

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 ——

岩 本 誠 吾

The Relationship of Legal Regulations among AI Civilian Use, AI Military Use and AI Autonomous Weapons

—FY 2024—

IWAMOTO Seigo

1. はじめに—三領域の AI 規制
2. 民生用 AI の規制
 - (1) AI 規制形態
 - (a) ソフト・ロー規制
 - (b) ハード・ロー規制
 - (2) 国連内での AI 規制
 - (a) 米国主導の国連総会決議
 - (b) 中国主導の国連総会決議
 - (c) 「未来のための約束」
3. 軍事用 AI の規制
 - (1) AI 意思決定支援システムの実態
 - (a) ウクライナ軍の AI システム
 - (b) イスラエル軍の AI システム
 - (2) 軍事用 AI の規制文書
 - (a) REAIM 2023・2024
 - (b) 米国の AI 軍事利用政治宣言
 - (c) 国連総会決議
4. AI 自律兵器の規制

- (1) CCW 内での AI 自律兵器規制
 - (a) GGE 第 1 会期
 - (b) GGE 第 2 会期
 - (c) CCW 締約国会議
 - (2) 国連内での AI 自律兵器規制
5. まとめにかえて
- (1) 三領域の AI 規制の相違点と共通点
 - (2) 日本の立場

* 添付翻訳資料

- ①「REAIM 行動ブループリント」
- ②LAWS・GGE「ローリング・テキスト 状況報告日：2024 年 11 月 8 日」

1. はじめに一三領域の AI 規制

人工知能（AI）¹⁾ およびその自律技術は、【図 1：AI 利用領域と規制状況】が示すように、現在、一般的な民生領域、軍事領域全般、そして、軍事領域の中の兵器領域の三領域において、すでに利活用されているかまたはその利活用のため研究開発が加速度的に推進されている。その現状に呼応して、各領域で非法的または法的規制が検討され制定されている。というのも、AI は、人間の社会生活における利便性の向上に役立つという利点とともに予測不能な潜在的リスクを併せ持つことから、AI の研究開発や利活用を推進する（アクセルを踏む）とともにそのリスクを禁止・規制する（ブレーキを掛ける）動きが同時進行するのである。

もはや AI の存在を否定できない現代社会における三領域での AI 規制の共通課題は、如何にしてヒトと AI を共存・共生させるか、如何にしてヒトと AI のより良い関係性を築くかである。当該三領域に通底する AI 規制の基本理念は、後述するように、人間の尊厳性であり人間中心主義である。

本稿の目的は、3 領域での AI 利用に関する 2024 年を中心に規制動向²⁾ を概観した後に、その相互関係を分析するとともに、今後の AI 規制、特に AI 自律兵器規制の方向性を考察することである。

1) 人工知能（AI：Artificial intelligence）は、「人間と同じ知的作業をする機械を工学的に実現する技術」と言われている。そして、AI は、意識を持ち意味を理解する「強い AI」と人間の思考を表面的に模倣する「弱い AI」、さらに、特定の用途に使われる「特化型 AI（Narrow AI）」と人間のように様々な状況や問題に対応できる「汎用型 AI（AGI：Artificial General Intelligence）」に分類される。人工知能学会「教養知識としての AI [第 1 回] AI ってなに？」https://www.ai-gakkai.or.jp/resource/ai_comics/comic_no1/、2024 年 12 月 14 日閲覧。現在の AI はすべて弱い AI であり特化型 AI であって、強い AI や汎用型 AI は存在していない。

2) 2022 年度および 2023 年度の当該規制動向については、以下を参照。岩本誠吾「特定通常兵器条約（CCW）」

	AI 自律兵器：無規制状態 * CCW・国連総会で審議中 * 人的制御・判断・関与必須 * 2層アプローチ（禁止と規制）
	軍事用 AI：軍事利用政治宣言（ソフト・ロー）
民生用 AI： 国内面・国際面で倫理原則・行動規範（ソフト・ロー）から 法規範（ハード・ロー）へ（EU の AI 規制法が可決・成立・発効） 人間中心主義、リスク・ベース・アプローチ（禁止と規制）	

【図1：AI 利用領域と規制状況】（著者作成）

2. 民生用 AI の規制

（1）AI 規制形態

（a）ソフト・ロー³⁾ 規制

民生用 AI の規制議論は、【年表1：民生用 AI 規制動向】が示すように、2019 年以降、国際社会において活発に行われるようになった。G7、G20、EU（欧州連合）、OECD（経済協力開発機構）および UNESCO（国連教育科学文化機関）など国際会合や国際機関は、採択文書において、主体はあくまで人間であり、AI は補完的な役割に徹し人間の監視下で取り扱われるべきであると位置付けた（人間中心アプローチ、human-centric approach⁴⁾）。個別国家、たとえば、中国もロシアも国内文書で同様のアプローチを採用している。

