエル国防軍は、2021 年 5 月 10 日にハマスなどの組織の武装勢力がガザ地区からイスラエルに向けて数千発のロケット弾を発射したことを受けて、同日以降 20 日までの 11 日間にわたって、ガザ地区におけるこれら組織の軍事拠点等を攻撃する軍事作戦を行った。“Operation Guardian of the Walls,” 2021.5.21. Gov.il Website <<https://www.gov.il/en/pages/operation-guardian-of-the-walls-10-may-2021>>

(45) 「イスラエル軍「初の AI 戦」ガザ紛争 迎撃目標設定に活用」『日本経済新聞』2021.6.27.

(46) Anna Ahronheim, “Israel’s Operation against Hamas Was the World’s First AI War,” *Jerusalem Post*, May 27, 2021. <<https://www.jpost.com/arab-israeli-conflict/gaza-news/guardian-of-the-walls-the-first-ai-war-669371>>

(47) Williams, *op.cit.*(43)

(48) Yonah Jeremy Bob, “IDF Unit 8200 Commander Reveals Cyber Use to Target Hamas Commander,” *Jerusalem Post*, February 13, 2023. <<https://www.jpost.com/israel-news/article-731443>>

(49) Tal Mimran and Gal Dahan, “Artificial Intelligence in the Battlefield: A Perspective from Israel,” 2024.4.20. *Opinio Juris* Website <<https://opiniojuris.org/2024/04/20/artificial-intelligence-in-the-battlefield-a-perspective-from-israel/>>

兵士はゴスペルから潜在的な標的の情報を受け取り、それをチェックリストに従って検討することになる。その際、兵士は「素早く作業し、標的について深く掘り下げる時間はない」という⁽⁵⁰⁾。

また、イスラエル国防軍はラベンダー (Lavender) と呼ばれる AI-DSS を配備しており、特に 2023 年から 2024 年にかけてのパレスチナ武装勢力との紛争⁽⁵¹⁾の初期段階で、重要な役割を果たしたとされる⁽⁵²⁾。ゴスペルが建物等の物体を対象とするのに対し、ラベンダーは、ハマスなどの組織の軍事部門に所属する人物を潜在的な標的としてマークするように設計されている。ラベンダーの利用方法に関しても懸念が指摘されており、+972 マガジンの報告によれば、ラベンダーの出力結果に対して人間が行う検証は、「AI が選択した標的が女性ではなく男性であることを確認すること」に限定され、攻撃前の標的一人当たりの検証にかかる平均時間は 20 秒程度であったという。また、システムは約 10% のケースで誤りを犯し、武装勢力とわずかしかつながりのない又は全くつながりのない人物を標的として推奨することもあったとされ、このような誤りがあることを認識した上でラベンダーによる標的リストの全面的な利用が承認されたという⁽⁵³⁾。

イスラエル国防軍は、2024 年 6 月に声明を発表し、ゴスペルとラベンダーについてメディア等で指摘された懸念に対して反論する形で、両システムは「情報分析官が既存の情報源を包括的かつ効果的に相互参照するためのツール」に過ぎず、標的の特定は国際法の規則に従って行われ、「常に、訓練され権限を持つ人間の情報分析官が行う」と主張している⁽⁵⁴⁾。

(3) GIS Arta ほか (ウクライナ)

ウクライナの防衛産業においては、2022 年 2 月のロシアによる侵攻を受けて、それまで散発的にしか利用されていなかった AI の利用が大幅に進んだとされる⁽⁵⁵⁾。AI-DSS に関しても、ウクライナのデジタル変革省が主導する防衛技術産業の支援のためのイニシアチブであるブレイブ・ワン (Brave1)⁽⁵⁶⁾の一環として、多くのシステムが開発されている⁽⁵⁷⁾。GIS Arta (Geographic Information System of Artillery) はその一つで、ウクライナの開発者グループによって開発された、砲撃のための指揮統制システムである⁽⁵⁸⁾。衛星、偵察用ドローン、スマートフォン、GPS、レーダーなどから収集したロシア軍の標的の位置情報をリアルタイムで共有し、アルゴリズムに従って、標的を攻撃するのに最も適した砲兵部隊とペアリングする。これによって、タイムズの記事によれば、標的を発見してからその標的が砲火を浴びるまでの時間は、従来の

(50) Yuval Abraham, “‘A Mass Assassination Factory’: Inside Israel’s Calculated Bombing of Gaza,” November 30, 2023. +972 Magazine Website <<https://www.972mag.com/mass-assassination-factory-israel-calculated-bombing-gaza/>>

(51) 2023 年 10 月 7 日にハマス等のパレスチナ武装勢力がイスラエル南部への大規模な奇襲攻撃を行い、これを発端としてイスラエル国防軍はガザ地区に対する空爆及び地上作戦を開始した。以降、2025 年 8 月現在まで両者間における武力衝突が続いている。

(52) Yuval Abraham, “Lavender: The AI Machine Directing Israel’s Bombing Spree in Gaza,” April 3, 2024. +972 Magazine Website <<https://www.972mag.com/lavender-ai-israeli-army-gaza/>>

(53) *ibid.*

(54) Israel Defense Forces, “The IDF’s Use of Data Technologies in Intelligence Processing,” June 18, 2024. <<https://www.idf.il/210062>>

(55) Vitaliy Goncharuk, “Survival of the Smartest? Defense AI in Ukraine,” Borchert et al., eds., *op.cit.*(34), p.375.

(56) ブレイブ・ワンは、ウクライナのデジタル変革省が主導する、政府機関の統合調整プラットフォームである。ウクライナの防衛技術プロジェクトに組織、情報及び財政上の支援を提供し、防衛技術産業の全ての利害関係者の協力を促進するために創設された。“Ukrainian Defense Innovations: United Coordinational Platform for Defense Tech Powered by the Government: Brave1.” Brave1 Website <<https://brave1.gov.ua/en/>>

(57) Goncharuk, *op.cit.*(55), p.382.

(58) Charlie Parker, “Uber-style Technology Helped Ukraine to Destroy Russian Battalion,” *Times (Online)*, May 14, 2022.

20 分から 1 分程度に短縮されたという⁽⁵⁹⁾。

このほか、ウクライナ政府は、標的選定、地雷除去、ロシアの戦争犯罪の証拠収集など、様々な目的で、米国の AI 開発企業（パランティア・テクノロジーズ（Palantir Technologies））が提供する AI を用いたソフトウェアを利用していることが報じられている⁽⁶⁰⁾。こうした中、ウクライナでの戦争が、AI を用いた軍事用ソフトウェアの「戦場での性能等を評価するのに適したテストベッド」として利用されているとも指摘されている⁽⁶¹⁾。

Ⅲ AI-DSS の利用に伴うメリットとリスク

1 メリット

武力行使に関する意思決定のために AI-DSS を利用するメリットとしては、大まかに、意思決定のスピードと質の向上があると考えられている⁽⁶²⁾。意思決定のスピードの向上に関しては、AI による大量の情報の迅速な分析処理によって、情報収集から標的を発見・特定し攻撃するまでの意思決定プロセスにかかる時間が短縮されるため、作戦をより迅速に実行できるようになる⁽⁶³⁾。また、意思決定の質の向上に関しては、複数のデータセットの一元的な分析を通して、敵の行動を予測したり自らの潜在的な行動の結果を評価したりするための情報など、指揮官が「より良い情報に基づいた決断」をすることが可能となる⁽⁶⁴⁾。さらに、戦場の状況認識が向上することで、部隊は、状況の変化に素早く対応しながら、「より複雑な作戦を実行できる」ようになるとされる⁽⁶⁵⁾。

軍事分野における意思決定のスピードの向上には、理論上、幾つかの基本的な効果も伴い得ることが指摘されている。まず、標的の発見・特定から攻撃までの時間の短縮は、通常、敵を打倒する上での優位性の確保につながる⁽⁶⁶⁾。作戦全体では、軍が特定し、攻撃できる標的の数を増やすことも可能となる⁽⁶⁷⁾。また、戦場では、標的が移動したり、文民が接近してきたりする可能性が低くなるという点で、時として、「兵力の展開をより精密にすることに資する」可能性がある⁽⁶⁸⁾。

(59) *ibid.* なお、GIS Arta のシステムが統合している具体的な AI 技術の種類などは不明であり、AI システムに該当するかにしては異論もある。例えば、戦略国際問題研究所（Center for Strategic and International Studies: CSIS）ワドワーニ AI センター（Wadhvani AI Center）のディレクターであるグレゴリー・アレン（Gregory Allen）氏は、GIS Arta は「伝統的なソフトウェア」だとしている。Arthur Holland Michel, “Inside the Messy Ethics of Making War with Machines: AI Is Making Its Way into Decision-making in Battle. Who’s to Blame When Something Goes Wrong?” August 16, 2023. MIT Technology Review Website <<https://www.technologyreview.com/2023/08/16/1077386/war-machines/>>

(60) Jeffrey Dastin, “Ukraine Is Using Palantir’s Software for ‘Targeting,’ CEO Says,” *Reuters*, February 2, 2023. <<https://www.reuters.com/technology/ukraine-is-using-palantirs-software-targeting-ceo-says-2023-02-02/>>; Vera Bergengruen, “How Tech Giants Turned Ukraine into an AI War Lab,” *Time*, February 8, 2024. <<https://time.com/6691662/ai-ukraine-war-palantir/>>

(61) Goncharuk, *op.cit.*(55), p.376. 他に、次を参照。Bergengruen, *ibid.*

(62) ICRC and Geneva Academy, *op.cit.*(22), pp.11-12; Tobias Vestner, “From Strategy to Orders: Preparing and Conducting Military Operations with Artificial Intelligence,” Geiß and Lahmann, eds., *op.cit.*(20), p.120; UNIDIR’s Security and Technology Programme, *op.cit.*(19), pp.13-14.

(63) Scharre, *op.cit.*(20), p.265; ICRC and Geneva Academy, *ibid.*

(64) Vincent Boulanin, “Risks and Benefits of AI-enabled Military Decision-making,” Geiß and Lahmann, eds., *op.cit.*(20), p.106.

(65) Scharre, *op.cit.*(20), p.265. 他に、次を参照。*ibid.*, pp.107-108.

(66) Boulanin, *ibid.*, p.101; ICRC and Geneva Academy, *op.cit.*(22), p.12.

(67) Arthur Holland Michel, *Decisions, Decisions, Decisions: Computation and Artificial Intelligence in Military Decision-making*, Geneva: ICRC, April 2024, p.26. <<https://shop.icrc.org/download/ebook?sku=4757/002-ebook>>

(68) Kenneth Anderson and Matthew C. Waxman, “Debating Autonomous Weapon Systems, Their Ethics, and Their Regulation

2 リスク

AI-DSS は、上述のようなメリットを持ち得る一方、AI の技術上の限界に伴うリスク及び人間とシステムとの相互作用に伴うリスクも抱えている。

(1) 技術上の限界に伴うリスク

機械学習ベースの AI システムには、その精度や有効性が、学習データの質と量に依存するという技術上の限界がある⁽⁶⁹⁾。AI は、データ内に存在するバイアス、データ汚染などの影響を受けやすく、また、欺瞞（まん）的情報に対しても脆（ぜい）弱である。そのため、学習に当たっては、可能な限りバイアス等の少ない、膨大なデータセットを使うことが求められている⁽⁷⁰⁾。軍事分野においては、データの利用可能性（data availability）も問題となり得る。スポーツや災害といった民生分野のデータと比べると、武力紛争の意思決定タスクに関わるデータに関しては、質の高い代表的な学習データを得ることがより難しいためである⁽⁷¹⁾。

こうした限界に伴い、AI-DSS の利用において生じ得るリスクとして、意図的でなくとも、例えば、データに性別や民族といった要素に関するバイアスが潜んでいれば、「特定の民族グループに属する人々を標的として特定」したり、「全ての文民男性を戦闘員とみなし」たりする結果を招く可能性があると考えられる⁽⁷²⁾。また、学習データと実際の配備環境との間にデータ分布上のずれがあったり、運用開始後に環境がわずかに変化したりするだけで、システムの性能は低下し、意図しない結果を招く可能性もある⁽⁷³⁾。特に、武力紛争は複雑な敵対的環境下にある。意図的に予測不可能で欺瞞的な行動を取り、常に行動を修正する敵が存在する中で、学習データは「時代遅れ」になりやすいとされる⁽⁷⁴⁾。

さらに、こうしたリスクは、AI-DSS の運用後も引き続き影響を与えると考えられている。従来型の兵器や兵器システムが開発段階で「学習」を停止するのに対し、AI を用いたシステムは、運用後も継続的なデータ学習を通じて性能を向上させることが想定されているためである。継続的な学習によって、実際の配備環境の条件に徐々に適合され、良い性能を発揮する可能性がある一方、新たな環境と相互作用することで、開発者が予見できず、試験や認証も受けていないような創発的な挙動を獲得したり、性能が低下したりする可能性もあるとされる⁽⁷⁵⁾。

(2) 人間とシステムの相互作用に伴うリスク

人間とシステムの相互作用の観点では、現行の多くの AI の作用が「ブラックボックス」状態であり、人間にとって、「システムがどのような出力をするか、なぜ所定の出力をしたのか」

under International Law,” Roger Brownsword et al., eds., *The Oxford Handbook of Law, Regulation and Technology*, Oxford: Oxford University Press, 2017, p.1103. <<https://doi.org/10.1093/oxfordhb/9780199680832.013.33>>

(69) ICRC and Geneva Academy, *op.cit.*(22), p.16; James Johnson, *Artificial Intelligence and the Future of Warfare: The USA, China, and Strategic Stability*, Manchester: Manchester University Press, 2021, pp.23-25.

(70) ICRC and Geneva Academy, *ibid.*; Vestner, *op.cit.*(62), p.121; Boulanin, *op.cit.*(64), pp.103-104.

(71) 例えば、通常は秘匿を前提とする敵の移動式ミサイル発射機に関するデータを体系的に得ることは難しい。ICRC and Geneva Academy, *ibid.*; Vestner, *ibid.*; Boulanin, *ibid.*; Johnson, *op.cit.*(69), p.24.

(72) Jimena Sofia Viveros Álvarez, “The Risks and Inefficacies of AI Systems in Military Targeting Support,” September 4, 2024. Humanitarian Law & Policy Blog <<https://blogs.icrc.org/law-and-policy/2024/09/04/the-risks-and-inefficacies-of-ai-systems-in-military-targeting-support/>>

(73) Scharre, *op.cit.*(20), pp.58, 64, 230, 233; *ibid.*

(74) Michel, *op.cit.*(67), pp.43-44, 49. 他に、次を参照。Paoli et al., *op.cit.*(9), p.12.

(75) ICRC and Geneva Academy, *op.cit.*(22), p.17.

を知ることが難しいという限界がある⁽⁷⁶⁾。この AI における予測可能性と理解可能性の欠如から、人間が AI-DSS による出力の質を適切に評価するのに限界が生じ、意思決定の質に影響を及ぼすリスクが懸念されている⁽⁷⁷⁾。意思決定者の立場からすると、AI-DSS を利用するに当たり、完全には理解できないその出力結果を信頼すべきかを判断することになる。

このリスクは、システムの出力結果に過度に依存する人間の傾向など、人間の認知的なバイアスによって悪化することも指摘されている⁽⁷⁸⁾。システムが導入されて一たび効果を発揮すると、人間の意思決定者はシステムの出力結果を正しいものとして受け入れがちになり（自動化バイアス）、矛盾する情報を無視したり、裏付ける情報を探さなかったりする可能性がある⁽⁷⁹⁾。一方、AI やアルゴリズムへの心理的抵抗感から、AI の出力結果を常に過小評価すること（アルゴリズム嫌悪）、不作為に陥ったり信頼性の低い他の情報に頼ったりすることにつながり、弊害をもたらす可能性がある⁽⁸⁰⁾。

さらに、軍事作戦のテンポが速くなり、より迅速な意思決定のために AI-DSS の利用が進められている状況では、人間が AI システムに批判的に関わる能力は一層制約されかねないと言われる。意思決定者は時間的プレッシャーにさらされる上、状況によっては、出力結果をクロスチェックしたり、異議を唱えたりするための時間自体が限られるためである⁽⁸¹⁾。こうした制約の中で、人間の意思決定者が、AI-DSS の能動的な管理者ではなく、「単にボタンを押すだけ」の受動的な監督者となりかねないことが懸念されている⁽⁸²⁾。

3 国際人道法の義務を履行する上での有効性

武力を行使するに当たり、紛争当事者は、目的の達成を追求するのと同時に、国際人道法の諸規則を遵守する義務を負う。AI-DSS を武力行使の文脈で用いる場合も例外ではなく、軍事目標とそれ以外の文民や民用物等を区別して、軍事行動の対象を軍事目標に限定し（区別原則）、非軍事目標である文民又は民用物への巻き添え被害を最小限にとどめ（均衡性原則）、そのために必要な全ての実行可能な予防措置を講じる（予防原則）という基本原則を始めとして、国際人道法に基づく義務の適用を受けることになる⁽⁸³⁾。こうした中、AI-DSS の利用に伴うリス

(76) *ibid.*, pp.16-17.