それら国際・国内の採択文書は、AI の取扱いに関する抽象的で大まかな原則や方向性を示唆するために、まずは、AI 倫理指針（EU）、AI 勧告（OECD⁵⁾、AI 倫理勧告（UNESCO）、AI 倫理規範（中国）、

締約国会議における自律型致死兵器システム（LAWS）規制議論の動向—2022 年度—『京都産業大学世界問題研究所紀要』38 巻 2023 年 3 月および同「自律型致死兵器システム（LAWS）規制議論の動向と今後の展望—2023 年度—」『産大法学』58 巻 2 号 2024 年 6 月。

3) ソフト・ロー（soft law）とは、法的拘束力（法的義務）の確立していない形成途上の法未満の規範である。その対義語であるハード・ロー（hard law）とは、法的拘束力の確立した法規範である。

4) 岩本誠吾「自律型致死兵器システム（LAWS）規制—多層的・多角的アプローチからの規制の試み」『軍縮研究』11 巻 1 号 2022 年 5 月 13–16 頁。G7、G20、EU、OECD の参加国・当事国は、重複を除くと 56 か国となる。

5) OECD は、2019 年に AI 原則を策定したが、生成 AI の急速な発展に対応すべく 2024 年 5 月 3 日に AI 原則を改訂した（OECD/LEGAL/0449, amended on: 03/05/2024, <https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449>, accessed 20 December 2024）。改訂 AI 原則は、生成 AI を念頭に「AI によって増幅された誤情報や偽情報に対処」すること（Section 1, IV, 1.2, a)）が明記され、AI システムが過度な損害を引き起こすかまたは望ましくない動作を示す危険性がある場合に、必要に応じてそれらを「停止、修復および/または安全に廃棄」（Section 1, IV, 1.4, b)）できるメカニズムを整備すべきと勧告した。OECD の AI 原則賛同国は、2019 年時点では加盟国 37 か国・非加盟国 8 か国の計 45 か国であったが、2024 年 12 月現在、加盟国 38 か国・非加盟国 9 か国の計 47 か国へと若干増加している。

【年表1：民生用 AI 規制動向】（著者作成）

年	月 日	日本、国際機構、国際会議および各国の一般的な AI 関連事項
2019	3/29 4/8 5/22 6/17 6/29 10/10	日本・内閣府「人間中心の AI 社会原則」 EU「信頼できる AI 倫理指針」 OECD「AI 理事会勧告」 中国「次世代 AI ガバナンス原則」公表 G20 大阪首脳宣言（G20 AI 原則） ロシア「2030 年までの AI 開発国家戦略」発表
2020	10/20 10/20	欧州議会 AI 倫理枠組み決議 G20 リヤド首脳宣言（G20 AI 原則推進国内政策例）
2021	4/21 9/25 10/26 11/23	EU 欧州委員会が「AI 規制法案」公表 中国「次世代 AI 倫理規範」公表 ロシア「AI 倫理準則」策定 UNESCO「AI 倫理勧告」採択
2022	11/30	オープン AI 社がチャット GPT（生成 AI）公表
2023	4/30 5/20 6/14 8 10/26 10/30 10/30 11/1～2 12/1 12/6 12/23	G7 デジタル・技術相会合関係宣言（AI 利用 5 原則合意） G7 広島首脳コミュニケ（広島 AI プロセス） 欧州議会が AI 規則法案採択 中国が生成 AI サービス管理暫定弁法を施行 国連事務総長が AI 諮問機関設立（39 名） 米国の AI 規制に関する大統領令 G7 首脳声明、AI 開発者向け広島プロセス国際指針・行動規範 英国で AI 安全サミット開催（11/2 に 28 か国 + EU がプレッチリー宣言賛同） 米国が AI 安全研究所設立の表明（11/1） G7 デジタル・技術閣僚声明、全 AI 関係者向け広島プロセス国際指針 G7 広島 AI プロセス包括的政策枠組みに最終合意 国連 AI 諮問機関による中間報告書公表
2024	3/13 3/21 5/3 5/21 5/21～22 6/12～15 7/1 9/19 9/22～23	欧州議会が生成 AI を含む AI 規制法を可決（523:46:49） 米国主導の「AI システム利活用」国連総会決議（A/RES/78/265） OECD「AI 理事会勧告」改訂版採択（38 か国 + 9 か国 + EU 賛同） 欧州（EU）AI 規制法の成立（8/1 発効） 第 2 回 AI ソウル・サミット「ソウル宣言」（10 か国 + EU 賛同） G7 プーリア・サミット（伊）首脳コミュニケ 中国主導の「AI 能力構築の国際協力強化」国連総会決議（A/RES/78/311） 国連 AI 諮問機関が最終報告書「人類のための AI 統治」公表 国連未来サミット「未来のための約束（Pact for the Future）」（A/RES/79/1）採択（グローバル・デジタル・コンパクトを含む）
2025	2/10～11	第 3 回 AI アクション・サミット（フランス）開催予定