(77) *ibid.* なお、近年、AI の挙動をより透明に、解釈可能に、説明可能にするための「説明可能な AI」(explainable AI) に向けた研究も行われている。例えば、次を参照。Scharre, *op.cit.*(20), pp.236-237.

(78) ICRC and Geneva Academy, *ibid.*, p.17.

(79) Paoli et al., *op.cit.*(9), p.12; Boulanin, *op.cit.*(64), p.112.

(80) Paoli et al., *ibid.*; Boulanin, *ibid.*; Larry Lewis and Diane Vavrichek, “Military Applications of AI: Current Situation and Future Considerations,” Geiß and Lahmann, eds., *op.cit.*(20), p.69.

(81) Álvarez, *op.cit.*(72); ICRC and Geneva Academy, *op.cit.*(22), p.17; Michel, *op.cit.*(67), p.45; Klaus, *op.cit.*(30)

(82) Álvarez, *ibid.* 他に、次を参照。ICRC and Geneva Academy, *ibid.*

(83) Arthur Holland Michel, “Background on Decision Support Systems in Military Decision-making on the Use of Force,” ICRC and Geneva Academy, *ibid.*, p.25. なお、国際人道法における兵器関連の規則は、本節で挙げた①区別原則（議定書 I 第 48 条）、均衡性原則（同第 51 条 5 項 (b)）及び予防原則（同第 57 条）の三原則を始めとする戦闘の方法（害敵方法）に関する規則と、②戦闘の手段（害敵手段）に関する規則（IV 章参照）に分けられる。①は特定の状況下で特定の標的に対する特定の攻撃が合法か否かに関するもので、②は兵器そのものの合法性に焦点を当てているという違いがある。例えば、いかなる兵器も、無差別に文民を攻撃するなど不正に使用される可能性があるが、それは①の害敵方法によるものと考えられる。一方、兵器が持つ性質や性能によって（通常において予期される使用方法によっても）軍事目標と非軍事目標とを区別することのできないような兵器があり、②はこのような観点から害敵手段を規制するものと考えられる。②の検討に当たっては、①を背景として考慮する必要があるとしつつ、①を標的化法（Targeting Law）、②を兵器法（Weapons Law）と呼んで区別することがある。William H. Boothby, *Weapons and the Law of Armed Conflict*, Second edition, Oxford: Oxford University Press, 2016, pp.1-2, 34-35; “Professor William Boothby - Written Evidence (AIW0003),” (AI in Weapon Systems Committee), March 2023. UK Parliament Website <<https://committees.parliament.uk/writtenevidence/119662/pdf/>>; Tim McFarland, *Autonomous*

クも踏まえて、武力行使に関する意思決定のための AI-DSS の利用が、意思決定者による国際人道法の義務を履行する能力にどのような影響を及ぼし得るかについて、議論がなされてきた。

意思決定者が国際人道法の義務を履行する能力に対して AI-DSS が何をもたらすかという点について、一方では、特定の場合に、国際人道法の遵守を促進する可能性があるとの見方がある。ICRC は、AI-DSS が「利用可能な情報の収集と分析を迅速かつ広範に行うことで、文民へのリスクを最小限に抑えることを促し」得るもので、意思決定者が、そうした AI-DSS を用いることによって、国際人道法の遵守に必要な十分な情報を得て、「より適切な判断を下せるようになる」可能性があることを認めている⁽⁸⁴⁾。ICRC の法律顧問であるウェン・チョウ (Wen Zhou) 氏らによれば、例えば、利用可能な文民及び民用物に関する既存のデータを最大限に活用したり、文民への偶発的な被害を回避し又は最小限に抑えるための攻撃手段や方法を提案したりするような使い方が考えられるという⁽⁸⁵⁾。AI はその利用方法によっては人間による国際法の遵守能力の向上に役立つとする認識は、2024 年 9 月の「REAIM 行動のための青写真」(I 章 1 節 (1) 参照) や 2024 年 12 月の国連総会決議 (I 章 1 節 (3) 参照) にも示されており、国際的に広く共有されつつある。

同時に、意思決定者が国際人道法の義務を履行する上での AI-DSS の全般的な有効性には、疑問も呈されている。その背景には、武力行使を規制する国際人道法の重要な規則の多くが、定性的で文脈に応じた価値判断の適用を前提としていることがある⁽⁸⁶⁾。例えば、均衡性原則においては、「予期される具体的かつ直接的な軍事的利益」と比較して、「巻き添えによる文民の死亡、文民の傷害、民用物の損傷」が「過度」であるかを評価する必要がある⁽⁸⁷⁾。また、区別原則においては、文民であっても「敵対行為に直接参加」していれば軍事目標となる⁽⁸⁸⁾。しかし、どのような行為がそれに該当するのかが明確に定義されていない中で、その文民が敵対行為に直接参加しているか否かを評価し、「文民であるか否かについて疑義がある場合には、文民とみなす」ものとされている⁽⁸⁹⁾。そもそも、人物や物体が軍事目標になるかは、その時点において要件を満たすか否かで変わり得る⁽⁹⁰⁾。このような文脈に依存する価値判断を行うことは、人間の意思決定者にとっても難しいが、判断基準をアルゴリズムであらかじめ包括的

Weapon Systems and the Law of Armed Conflict: Compatibility with International Humanitarian Law, Cambridge: Cambridge University Press, 2020, p.85.

(84) ICRC, “Artificial Intelligence and Machine Learning in Armed Conflict: A Human-centred Approach,” *International Review of the Red Cross*, vol.102 no.913, March 2021, p.468. <<https://international-review.icrc.org/sites/default/files/reviews-pdf/2021-03/ai-and-machine-learning-in-armed-conflict-a-human-centred-approach-913.pdf>>; ICRC and Geneva Academy, *ibid.*, pp.12-13, 37.

(85) Wen Zhou and Anna Rosalie Greipl, “Artificial Intelligence in Military Decision-making: Supporting Humans, Not Replacing Them,” August 29, 2024. Humanitarian Law & Policy Blog <<https://blogs.icrc.org/law-and-policy/2024/08/29/artificial-intelligence-in-military-decision-making-supporting-humans-not-replacing-them/>> 他に、次を参照。ICRC, “Submission to the United Nations Secretary-General on Artificial Intelligence in the Military Domain,” 11 April 2025, pp.6, 10. <https://www.icrc.org/sites/default/files/2025-04/ICRC_Report_Submission_to_UNSG_on_AI_in_military_domain.pdf>; UNIDIR’s Security and Technology Programme, *op.cit.*(19), p.14.

(86) ICRC and Geneva Academy, *op.cit.*(22), p.15.