AI 倫理準則（ロシア）など非法文書（ソフト・ロー）形式を採用した⁶⁾。それらに共通する具体的内容として、バイアス（偏向）を軽減・回避する公平性、AI 本体に対する信頼性や安全性、不具合の原因究明を可能にする透明性と追跡可能性（説明可能性）、AI による意図せざる行動を解除・停止させる制御可能性、そして、AI 使用に対する人間の責任性などが規定されている。特に、AI 関連の国際

6) 注4「前掲論文」12-19頁；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」『防衛法研究』46号 2022年9月 28-31頁。

文書で定型句として、「安全、安心で、信頼できる AI (safe, secure, and trustworthy AI)」が頻繁に使用される⁷⁾。

2023 年 5 月の G7 広島サミットの結果、AI 規制の在り方を議論するために立ち上げられた広島プロセスでは、その成果文書として、2023 年 10 月 30 日に「高度な AI システムを開発する組織向けの広島プロセス国際指針 (Principle)」と「高度な AI システムを開発する組織向けの広島プロセス国際行動規範 (Code of Conduct)」および同年 12 月 1 日に「全ての AI 関係者向けの広島プロセス国際指針」が採択された。それらの文書には、リスク・ベース・アプローチ (risk-based approach) が明記されている。そして、同年 12 月 6 日の G7 首脳声明において、信頼できる AI に対する「政策手段が、G7 メンバー間で異なり得ることを認識しつつ」、換言すれば、各国が現時点で採用する立場（国内法の立法化か、それに至らない原則・指針レベルか）の相違はあるとしても、初の包括的な国際的枠組みとして、上記の三つの成果文書を含む「広島 AI プロセス包括的政策枠組み (Framework)」が承認された⁸⁾。

それとは別の新たな動きとして、2023 年 11 月に英国主催の「AI 安全サミット」が採択したブレッチェリー宣言⁹⁾ と、その後継である 2024 年 5 月の韓国主催の「AI ソウル・サミット (オンライン形式)」が採択したソウル宣言¹⁰⁾ が指摘できる。ブレッチャー宣言では、AI は機会と同時に重大なリスクを孕んでいること、AI のリスクは本質的に国際的であり、国際協力を介して対処するのが最善であること、共通の懸念事項である AI の安全リスクの科学的な共通理解を構築すること、AI の安全性を確保するために各国間でリスク・ベース政策を採用することなどが規定された。ソウル宣言では、AI の安全性、革新性および包摂性は相互に関連する目標であること（2 項）、AI の安全、安心で信頼できる設計、開発、配備および利用を確保するために、リスク・ベース・アプローチに沿った AI ガバナンス枠組み間の相互運用可能性が重要であること（3 項）、そして、人間中心の AI を活用し、民主主義的価値・法の支配・人権・基本的自由およびプライバシーを保護し、AI の安全性、革新性および包摂性を促進するために国際協力を強化すること（5 項）が規定された。ソウル宣言の最後の部分で、AI 安全性サミットの第 3 回目として 2025 年にフランスで AI 行動サミット (AI Action Summit) の開催が予定された。

もう一つの新たな動きとして、グテーレス国連事務総長が 2023 年 10 月に AI のリスク、機会および

7) たとえば、G7 Italia2024, *Apulia G7 Leaders' Communiqué*, June 14 2024, <https://www.g7italy.it/wp-content/uploads/Apulia-G7-Leaders-Communique.pdf>, accessed December 14, 2024.

8) G7 Leaders' Statement, December 6, 2023, <https://www.mofa.go.jp/mofaj/files/100591757.pdf>, accessed 14 December 2024.

9) GOV.UK, *Policy paper The Bletchley Declaration by Countries Attending the AI Safety Summit, 1-2 November 2023*, 1 November 2023, <https://web.archive.org/web/20240830171000/https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023>, accessed 28 December 2024. 賛同国は 28 か国と EU であり、その中に中国も参加している。

10) *Seoul Declaration For Safe, Innovative and Inclusive AI by Participants Attending the Leaders' Session of the AI Seoul Summit, 21st MAY 2024*, <https://www.mofa.go.jp/mofaj/files/100672517.pdf>, accessed 28 December 2024.

国際的ガバナンスの検討のために AI 諮問機関 (AI Advisory Body) を立ち上げたことである。当該機関は、同年 12 月に中間報告書を公表し、2024 年 9 月に最終報告書「人類のための AI 統治」¹¹⁾ を公表した。そこでは、7 つの提言¹²⁾ が含まれ、その一つに「気候変動に関する政府間パネル (IPCC)」をモデルとした新たな専門家組織「AI 国際科学パネル (An international scientific panel on AI)」の設置が提案された。

これら二つの動きは、AI に関する具体的な条文などの規範策定には直接関連しないとしても、AI に関する共通認識の醸成や AI の実態およびその規制を議論する今後の組織的体制作りに役立ち、間接的に AI 規制の規範形成に貢献するものと言える。

(b) ハード・ロー規制

AI 規制の次の段階として、当該原則や倫理規範に基づいたより具体的かつ詳細な法文書 (ハード・ロー) 形式の条文案が審議され策定されてきた。たとえば、生成 AI を含む AI 規制法が、欧州議会により 2024 年 3 月 13 日に可決され、同年 8 月 1 日に発効するに至った¹³⁾。この欧州 AI 法は、リスク・ベース・アプローチが採用され、AI システムのリスクを 4 段階に区分し、段階ごとに対応策を規定している。具体的には、容認できないリスク (使用禁止)、高リスク (厳格な使用条件付与)、限定リスク (透明性の使用条件付与)、最小リスク (規制なく自由に利用可能) に区分される。米国も、2023 年 10 月に初めて法的拘束力のある行政措置として「AI 規制に関する大統領令」¹⁴⁾ を発した。当該大統領令は、安全性の新基準、プライバシー保護、公平性の推進、消費者・患者・学生の権利保護などを規定した¹⁵⁾。近年、日本¹⁶⁾ を含む諸国家内でも AI の国内法規制の動きが見られる。

(2) 国連内での AI 規制

(a) 米国主導の国連総会決議

国連総会は、2024 年 3 月に AI 規制に関する初めての決議「持続可能な開発のために安全、安心で

11) United Nation AI Advisory Body, *Governing AI for Humanity*, September 2024, https://www.un.org/sites/un2.un.org/files/governing_ai_for_humanity_final_report_en.pdf, accessed 28 December 2024.

12) ① AI 国際科学パネルの設置、②政府間・関係者間での年 2 回の AI ガバナンスに関する政策対話の開始、③ AI 標準の意見交換組織の設置、④能力開発ネットワークの創設、⑤ AI 格差の是正のための AI グローバル基金の設立、⑥グローバル AI データ枠組みの構築、⑦国連事務局内に AI 事務室の設置。

13) European Commission, “European Artificial Intelligence Act comes into force,” *Press Release*, Aug 1, 2024, https://ec.europa.eu/commission/presscorner/detail/en/ip_24_4123, accessed 14 December 2024.

14) The White House, *Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*, October 30, 2023, <https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/>, accessed December 14 2024.

15) JETRO「ビジネス短信：バイデン米政権、AI の安全性に関する新基準などの大統領令公表」2023 年 11 月 01 日、<https://www.jetro.go.jp/biznews/2023/11/495833ae70119dbf.html>, 2024 年 12 月 14 日閲覧。

16) 2024 年 12 月現在、日本は AI を包括的に規制する法令は存在せず、法的拘束力のないガイドラインが規定されているが、日本政府は「今後、AI に関する様々なリスクや、規格やガイドライン等のソフト・ローと法律・基準等のハード・ローに関する国際的な動向等も踏まえ、制度の在り方について検討する」と規定し、AI 法規制に向けた検討を始める方針を示した。閣議決定『統合イノベーション戦略 2024』2024 年 6 月 4 日 6 頁。

信頼できる AI システムの機会を掴む」¹⁷⁾ をコンセンサスで採択した。本決議の採択に至る背景は、以下のように説明される¹⁸⁾。2023 年末まで国連総会は技術的發展と国連の持続可能な開発目標へのその組込みに関する一連の決議を採択していたけれども、それらのいずれの決議も AI を特に取り扱っていなかった。近年、生成 AI や大規模言語モデル¹⁹⁾ の画期的な登場以降、その規制は徐々に進展しているとはいえ、AI 導入がグローバルな性質を有するために、その規制には国際的アプローチが必要となる。しかし、諸国家がその共通基盤を見つけるのが容易ではなかったという。

そのような状況の中で、国連総会が、米国主導で 120 か国が共同提案した決議案をコンセンサスで採択した。本決議はあくまで非軍事領域に限定されている（前文 6 項）が、本決議案を提出した米国の国連大使は、採択に先立ち、「この決議に結実した包摂的で建設的な対話が、例えば平和と安全や、AI の自律性の責任ある軍事利用など、他の分野での AI を巡る課題に関する今後の議論のモデルとなる」との期待を表明した²⁰⁾。

では、軍事領域での AI 利用規制に参考になるかもしれない当該国連総会決議の概要は、次の通りである。総会は、非軍事的領域に限定した「安全で (safe)、安心で (secure)、信頼できる (trustworthy) 人工知能 (AI) システム」は、「そのライフ・サイクル（事前設計、設計、開発、評価、実験、設置、使用、販売、調達、運用、廃棄）の各段階において、人間中心で (human-centric)、信頼でき (reliable)、説明可能で (explainable)、倫理的で (ethical)、包摂的 (inclusive)」であり、「人間を中心に置きながらすべての人権や基本的自由の享有を促進し保護する潜在的能力がある」と明記する（前文 6 項）。また、適切な安全装置がないかまたは国際法と両立しない方法での AI システムの不適切なもしくは悪意のある設計、開発、設置および使用は、人権および基本的自由の保護・促進・享受を損なうリスクがあると指摘する（前文 7 項）。