(87) 議定書 I 第 51 条 5 項 (b)

(88) 議定書 I 第 51 条 3 項

(89) 議定書 I 第 50 条 1 項 なお、ICRC は、戦闘への直接的な参加という概念について、同機関による解釈を推奨事項として示した詳細なガイダンスを公表している。Nils Melzer, *Interpretive Guidance on the Notion of Direct Participation in Hostilities under International Humanitarian Law*, Geneva: ICRC, May 2009, pp.75-76. <<https://shop.icrc.org/download/ebook?sku=0990/002-ebook>>

(90) 黒崎将広「科学技術・イノベーションと国際安全保障法制—自律型兵器システムの適正な研究開発利用のための法的視座—」『オペレーションズ・リサーチ』67 巻 10 号, 2022.10, pp.521-522. <https://orsj.org/wp-content/corsj/or67-10/or67_10_520.pdf>

に満足に定義できる可能性は低く、ICRC も、少なくとも現在の技術水準で、「AI-DSS の機械処理プロセスに組み込む（encoding）には、おそらく適していない」ことを指摘している⁽⁹¹⁾。

実際には、意思決定者が国際人道法の義務を履行する上での AI-DSS の有効性の程度は必ずしも一律に決まるものではなく、利用環境や対象とする標的の種類等の要素によっても左右されると考えられる。例えば、ストックホルム国際平和研究所（SIPRI）上級研究員であるヴィンセント・ブーラニン（Vincent Boulanin）氏らは、作戦環境が予測可能で急変する可能性が低い環境（例えば、文民が居住しない砂漠などの開放的な環境）では AI システムは比較的有効に機能するかもしれないが、複雑で動的な環境（例えば、文民が行き交う市街地）ではその有効性は限定されるであろうことを指摘している⁽⁹²⁾。また、現在の技術では、人的目標に関する区別（例えば、文民であるか戦闘員であるかの区別）を確実に行うことは難しいとしても、エコノミスト誌の記事が指摘するように、物的目標のうち、戦車や艦船などの大型の軍事装備品など、軍事目標を比較的容易に区別できる場合もあると考えられる⁽⁹³⁾。

4 リスク軽減措置の提案

AI-DSS の利用に伴ってもたらされるリスクの程度は、結局のところ、技術的な対応に加え、人間による利用方法（どのようなシステムを、何を対象に、どのような環境で、どのように使用するか等）に依存すると考えられる。そのため、AI-DSS のリスクを軽減する上での中心的な課題としては、技術的な対応を可能な限り行うことに加えて、様々なリスクを理解した上で人間がシステムを適切に利用できるかが問われることになるだろう⁽⁹⁴⁾。英国のアバディーン大学政治・国際関係学部上級講師のジェームズ・ジョンソン（James Johnson）氏は、AI システムのリスクについて、利用者に対する訓練や文化、システムのインターフェース、軍事的な教義（ドクトリン）への反映といった点を含め、システムが「どのように実用化され、運用されるかに大きく依存することになるであろう」としている。そして、人間が AI システムを信頼して利用するに当たっては、「人間とマシンそれぞれの相対的な強みを理解し、どのような組合せが最も効率よく機能するか」、「人間とマシンを我々の戦争遂行システムにいかに効果的に統合するか」といった点が重要になると指摘している⁽⁹⁵⁾。

では、意思決定支援システムに AI を組み込むことによってもたらされる様々なリスクに対応し、適切に利用していくには、具体的にどのような対応措置がとられるべきであろうか。この点に関しては、現時点で国際的な議論がなお進行中である。一つの対応措置で解決できるも

(91) ICRC and Geneva Academy, *op.cit.*(22), p.15. 他に、次を参照。Vincent Boulanin and Maaïke Verbruggen, “Article 36 Reviews: Dealing with the Challenges Posed by Emerging Technologies,” *SIPRI Background Paper*, December 2017, pp.20-21. <https://www.sipri.org/sites/default/files/2017-12/article_36_report_1712.pdf>; Michel, *op.cit.*(59); Zhou and Greipl, *op.cit.*(85) なお、将来において、AI がその技術的な進展等によって国際人道法の諸規則に沿った武力行使を十分に支援できるようになる可能性に触れた文献として、次を参照。Boothby, *op.cit.*(83), pp.251-252; Masahiro Kurosaki, “Towards the Special Computer Law of Targeting: ‘Fully Autonomous’ Weapons Systems and the Proportionality Test,” Claus Kreß and Robert Lawless, eds., *Necessity and Proportionality in International Peace and Security Law* (Lieber Studies, vol.5), New York: Oxford University Press, 2021, pp.409-435.

(92) Boulanin and Verbruggen, *ibid.*, p.21. 他に、次を参照。“Model Major-general,” *Economist*, vol.451 no.9402, June 22, 2024, pp.18-21.

(93) “Model Major-general,” *ibid.*

(94) Johnson, *op.cit.*(69), pp.182, 212; *ibid.*; Noah Sylvia, “The Israel Defense Forces’ Use of AI in Gaza: A Case of Misplaced Purpose,” 4 July 2024. Royal United Services Institute (RUSI) Website <<https://rusi.org/explore-our-research/publications/commentary/israel-defense-forces-use-ai-gaza-case-misplaced-purpose>>

(95) Johnson, *ibid.* ジョンソン氏は、元 NATO 事務次長のローズ・ゴットモラー（Rose Gottemoeller）氏の発言を参照する形でこのように記している。

のでもなく、複数のガバナンスのレベルで、AI の様々な段階に応じて必要な対応を組み合わせる必要があると考えられている。国連軍縮研究所（UNIDIR）は、2020 年の報告書で、AI-DSS の課題に対応するには、ワンストップの解決策はなく、様々な利害関係者間でそれらの専門知識を活用することによる「対応策とインセンティブ構造の網が必要となる」と指摘している⁽⁹⁶⁾。さらに、2025 年 7 月の報告書では、より具体的に、軍事分野における AI の効果的なガバナンスに向けて、「多国間、地域、国家レベルの協調した行動を組み合わせ」て各種の措置を実施することを提案している⁽⁹⁷⁾。そこで本節では、以降、AI-DSS の利用をめぐる対応措置に関する関連機関及び研究者による主要な提案や考察の幾つかを紹介する。

一つには、AI-DSS の特定の用途での利用に制約を設けることに関する提案がある。制約の必要性がしばしば指摘される用途として、核の指揮統制システムへの AI の統合など、核兵器の使用に関する意思決定がある。核兵器の使用に関しては安全性が最重視される領域で、過去の有用なデータも乏しく、誤作動等のリスクが非常に高いこと等から、核の指揮統制に直接的に影響するような意思決定の役割は AI に委ねるべきではないといった指摘がみられる。例えば、ICRC は、2025 年 4 月の AI-DSS の規制に関する予備的な勧告の中で、「このようなツール [AI-DSS] は決して核兵器の指揮統制システムに組み込まれてはならない」と主張している⁽⁹⁸⁾。アバディーン大学のジョンソン氏は、「核の早期警戒用の機械学習（ML）アルゴリズムの設計と訓練は、ほぼ完全にシミュレーション・データに依存することになり、安全性が最も重要視される核の領域では非常にリスクが高い」とし、当面の間、核兵器の使用等に関する戦略的意思決定は「基本的に人間の努力に頼り続けるしかないであろう」としている⁽⁹⁹⁾。

AI-DSS の利用方法に一定の制約を設けることに関する提案もある。例えば、ICRC は、AI-DSS における予測可能性の欠如等を考慮し、AI-DSS をあらかじめ厳密にテストされたタスクや文脈に限定して利用することを提案している⁽¹⁰⁰⁾。また、AI-DSS の運用環境を制限することも選択肢となり得る。SIPRI のブーランニン氏は、システムが標的の位置を検出して攻撃することを可能とする特定の領域、場所の種類（例えば、居住者のいない地域）、特定の時間枠などを設定することが考えられ、これらはシステムに直接プログラムし得るとしている⁽¹⁰¹⁾。このほか、ICRC は、予備的な勧告の中で、人間を標的とする自律的なシステムと AI-DSS とを組み合わせる利用することを制限するよう提案している⁽¹⁰²⁾。

人間の意思決定者が AI-DSS に批判的に関与するための措置も提案されている。まず、利用者に対して、AI-DSS の出力結果を批判的に検討するための継続的な教育や訓練を行う必要性が広く指摘されている。UNIDIR は、AI のガバナンスに向けて国家レベルで行うべき事項として、「AI に関する軍人のための広範な訓練プログラムを開発し、AI に精通した新世代の将

⁽⁹⁶⁾ Paoli et al., *op.cit.*(9), p.35.

⁽⁹⁷⁾ 各レベルで行うべき政策行動として、例えば、①多国間レベルでは、軍事領域における責任ある AI のための包括的な基本原則の策定とその国際的な自主規範又は指針への発展、信頼醸成措置の策定、能力開発プログラムの開発・実施、②地域レベルでは、地域特有の信頼醸成措置、規範、指針の策定、③国家レベルでは、安全保障と防衛における AI に関する包括的な国家戦略又は政策の策定、強固なガバナンス構造と審査プロセスの確立、透明性と説明責任に関する施策の実施、訓練プログラムの開発、軍事作戦の指針の見直しなどを挙げている。

UNIDIR's Security and Technology Programme, *op.cit.*(19), pp.7, 18, 22, 27.

⁽⁹⁸⁾ ICRC, *op.cit.*(85), p.10.

⁽⁹⁹⁾ Johnson, *op.cit.*(69), p.173. 他に、次を参照。"Model Major-general," *op.cit.*(92)

⁽¹⁰⁰⁾ ICRC and Geneva Academy, *op.cit.*(22), p.22; ICRC, *op.cit.*(85), p.10.

⁽¹⁰¹⁾ Boulanin and Verbruggen, *op.cit.*(91), p.23.

⁽¹⁰²⁾ ICRC, *op.cit.*(85), p.10.

官や専門家を育成することにより、人的資本と訓練に投資する」ことを挙げ、これには「技術的な訓練だけでなく、作戦での AI 利用の倫理的・法的側面に関する訓練も含まれる」としている⁽¹⁰³⁾。また、システムの開発者側の対応として、ICRC は、情報量が過剰となって利用者が適切に関与できなくなることがないように一貫性のある形で情報を提示し、出力結果を批判的に検討することが可能なユーザーインターフェース及びプラットフォームの設計を行うこと、明確で活用しやすい実践的な利用のガイダンスを提供することを提案している⁽¹⁰⁴⁾。ICRC は、武力行使の際に国際人道法の下で要求される定性的評価を人間が行えるよう、意思決定プロセスを特定の時点で減速させる仕組みを組み込むことも提案している⁽¹⁰⁵⁾。

データに関するガバナンスや品質管理の必要性も指摘されている。UNIDIR は、国家レベルで防衛組織や規制当局が行うべきこととして、「全ての軍事用 AI アプリケーションに対して、強固なデータプラクティスとガバナンスの枠組みを導入する」ことを挙げ、これには「質が高く、代表的かつ分解可能なデータセットを整備 (curating) することへの投資、データの検証と更新の手順の確立、データセキュリティ対策の実施」が含まれるとしている⁽¹⁰⁶⁾。

AI-DSS の研究・開発段階からの適切な試験及び評価、運用後における継続的な性能監視や再評価を行うことも求められている⁽¹⁰⁷⁾。また、その国際人道法の側面からの合法性の評価に向けて、兵器等の「法的審査」の枠組みを活用することが議論されており、次のⅣ章でその概要、課題等を検討する。

Ⅳ AI-DSS への法的審査の適用に向けた課題と展望

1 法的審査の概要

(1) 法的審査とは

国家が AI-DSS の利用に伴う各種のリスクに対応し、意思決定者による国際法の遵守を確保するためには、システムの研究・開発段階から、その性能を検証し信頼性を確保する取組が重要になる。そのために活用することが期待されている既存の国際人道法上の枠組みとして、新たな兵器等の導入に当たってその使用の合法性を審査するための法的審査がある。法的審査は、36 条審査 (article 36 review)、兵器審査 (weapons review) などとも呼ばれる。

議定書Ⅰの第 36 条は、法的審査について、次のとおり規定している。

「締約国は、新たな兵器、戦闘の手段又は方法 (a new weapon, means or method of warfare) の研究、開発、取得又は採用 (study, development, acquisition or adoption) に当たり、その使用がこの議定書又は当該締約国に適用される他の国際法の諸規則により一定の場合又は全ての場合に禁止されているか否かを決定する義務を負う。」

この条文に基づき、締約国は、新たな兵器、戦闘の手段又は方法が議定書Ⅰの規定とその他

⁽¹⁰³⁾ UNIDIR's Security and Technology Programme, *op.cit.*(19), pp.8, 28-29. 他に、次を参照。 *ibid.*, p.11; Klaus, *op.cit.*(30); Michel, *op.cit.*(67), p.31.

⁽¹⁰⁴⁾ ICRC, *ibid.*, p.9; ICRC and Geneva Academy, *op.cit.*(22), pp.19-20.

⁽¹⁰⁵⁾ ICRC, *ibid.*, p.11; Ruben Stewart and Georgia Hinds, "Algorithms of War: The Use of Artificial Intelligence in Decision Making in Armed Conflict," October 24, 2023. Humanitarian Law & Policy Blog <<https://blogs.icrc.org/law-and-policy/2023/10/24/algorithms-of-war-use-of-artificial-intelligence-decision-making-armed-conflict/>>

⁽¹⁰⁶⁾ UNIDIR's Security and Technology Programme, *op.cit.*(19), pp.8, 28.

⁽¹⁰⁷⁾ ICRC and Geneva Academy, *op.cit.*(22), p.18; *ibid.*, pp.8, 27.

の適用可能な国際法の規則に適合しているか否かを評価する義務を負っている⁽¹⁰⁸⁾。

法的審査の義務は、議定書 I 第 35 条 1 項に定める、紛争当事者が戦闘の方法及び手段を選ぶ権利は無制限ではないという基本原則に紐（ひも）付いている⁽¹⁰⁹⁾。国際人道法上、戦闘の手段（害敵手段）として用いられるあらゆる兵器及び兵器システムは、①特定の軍事目標のみを対象とすることができないなどの無差別な戦闘の方法及び手段の禁止（同第 51 条 4 項 (b)(c)）、②人的軍事目標（戦闘員等）に過度の傷害又は無用の苦痛を与える兵器や戦闘の方法の禁止（同第 35 条 2 項等）、③自然環境に対して広範、長期的かつ深刻な損害を与える戦闘の方法及び手段の禁止（同第 35 条 3 項及び第 55 条）の各原則に服する⁽¹¹⁰⁾。また、毒物又は施毒兵器、対人地雷、盲目化レーザー兵器といった個別の兵器の使用等を禁止する条約が制定されている。ただし、こうした禁止原則や条約の存在にもかかわらず、新たに開発や取得を行う兵器の使用が国際法上許されるか否かの判断は、実際には、個別の国家の判断に基づく履行確保に委ねられるところが多い⁽¹¹¹⁾。法的審査は、この国家による履行確保を補強するものとして、国家に対し、新たな兵器の使用前にその合法性を審査する義務を締約国に課しているものと考えられている⁽¹¹²⁾。

従来の法的審査において、これらの禁止原則や個別の条約を始めとする害敵手段に関する法規制は、兵器そのものの合法性を評価するに当たり、最も一般的に適用される審査基準となってきた⁽¹¹³⁾。なお、いかなる兵器も不正に使用される可能性があることに関して、ICRC は、審査時に予期される兵器等の通常の使用に基づいて評価を行うことを推奨している⁽¹¹⁴⁾。

⁽¹⁰⁸⁾ 法的審査の義務が、慣習国際法（慣行と法的信念によって形成される国際法の法源）として議定書 I の非締約国まで拘束するか否かについては、評価は定まっていない。黒崎ほか 前掲注(1), pp.419-420。

⁽¹⁰⁹⁾ Yves Sandoz et al., eds., *Commentary on the Additional Protocols of 8 June 1977 to the Geneva Conventions of 12 August 1949*, Geneva: ICRC, 1987, paras 1463-1466. <<https://ihl-databases.icrc.org/en/ihl-treaties/api-1977/article-36/commentary/1987>>; Vincent Boulanin, “Implementing Article 36 Weapon Reviews in the Light of Increasing Autonomy in Weapon Systems,” *SIPRI Insights on Peace and Security*, no.2015/1, November 2015, p.3. <<https://www.sipri.org/sites/default/files/files/insight/SIPRIInsight1501.pdf>>

⁽¹¹⁰⁾ このうち、①及び②の禁止原則については、「逸脱不可能な国際慣習法原則」としてその普遍的な地位を確立しているとされる。一方、③の禁止原則については、慣習国際法として議定書 I の非締約国まで拘束するか否かをめぐり、各国の見解は一致していない。黒崎ほか 前掲注(1), pp.388-389. なお、①の禁止原則は、戦闘の方法（害敵方法）に関する規則のうちの区別原則（議定書 I 第 48 条）と関係しており、区別原則を具体化する内容の、軍事目標と非軍事目標（文民又は民物）とを区別しない無差別攻撃を禁止する規定（同第 51 条 4 項）の下に置かれている。

⁽¹¹¹⁾ 同上, p.419.