そのような認識の上で、国連総会は、国家間や国内の AI その他のデジタル格差を埋めること（本文 1 項）や、そのために安全で安心かつ信頼できる AI システムを推進することを決意し（同 2 項）、当該 AI システムの規制やガバナンスに関するアプローチや枠組みの策定を加盟国に奨励する（同 3 項）。全加盟国は、国際人権法を遵守して運用できないかまたは人権の享有に過度なリスクをもたらす AI システムの使用を差し控えるよう要請され（同 5 項）、さらに、各国国内の規制およびガバナンスのアプローチや枠組みを実施しながら、包摂的で公平な方法で当該 AI システムの促進が奨励され

17) Resolution adopted by the General Assembly on 21 March 2024, “Seizing the opportunities of safe, secure, and trustworthy artificial intelligence system for sustainable development,” A/RES/78/265, 1 April 2024.

18) Annika Knauer, “The First United Nations General Assembly Resolution on Artificial Intelligence,” *Blog of the European Journal of International Law*, April 2, 2024, <https://www.ejiltalk.org/the-first-united-nations-general-assembly-resolution-on-artificial-intelligence/>, accessed 15 November 2024.

19) オープン AI 社のチャット GPT、アンソロビック社のクロード、グーグル社のジェミニ、マイクロソフト社のコパイロットなど。

20) Vibhu Mishra, “General Assembly adopts landmark resolution on artificial intelligence,” *UN News*, 21 March 2024, <https://news.un.org/en/story/2024/03/1147831>, accessed 27 December 2024.

た（同6項）²¹⁾。AIシステムの進化に合わせてAIガバナンス領域の議論をする重要性も指摘された（同8項）。

（b）中国主導の国連総会決議

米中間でAIに関する国際ルール作りの主導権争いを示しているのか、2024年7月にもAI利用でのデジタル格差を解消するために中国主導の「AIの能力構築に関する国際協力強化」国連総会決議²²⁾が採択された。本決議は、前述のAI利活用に関する国連総会決議（A/RES/78/265）を踏まえて、中国がそれと異なる視点（AIの発展途上国支援に重点）から提案し、日米英仏露など約140か国が共同提案国となり、コンセンサスで採択された。本決議も、A/RES/78/265と同様に、「非軍事領域におけるAIの能力構築に関する国際協力に焦点を当てており、軍事目的でのAIの開発または使用に触れていない」（前文6項）。

その決議内容は、表題が示すように、発展途上国を含むすべての加盟国がAIの設計から使用まで「平等な機会を享受すべきであり」（同7項）、AIその他のデジタル格差を埋めるために、次のことを要請する。まず、AIのグローバル・ガバナンスの国際プロセスに発展途上国が参加し発言するのを支援する（本文1項）。AI能力構築のために発展途上国への資金提供や技術支援を増加させる（同2項）。そして、AIシステムのライフ・サイクル全般にわたって、公正、公開、包摂的かつ差別のないビジネス環境を提供する（同6項）。自発的で透明性のある協力の枠組みまたはイニシアティブを通じて、AIの能力構築に関する行動指向型の国際協力を向上させる（同13項）ことなどが要請された。

このように、中国主導の本決議は、デジタル格差の解消のためにAI能力の構築に対する国際協力に重点を置いているが、AIの前提に関して、先の米国主導決議と同じ立場を採用している。本決議においても、AIシステムは「安全で、安心で、信頼できる」ものでなければならないし、具体的には、「そ

21) その実施方法として17項目が列挙された。具体的には（a）国内の規制やガバナンスの枠組みの策定・実施の推進、（b）AIの設計・開発中にその脆弱性やリスクの国際的に相互運用可能な識別・分類・評価・実験・防止・軽減のための技術革新を促す効果的措置の奨励、（c）AIシステムの配備後にその技術的脆弱性・誤用・事故に関するエンド・ユーザーや第三者によるフィードバック・メカニズムの組込の奨励、（d）AIシステムの適切な民生利用の中核機能・能力・限界・領域に関する国民認識の向上、（e）AIシステムのライフ・サイクル全般でリスクの監視・管理メカニズム・データ保護メカニズム・影響評価の開発・実施・開示の促進、（f）AIシステムのライフ・サイクル全般で物理的保証・AIシステム保証・リスク管理など効果的な安全装置の開発・実装への投資強化、（g）信頼できるコンテンツ認証や来歴メカニズムなど効果的・アクセス可能・適用可能・国際的に相互運用可能な技術的ツール・標準・実行の開発・配備の奨励、（h）バイアス・誤用・他の危害からの個人保護のためにAIシステムの訓練・実験用の効果的に相互運用可能な枠組み・実行・標準の開発・実装の促進、（i）知的財産権の尊重のための安全装置の奨励、（j）システムの実験・評価時のプライバシーや個人データの保護、（k）エンド・ユーザーに影響を及ぼすAI決定支援システムのライフ・サイクルを通じての透明性・予測可能性・信頼性・理解可能性の促進およびAIシステムの自動決定で不利益を被る者への人間による意思決定の代替手段または効果的な救済や説明責任による人間による監視の促進、（n）AIシステムの潜在的リスクを軽減し利益を最大化するためのベスト・プラクティス、政策、アプローチに関する情報共有の強化。

22) Resolution adopted by the General Assembly on 1 July 2024, “Enhancing international cooperation on capacity-building of artificial intelligence,” A/RES/78/311, 5 July 2024.