⁽¹¹²⁾ Sandoz et al., eds., *op.cit.*(109), paras 1463-1466; Boulanin and Verbruggen, *op.cit.*(91), p.vii; 同上

⁽¹¹³⁾ Boothby, *op.cit.*(83), pp.347-349; Boulanin and Verbruggen, *ibid.*, p.4.

⁽¹¹⁴⁾ ICRC は、議定書 I のコメンタリー（条約注釈書）及び法的審査の手續に関するガイドにおいて、ほぼ全ての兵器は禁止されるような方法で不正に使用される可能性があるため、全ての潜在的な不正使用の可能性を予期又は分析する必要はなく、「ある兵器の通常又は予期される使用が、一定の場合又は全ての場合に禁止されるか否か」を評価すればよいとする解釈を示している。Sandoz et al., eds., *op.cit.*(109), paras 1466, 1469; Kathleen Lawand et al., *A Guide to the Legal Review of New Weapons, Means and Methods of Warfare: Measures to Implement Article 36 of Additional Protocol I of 1977*, Revised, Geneva: ICRC, November 2006, p.10. <https://www.icrc.org/sites/default/files/external/doc/en/assets/files/other/icrc_002_0902.pdf> なお、ICRC はスイス国内法上の法人であり、政府間国際機関ではないが、19 世紀後半以来、関係条約に基づき武力紛争犠牲者保護のために重要な役割を担ってきた。条約上の任務以外にも、ジュネーブ諸条約その他の関係条約の起草を中心となって行ったほか、近年でも国際人道法に関する調査研究を積極的に行うなどしており、ICRC が刊行するコメンタリーや報告書は、締約国による解釈適用に強い影響を与えている。黒崎ほか 前掲注(1), pp.580-581. 本章では、以降、ICRC が刊行した議定書 I のコメンタリー及び法的審査の手續に関するガイドを、一つの有力な解釈として参照するものとする。

(2) 履行の現状

第 36 条は、具体的な審査の実施方法は定めていない⁽¹¹⁵⁾。したがって、審査の実施方法に関する確立された基準はなく、その形式、作業方法、権限の範囲やレベルは国ごとに異なっている⁽¹¹⁶⁾。また、審査の履行に対する国際的な監視なども行われない中で、各国の国内手続に基づいて行われるため、法的審査においても、最終的な解釈権を有する国家に大きな裁量が委ねられているのが現状である⁽¹¹⁷⁾。こうした中、国家によっては、法的審査の実施に他国よりも意欲的でなかったり、実施する能力が低かったりする可能性もある⁽¹¹⁸⁾。

法的審査を実施している国は、(条約上の義務であるにもかかわらず) 議定書 I の締約国を含めて少数にとどまっていることも指摘されてきた⁽¹¹⁹⁾。SIPRI とアジア太平洋法及び安全保障研究所 (APILS) の研究者それぞれによる近年の調査によれば、議定書 I の締約国で第 36 条に基づく審査制度を有していることが分かっている国としては、英国、ベルギー、ドイツ、フランス、オランダ、ニュージーランド、ノルウェー、スウェーデン、フィンランド、スイス、デンマーク、エストニア、オーストラリア、ブラジル、カナダなどがある⁽¹²⁰⁾。なお、米国やイスラエルのように、非締約国であっても事実上の発達した法的審査制度を有している国もある⁽¹²¹⁾。

⁽¹¹⁵⁾ Lawand et al., *ibid.*, p.20.

⁽¹¹⁶⁾ Boulanin and Verbruggen, *op.cit.*(91), p.1. この点に関し、各国に画一的な法的審査のシステムを課すことは、各国固有の兵器取得プロセスに法的審査の手続を組み込むのを阻害することになるとする見方がある。Justin McClelland, “The Review of Weapons in Accordance with Article 36 of Additional Protocol I,” *International Review of the Red Cross*, vol.85 no.850, June 2003, p.414. <https://www.icrc.org/sites/default/files/external/doc/en/assets/files/other/irrc_850_mcclelland.pdf>

⁽¹¹⁷⁾ Boulanin, *op.cit.*(100), p.2; 福井康人「新たな兵器の合法性審査を事例として—国際人道法・国際刑事法と軍縮国際法との協働—」日本軍縮学会編『軍縮・不拡散の諸相—日本軍縮学会設立 10 周年記念—』信山社, 2019, pp.88-89.

⁽¹¹⁸⁾ *ibid.*

⁽¹¹⁹⁾ Boothby, *op.cit.*(83), p.343; Boulanin and Verbruggen, *op.cit.*(91), p.1; 黒崎ほか 前掲注(1), p.420.

⁽¹²⁰⁾ SIPRI のブーランニン氏は、2017 年、ベルギー、ドイツ、オランダ、英国、米国など主要な 9 か国を対象に、質問状を送付し、各国からの回答に基づいて法的審査の手続の概要をまとめた報告書を刊行している。Vincent Boulanin and Maaïke Verbruggen, “SIPRI Compendium on Article 36 Reviews,” *SIPRI Background Paper*, December 2017. <https://www.sipri.org/sites/default/files/2017-12/sipri_bp_1712_article_36_compendium_2017.pdf> また、安全保障を実現するための手段としての国内法と国際法の発展を促進することを目的とし、オーストラリアを拠点として活動するアジア太平洋法及び安全保障研究所 (Asia-Pacific Institute for Law and Security: APILS) は、自国の法的審査の手続等を公表している国の情報を収録したポータルサイトを構築している。2025 年 8 月現在、同サイトには、審査プロセスに関する詳細を公表している国として 17 か国、(審査プロセスに関する詳細な情報は得られていないものの) 審査の実施を公的に表明している国として 8 か国の情報が掲載されている。“Information Portal on the Legal Review of Weapons.” APILS Website <<https://apils.org/legal-review/>> なお、日本における法的審査については、防衛装備品の研究・開発・調達・配備などの検討を行う防衛省の会議に法務官が出席し、その意見が反映されることによって、合法性の評価が行われるとされる。また、日本の防衛省は、2025 年 6 月、防衛分野における責任ある AI 適用のため、AI 技術を適用した装備品等のライフサイクルのうちの構想段階から研究・開発段階までの間において、当該装備品等のリスクの程度を判断し、必要に応じて「法的・政策的審査」と「技術的審査」を行うことを定めたガイドラインを公表している。福井 前掲注(107), p.75; 防衛省「装備品等の研究開発における責任ある AI 適用ガイドライン 第 1 版」[2025.6]. <https://www.mod.go.jp/atla/soubiseisaku/ai_guideline/ai_guideline_ver.01.pdf>

⁽¹²¹⁾ 米国は 1977 年 12 月に議定書 I に署名したが、2025 年 8 月現在まで未批准である。イスラエルは議定書 I に署名も批准もしていない。一方、米国やイスラエルには、「兵器の合法性に関する審査プロセスを長年高度に発展させ、形式化してきた」事実があり、これらのプロセスは「第 36 条が議定書 I の締約国に求めているものと類似している」とされる。なお、米国は、議定書 I の採択 (1977 年) より前の 1974 年からこうした審査を実施していることを明らかにしている。Kenneth Anderson et al., “Adapting the Law of Armed Conflict to Autonomous Weapon Systems,” *International Law Studies*, vol.90, 2014, p.398. <<https://digital-commons.usnwc.edu/cgi/viewcontent.cgi?article=1015&context=ils>>; U.S. Department of Defense, *Law of War Manual*, June 2015, Updated July 2023, paras. 6.2.3, 19.20.1.2. <<https://media.defense.gov/2023/Jul/31/2003271432/-1/-1/0/DOD-LAW-OF-WAR-MANUAL-JUNE-2015-UPDATED-JULY%202023.pdf>>; 黒崎ほか 前掲注(1), p.420.