のライフ・サイクル（事前設計、設計、開発、評価、実験、配備、使用、販売、調達、運用、廃棄）の各段階において、人間中心、信頼でき、説明可能で、倫理的で、包括的であること」を繰り返している（前文6項）。

（c）「未来のための約束」

国連は、2020年の創設75周年にあたり、グローバル・ガバナンスの強化を誓い、国連事務総長に「我々の共通の課題（our common agenda）を促進し、現在および未来の課題に対処する」ための提言を要請した²³⁾。その要請に従って、2021年に、国連事務総長は報告書「我々の共通の課題」²⁴⁾を作成し、その実行に向けてハイレベルでのマルチ・ステークホルダーによる「未来サミット（Summit of the Future）」の開催を提言した²⁵⁾。2024年9月22～23日に開催された国連未来サミットは、その成果文書として「未来のための約束」²⁶⁾ならびに付属文書Ⅰ「グローバル・デジタル・コンパクト」²⁷⁾および付属文書Ⅱ「未来世代への宣言」²⁸⁾を採択した。

デジタル協力とAIガバナンスに関する初めての包括的枠組みである「グローバル・デジタル・コンパクト」では、目標の一つとして「5. 人類の利益のためにAIの国際的ガバナンスを向上させること」が挙げられた²⁹⁾。原則の一つとして「(i) AIを含めて安全、安心、そして信頼できる新興技術（safe, secure and trustworthy emerging technologies, including artificial intelligence）」によって、開発を急速に促進する新たな機会が提供される。効果的な人間による監視（effective human oversight）を受けつつ、デジタル技術や新興技術のライフ・サイクル（事前設計、設計、開発、評価、実験、設置、使用、販売、調達、運用、廃棄）に対する責任・説明責任³⁰⁾があり、透明性があり、そして、人間中心的なアプローチ（responsible, accountable, transparent, and human-centric approach）は、我々の協力により促進される³¹⁾と明記された。さらに、「コミットメント（誓約）と行動」の「第5目標：人類の利益のための国際的AIガバナンスの向上」において「AIガバナンスに対するバランスの取れた包括的なリスク・ベース・アプローチの必要性」³²⁾が承認された。

民生用AIに関連する三つの国連総会決議は、法文書ではないが、加盟国193か国の政治的意思を

23) *Declaration on the commemoration of the seventy-fifth anniversary of the United Nations*, A/RES/75/1, 28 September 2020 (adopted on 21 September 2020), par. 20.

24) *Our Common Agenda Report of the Secretary-General*, A/75/982, 5 August 2021.

25) *Ibid.*, par. 103.

26) *The Pact for the Future*, A/RES/79/1, 22 September 2024, pp. 1–37. ここで用いられている「Pact」は、「協定」と訳す文献も散見されるが、本稿では、それが国際法上の義務として規定されていないことから、より一般的な「約束」と訳す。

27) *Annest I Global Digital Compact*, *ibid.*, pp. 37–52.

28) *Annex II Declaration on Future Generations*, *ibid.*, pp. 52–56.

29) *Ibid.*, p. 38, par. 7.

30) 責任（responsibility）は与えられた任務を行う義務を言い（用語法として、「責任がある」）、説明責任（accountability）は与えられた任務の結果に対する答責性を言う（用語法として、「責任を負う」）。

31) *Ibid.*, p. 39, par. 8.

32) *Ibid.*, p. 48, par. 50.

表明した文書であり、AIの利用および規制に関する今後の方向性をより明確に示唆するものであると言える。今後、民生用のAI規制は、国際的にも国内的にも、人間中心アプローチや禁止・規制のリスク・ベース・アプローチを共通基盤とし、政治的な原則やコミットメント（ソフト・ロー）を醸成し、次の発展段階として各国の国内法や国際条約（ハード・ロー）へと展開していくと予測される。

3. 軍事用 AI の規制

現在、AIの軍事利用は、①情報収集、監視および偵察、②兵站、③サイバー戦、④情報戦およびディープ・フェイク、⑤指揮・統制、⑥半自律型および自律型ヴィークル（車両、航空機・ドローン、艦船）、⑦自律型致死兵器システム（LAWS：Lethal Autonomous Weapons Systems）など、様々な領域で行われている³³⁾。本稿では、その中の⑤指揮・統制部門のAI利用を取り上げる。

(1) AI 意思決定支援システムの実態

(a) ウクライナ軍の AI システム

AIによる意思決定支援システム（AI-Decision Support System, AI-DSS）は、すでに軍民双方の領域で活用されている。民生用として、医学部門で早期胃がんの発見のためのAI搭載内視鏡画像診断支援ソフトウェア³⁴⁾や、スポーツ部門で器械体操の複雑な動きを分析し審判の判定を支援するために国際体操連盟の採用している判定支援システム（Judging Support System）³⁵⁾がある。

軍事用 AI-DSS として³⁶⁾、現在進行中のロシア・ウクライナ戦争においてウクライナ軍は、米国企業バランティア・テクノロジー社のデータ分析 AI システム「ゴッサム（Gotham）」³⁷⁾を活用している。当該システムは、偵察衛星・偵察機・一般人の SNS からの敵情報・被害状況など膨大な情報を収集・分析し、1日に300の攻撃目標（敵の位置・部隊規模）を特定し、自軍の位置・兵力・保有兵器の状況認識や敵の進軍方向の予測を含めて、効率的な攻撃方法を立案・提言する。軍司令官は、複数の軍事作戦案の中から最適解を選択し、前線の将校に命令を送信する。このように、同システムは、

33) Kelly M. Saler, “Artificial Intelligence and National Security,” *Congressional Research Service Report R45178*, Updated August 26, 2020, pp. 9–15.

34) AI メディカルサービス社の gastroAI-modelG。これは、20 万本の内視鏡検査の画像や動画を AI に読み込ませ、早期胃がんの画像にどの程度似ているかを数値化し、医師のがん検診を支援する。「AI メディカルサービス、早期胃がんに特化した AI 搭載の内視鏡画像診断支援ソフトウェアを発売」2024/03/04. <https://www.ai-ms.com/news/product/20240304>、2024 年 12 月 14 日閲覧。

35) 富士通の Human Motion Analytics。「国際体操連盟、富士通の Judging Support System を全 10 種目で利用開始」2023 年 10 月 5 日、<https://pr.fujitsu.com/jp/news/2023/10/5-1.html>、2024 年 12 月 14 日閲覧。

36) 日本も「AI により行動方針を分析し、指揮官の意思決定を支援する技術を装備品に反映するための研究を行う」と表明している。閣議決定『防衛力整備計画』2022 年 12 月 16 日 24 頁。

37) TBS NEWS DIG「こっそり歴史を変えた〜ウクライナ善戦の裏でアメリカ民間企業“魔法の力”【報道 1930】」2023 年 2 月 8 日、<https://newsdig.tbs.co.jp/articles/-/318447?page=3>、2024 年 12 月 14 日閲覧。

敵の位置情報の分析およびそれへの効果的な攻撃方法を軍の指揮官に提言する³⁸⁾。

(b) イスラエル軍の AI システム

2023 年 10 月 7 日から始まったイスラエル・ハマス戦争でのイスラエル軍は、対物破壊用標的をリスト化する AI システム「ハブソラ (Habsora, Gospel, 福音)」を活用している。それは、テロリストのアセットである標的（個人住宅、公共施設、インフラ、高層ビルなど）を power target（有力標的）と定義し、それらをリスト化する AI 標的設定システムである³⁹⁾。さらに、イスラエル軍は、対人殺傷用標的をリスト化する AI システム「ラベンダー (Lavender)」も活用している。それは、SNS、ドローン映像、携帯電話のデータの監視からハマスやパレスチナ・イスラム聖戦の戦闘員と「看做す」人物のプロフィールを作成するシステムであり、AI は殺害標的である戦闘員の可能性の程度を 1 から 100 までで評価する⁴⁰⁾。

軍事用 AI の利用によって、イスラエル・ハマス戦争での文民被害数が一層増加したと一般的に認識されている。その原因として、①ラベンダーによる人的標的化の 10% に誤りがあること、②人的標的を上級戦闘要員から下級戦闘要員にまで拡大したこと、③人間の爆撃許可のための標的確認は 1 人に 20 秒しかかけなかったこと、④国際法上の比例原則に関する巻き添え損害の許容範囲を下級戦闘要員 1 人に文民 15～20 人、上級戦闘員 1 人に文民 100 人に拡大したことなどが指摘される⁴¹⁾。このように、イスラエル軍の指揮官は、AI による分析の根拠となる生の情報データを十分に確認しないまま、ハブソラやラベンダーが作成する破壊・殺傷リストを疑うことなく全面的に信用し、指揮官自身が AI システムの認定に検討時間をかけずに単にゴム印を押すだけになっていたと批判された。これに対して、イスラエル軍は、ハブソラもラベンダーも情報分析者により使用されていることを認めるが、標的認定過程は広範な標的化過程の一部であり、標的を同定する情報将校と標的の攻撃を決定する将校は別である点などから、上記の報道や法的評価に反論している⁴²⁾。

イスラエル軍による AI-DSS の使用についての国際法的評価や AI-DSS それ自体の法的問題点の分析は別稿に譲るとして、ハブソラやラベンダーが実際に軍事領域で活用されている事実やそれが多くの

38) ウクライナは、ロシアの全面侵攻に際して、国内の AI 規制（制限）を防衛部門で適用外とし、防衛部門での責任ある AI 利用の効果的履行は国際法レベルでのみ可能であるとの立場を取っている。Ministry of Digital Transformation of Ukraine, *White Paper on Artificial Intelligence Regulation in Ukraine: Vision of the Ministry of Digital Transformation of Ukraine version for Consultation*, June 2024, p. 10.