法的審査には、このように履行に関する限界もあるが、近年、サイバー兵器や LAWS といった新興技術を用いた新たな兵器の規制に関心が集まる中で、国家が国際的義務に従って敵対行為を遂行できることを確保する方法になり得る枠組みとして、注目されるようになってい⁽¹²²⁾。AI-DSS を含む軍事分野における AI 利用に関しても、規制へのアプローチが模索される中で、法的審査の枠組みを活用し、AI システムの研究・開発段階からその合法性を審査する可能性が議論されるようになった。推進派の研究者らや関連機関の間において、例えば、ICRC の法律顧問であるネッタ・グサック (Netta Goussac) 氏は、法的審査が、敵対行為の実施を規制する国際人道法の規則を遵守できない AI システムの開発と使用に対する「強力な予防策となり得る」と述べている⁽¹²³⁾。また、UNIDIR も、AI のガバナンスに向けて国家レベルで行うべき事項に、AI システムの法的審査プロセスの確立を挙げ、そのために既存の第 36 条に基づく法的審査の枠組みを活用する可能性に言及している⁽¹²⁴⁾。また、国際的な規制に向けた取組においても、2024 年 9 月の「REAIM 行動のための青写真」(I 章 1 節 (1) 参照) には、効果的な法的審査の手続に関する情報交換及び協議を奨励する旨が盛り込まれ、また、2023 年 2 月の米国による「AI と自律性の責任ある軍事利用に関する政治宣言」(I 章 1 節 (2) 参照) にも、法的審査の実施への言及がなされている。

2 法的審査の適用に向けた課題

AI-DSS の審査のために議定書 I の第 36 条で定める法的審査の義務を活用するに当たっては、幾つかの主要な課題があることが指摘されている。ICRC が公表している議定書 I のコメンタリー (条約注釈書) 及び法的審査の手続に関するガイドも参照しながら、主要な課題を確認する。

(1) 審査の対象

まず、AI-DSS が第 36 条の定める法的審査の対象となり得るかという問題がある。議定書 I の第 36 条は、審査の対象について、「新たな兵器、戦闘の手段又は方法」としているが、それらの用語の定義は示されていない⁽¹²⁵⁾。AI-DSS は、ソフトウェアとそれを搭載したプラットフォームであり、兵器と連携させて使用される可能性があるが、兵器そのものではない。よって、AI-DSS が「戦闘の手段又は方法」に該当するのかが一つの焦点となる。

⁽¹²²⁾ サイバー兵器に関して、例えば、サイバー空間への国際法の適用に関する研究として広く参照されるタリンマニュアル 2.0 には、新たなサイバー戦の手段又は方法 (new means or method of cyber warfare) を法的審査の対象とする解釈を記している (ルール 110)。Michael N. Schmitt, ed., *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations*, Second edition, Cambridge: Cambridge University Press, 2017, pp.464-467. LAWS に関しては、2019 年 8 月の CCW の LAWS に関する政府専門家会合 (GGE) で採択された 11 項目から成る指導原則において、LAWS の開発・取得・採用に当たって、その使用が国際法により禁止されるか否かを判断すべきことが謳われた (e 項)。UN Doc. CCW/GGE.1/2019/3, *op.cit.*(7), p.13.

⁽¹²³⁾ Netta Goussac, “Safety Net or Tangled Web: Legal Reviews of AI in Weapons and War-fighting,” April 18, 2019. Humanitarian Law & Policy Blog <<https://blogs.icrc.org/law-and-policy/2019/04/18/safety-net-tangled-web-legal-reviews-ai-weapons-war-fighting/>>

⁽¹²⁴⁾ UNIDIR’s Security and Technology Programme, *op.cit.*(19), pp.7, 18, 22, 27. AI-DSS に第 36 条に基づく法的審査の適用を求める主要な見解として、本章 2 節の本文に挙げるものの他に、次を参照。ICRC, *op.cit.*(85), pp.2, 10; Damian Copeland, “Are AI-DSS a ‘Means or Method of Warfare’ Requiring Legal Review?” 18 June [2025]. Article 36 Legal Website <<https://www.article36legal.com/blog/are-aidss-means-or-methods-of-warfare>>

⁽¹²⁵⁾ ICRC は、「兵器又は戦闘の手段は、それが使用される戦闘の方法から切り離して評価することはできない」ものであり、「兵器の合法性は、その設計や意図された目的のみに依存するのではなく、戦場でどのように使用されるかにも依存する」とする解釈を示している。Lawand et al., *op.cit.*(114), p.10.

ICRC のガイドは、法的審査の対象について、対人・対物や殺傷性の有無は問わない全ての種類の兵器（兵器システムを含む。）、軍事ドクトリンや交戦規則等に従って兵器を使用する方法、研究開発された兵器か既製品かを問わず取得された兵器、初めて取得する兵器（技術的な新規性は問わない。）、機能変更する形で更新された兵器などが含まれるとしている⁽¹²⁶⁾。これに基づけば、武力行使の文脈で用いられる AI-DSS が審査の対象に明示的に含まれると解釈することは難しい。専門家の間には、「戦闘の手段又は方法」について、用語の定義がない中で合理的な解釈を適用する必要があるとして、手段と方法を一体として解釈し、それ自体は兵器でなくとも、それが属する部隊の攻撃能力に直接的な影響を与える装備品（例えば、地雷除去車両）を指すものとする説もある⁽¹²⁷⁾。

AI-DSS が法的審査の対象となるかについて、公式に見解を明らかにした国はまだないとされるが⁽¹²⁸⁾、推進派の研究者らの間では、対象とする場合の根拠について議論がなされている。イスラエルのツファット・アカデミック・カレッジ准教授のタル・ミムラン（Tal Mimran）氏は、イスラエルによるゴスペル等の AI-DSS の利用について取り上げた論考で、「戦闘の手段」を、軍事作戦を円滑にするために使用される装備品、システム、プラットフォーム、その他の関連機器に及ぶ広範な用語であると解釈し、攻撃のために配備されたツールである AI-DSS は、法的審査を受けるべき新たな戦闘の手段になり得ると指摘している⁽¹²⁹⁾。第 36 条を「戦争技術の審査」として概念化することを主張する立場から、オランダのアッセル研究所の研究者であるクラウディア・クロノフスカ（Klaudia Klonowska）氏は、AI-DSS を含めて、敵対行為の遂行に不可欠な全ての道具（instruments）及び方法は、攻撃能力を増強させることを意図し、第 36 条が未然に防ごうとしている不法な危害、損害、傷害につながる可能性があるため、法的審査の下に置かれるべきであると指摘している⁽¹³⁰⁾。

(2) 審査のタイミング

AI-DSS に関しては、法的審査の対象となった場合でも、既存の枠組みに基づく限り、審査のタイミングが問題になる。第 36 条は「研究、開発、取得又は採用」の段階における合法性の評価を求めている⁽¹³¹⁾。ここで注意を要する点として、条文上は、運用を開始した後の審査までは求めておらず、事前審査義務にとどまっていることがある⁽¹³²⁾。一方の AI-DSS は、運用後もデータ学習を継続して性能を向上させることが想定されている。このことは、システムの採用段階では問題がなかったとしても、以後にデータ学習を継続する限り、意図しない挙動を獲得する可能性をもたらすことになる⁽¹³³⁾。ICRC のガイドは、「既存の兵器を機能変更する形で更新したもの又は既に法的審査を通過した兵器をその後更新したもの」は審査対象に含む

⁽¹²⁶⁾ *ibid.*, pp.9-10.

⁽¹²⁷⁾ McClelland, *op.cit.*(116), pp.404-405.

⁽¹²⁸⁾ Copeland, *op.cit.*(124)

⁽¹²⁹⁾ Mimran and Dahan, *op.cit.*(49)

⁽¹³⁰⁾ Klaudia Klonowska, "Article 36: Review of AI Decision-support Systems and Other Emerging Technologies of Warfare," *Asser Research Paper*, 2021-02, April 2021, pp.13-14, 32. <<https://ssrn.com/abstract=3823881>>

⁽¹³¹⁾ ICRC のガイドは、審査のタイミングについて、兵器調達プロセスの全ての段階にわたるもので、特に研究の初期段階（構想、研究等）、開発段階（試作品の開発、試験等）及び取得段階（既製品の調達を含む。）を対象とする解釈を示している。Lawand et al., *op.cit.*(114), p.23.

⁽¹³²⁾ 黒崎ほか 前掲注(1), p.420.

⁽¹³³⁾ ICRC and Geneva Academy, *op.cit.*(22), p.17.

とする解釈を示しているが⁽¹³⁴⁾、現時点で、それが AI システムにおける運用後のデータ学習までも明示的に含むものと解釈されてはいない⁽¹³⁵⁾。

防衛大学校総合安全保障研究科教授の黒崎将広氏は、AI 技術の現状も踏まえ、「兵器審査の時期は特定の時期に限定されるべきではないだろう」とし、「事前審査に限らず、同システムの実戦投入後も継続してその動作性能を検証し改善するものでなければならない」と指摘している⁽¹³⁶⁾。SIPRI のブーランニン氏らは、AI システムによるデータ学習について、内部的な特性の変更が行われるもので、いつ新たな審査を必要とする変更とみなされるのか不明確であるとする。その上で、「その変更が当初の審査では考慮されなかった方法でシステムを動作させるようなもの、又は、システムや最終利用者による法的基準の遵守能力を変更するようなもの」であった場合には、改めて審査を行う必要があると指摘している⁽¹³⁷⁾。その一方、法的審査における既存の試験や評価の手法をそのまま AI システムに対しても適用することは現実的にも財政的にも難しく、審査のための新たな手法を開発する必要があることも指摘している⁽¹³⁸⁾。

(3) 審査の適切な実施

技術の新規性は、場合によっては法的審査の実施に課題をもたらす。新しい技術によって、「古い法的概念を新たに見直す必要が生じたり、新たなリスクをもたらしたりする可能性」があるためである⁽¹³⁹⁾。機械学習ベースの AI システムに関しても、データ学習がシステムの機能に予測不可能性をもたらし、システムがその運用前にはシミュレーションできない動的な環境と相互作用することなどから、法的審査において特有の課題を生じさせている⁽¹⁴⁰⁾。UNIDIR は、AI システムの合法性の審査に当たって既存の法的審査の枠組みを活用する際の課題として、例えば、「ハードウェアだけでなく、アルゴリズムやデータも審査する必要があるが、こうしたプロセスにはほとんど前例がない」とした上で、法的審査をどのように、どのような基準で、どの頻度で行うかが課題となっていることを指摘している⁽¹⁴¹⁾。

ICRC のチョウ氏らは、武力行使に関する意思決定の一部を構成する全ての AI-DSS の法的審査の実施を検討すべきとし、その際、国家がその政策上の措置又はグッド・プラクティスとして審査を行っていくことの意義に言及している⁽¹⁴²⁾。AI 関連の技術やツールが法的審査に服することとなった場合における、審査の際に考慮すべき特有の要素について検討する試みもみられるようになってきている⁽¹⁴³⁾。SIPRI のブーランニン氏らは、AI-DSS の審査に関して、技術

(134) Lawand et al., *op.cit.*(114), pp.9-10.

(135) なお、AWS への法的審査の適用に関しては、AWS の運用開始後における定期的な検証（ランタイム検証（runtime verifications））を行う可能性が議論されている。その文脈で、ICRC の解釈でいう「更新」（modification）の語の概念について、伝統的に人間が兵器システムに手作業で行ってきたような改修のみならず、システム自体が最適な出力を追求する過程で自動的に行う動作上の変更をも含めるように解釈を拡大し得るとする見解もある。Tim McFarland and Zena Assaad, “Legal Reviews of *in situ* Learning in Autonomous Weapons,” *Ethics and Information Technology*, vol.25 no.9, 2023, pp.4-5. <<https://doi.org/10.1007/s10676-023-09688-9>>

(136) 黒崎ほか 前掲注(1), p.421.

(137) Boulanin and Verbruggen, *op.cit.*(91), p.24.

(138) *ibid.*

(139) *ibid.*, p.6. 法的概念の見直しに関して、例えば、サイバー戦争技術の場合であれば、国際法における「兵器」「攻撃」「武力紛争」などの概念の再考を要するとされる。*ibid.*, p.33.

(140) Goussac, *op.cit.*(123)

(141) UNIDIR's Security and Technology Programme, *op.cit.*(19), p.22.

(142) Zhou and Greipl, *op.cit.*(85)

(143) Boulanin and Verbruggen, *op.cit.*(91), pp.21-23; Dustin A. Lewis, “Legal Reviews of Weapons, Means and Methods of Warfare Involving Artificial Intelligence: 16 Elements to Consider,” March 21, 2019. Humanitarian Law & Policy Blog

的な要素のみならず、システムへの人間の関与の要素も考慮することが不可欠であることを強調している。その際に特に重要なポイントは、システム自体による国際法の遵守能力というよりも、システムの利用者が国際法の要件を遵守していることを確保するに当たってシステムに十分な信頼性があるかどうかであるとし、例えば、システムがどのような情報を人間の利用者に送り、その情報の信頼性はどの程度か、システムの動作を訓練された利用者が理解できるか否かといった点を考慮すべきことを指摘している⁽¹⁴⁴⁾。英空軍法務部副部長を務めた経歴を持つ国際法学者のウィリアム・ブースビー（William H. Boothby）氏は、「兵器システムや戦闘の方法に AI を適用するとき、特にそれが標的決定プロセスにおける人間の関与の程度を排除し、又は制限する効果をもたらす場合、追加の基準の適用が必要となる」としている。具体的には、従来人間が担ってきた意思決定に係るタスクの一部を AI が担うことになる中で、従来型の兵器（意思決定に関するタスクを行う機能を持たない兵器）の法的審査において最も一般的に適用されてきた、兵器そのものの合法性を評価するための法規制に加えて、「兵器システムとその AI の設計上の使用方法が、[区別原則、均衡性原則、予防原則を始めとする害敵方法に関する規制である] 標的化法で要求される決定を行うことを可能にするかどうか」を検討する必要があるとする見解を示している⁽¹⁴⁵⁾。

おわりに

AI-DSS は、その利用方法によっては、意思決定者による国際法の遵守を支援し得る一方で、結果として意思決定者を国際法違反に導いてしまう可能性も有する。AI-DSS を利用するに当たっては、そうしたリスクを認識した上で、人間の適切な関与を確保し、法的審査を含めて、様々な段階での必要な対応措置を組み合わせる実施していくことが求められる。AI-DSS の法的審査の実施に向けては、AI システムのリスクを踏まえた適切な審査のタイミング、要素、手法等を検討していく必要がある。また、各国が進んで AI-DSS の法的審査を実施し、その取組を対外的に公表・共有していくことは、法的審査の義務の履行を強化するものとして推奨されよう。

今後、AI 技術の更なる進展や利用可能なデータの拡充を背景として、軍事分野における AI 利用が一層進んでいくことが見込まれる。AI の開発競争の中にあっても、軍事分野における

<<https://blogs.icrc.org/law-and-policy/2019/03/21/legal-reviews-weapons-means-methods-warfare-artificial-intelligence-16-elements-consider/>> なお、AWS に関しては、その固有の特性を考慮した形での審査を促進する目的などから、APILS の研究者らによって法的審査のグッド・プラクティスを策定する試みが進められている。兵器システムに統合された AI 関連の技術やツールの審査も考慮されており、AI システムのリスクを踏まえた内容は、AI-DSS に関しても参考となる部分があると考えられる。例えば、継続的なデータ学習を行う AI の審査のタイミングに係る内容については、法的審査の見直し（reconsideration）を要する変更として、人間とシステムの相互作用の性質と範囲、アルゴリズムとデータセット、意図されたミッションセット、意図された標的セット、予測される敵対的対抗手段（adversarial countermeasures）への変更が挙げられている。“Good Practices in the Legal Review of Weapons, Means and Methods of Warfare,” Version 0.6, March 2025. APILS Website <<https://apils.org/wp-content/uploads/2025-03-aws-legal-review-good-practices-v0-6.pdf>>

⁽¹⁴⁴⁾ Boulanin and Verbruggen, *ibid.*, p.21.

⁽¹⁴⁵⁾ Bill Boothby, “AI Warfare and the Law,” *International Law Studies*, vol.104, 2025.1, pp.134-136. <<https://digital-commons.usnwc.edu/cgi/viewcontent.cgi?article=3097&context=ils>> 標的化法については、前掲注83参照。関連して、LAWS における法的審査の基準の適用に関する問題について取り上げた最近の文献としては、次がある。岩本誠吾「国際人道法における新兵器の法的審査制度と AI 兵器の登場による当該制度の変容」『産大法学』199号, 2025.7, pp.127-156.

AI 利用に国家が責任を負うとともに、各国政府、国際機関、開発企業、市民社会などの利害関係者が互いに協力し、AI 利用に伴うリスクや課題に効果的に対応するための道筋を見いだしていくことが期待される。

(きゅうこ さとみ)