39) Yuval Abraham, “‘A mass assassination factory’: Inside Israel’s calculated bombing of Gaza,” *+972 Magazine*, November 30, 2023, <https://www.972mag.com/mass-assassination-factory-israel-calculated-bombing-gaza/>, accessed 20 December 2024.

40) Yuval Abraham, “‘Lavender’: The AI machine directing Israel’s bombing spree in Gaza,” *+972 Magazine*, April 3, 2024, <https://www.972mag.com/lavender-ai-israeli-army-gaza/>, accessed 20 December 2024. それに加えて、標的が自宅に戻ったことを追跡する AI ソフトウェア「ノパバはどこ？ (Where’s Daddy?)」も使用されたという。

41) *Ibid.*

42) Israel Defense Forces, “The IDF’s use of Data Technologies in Intelligence Processing,” *IDF Press Releases Regarding the Hamas-Israel War*, June 18, 2024, <https://www.idf.il/en/mini-sites/idf-press-releases-israel-at-war/june-24-pr/the-idf-s-use-of-data-technologies-in-intelligence-processing/>, accessed 30 December 2024.

【年表2：軍事用 AI 規制動向】（著者作成）

年	月 日	軍事用 AI 関連の国家、国際機構、国際会議
2020	2/24	米国防総省「AI 倫理原則」採択
2021	10/22	NATO 国防相理事会「AI 戦略（防衛 AI 責任ある使用原則）」合意
2023	2/15～16 11/9 11/13 12/4～5	オランダ・韓国共催 REAIM サミット（ハーグ） 「軍事領域での責任ある AI 行動要請」（60 か国賛同） 米国提唱の「AI・自律性の責任ある軍事利用政治宣言」構想公表 米国作成「AI・自律性の責任ある軍事利用政治宣言」（51 か国賛同） 米国提唱の AI・自律性軍事利用政治宣言の初会合（ニューヨーク） 国連・韓国共催「第22回軍縮・不拡散合同会議」
2024	2/26～27 3/19～20 7/2 9/9～10 11/6 12/24	REAIM 地域（アジア）協議（シンガポール） AI・自律性の責任ある軍事利用政治宣言第1回全体会議 日本「防衛省 AI 活用推進基本方針」発表 第2回 REAIM サミット（ソウル）「REAIM 行動ブループリント」（61 か国賛同） 国連総会第1委員会「軍事領域での AI」決議採択（165:2:6） 国連総会「軍事領域での AI」決議採択（A/RES/79/239, 159:2:5）
2025	?	国連総会「軍事領域での AI」議論予定

国際法の問題点を抱えている実態が公になった。他方、ウクライナ軍使用のゴッサム・システムに対する批判があまり聞こえてこないのは、戦況が詳細に報道されないこともあるが、入力データの信憑性、標的判定基準の厳格性、指揮官による攻撃命令手続きの慎重性などが存在したためなのかもしれない⁴³⁾。

（2）軍事用 AI の規制文書

（a）REAIM 2023・2024

前述の如く、AI による意思決定支援システムは、すでに実戦使用されている。それに伴い、軍事領域全般に関する AI 利用が国際会議の議題として取り上げられるようになった。【年表2：軍事用 AI 規制動向】が示すように、米軍や NATO 軍といった一部諸国の軍隊用の倫理原則はあったが、2023 年 2 月に開催されたオランダ・韓国共同主催の「軍事領域での責任ある AI サミット（REAIM 2023）」⁴⁴⁾が、国際会議で軍事用 AI 問題を取り上げる嚆矢となった。当該会議で中国を含む 60 か国（後に 57 か国）が賛同し採択された「REAIM Call to Action（REAIM 行動要請）」⁴⁵⁾は、非法文書形式で、AI の軍事利

43) もともと、ウクライナ軍は、越境攻撃してきたロシア軍の戦闘車両や戦闘員の識別が比較的容易であったために、AI-DSS を有効に活用できたように思える。他方、イスラエル軍は、ヒト（戦闘員か文民か）の認定がそもそも困難であり、ハマス戦闘員と文民が混在するガザ地区での市街戦であったために、多数の文民被害を発生させたと思われる。イスラエル軍による AI-DSS 利用は、市街戦であるがゆえに、国際法を遵守するためにもより厳格な標的認定基準が設定され、慎重に運用されなければならなかったと言える。

44) 注 2 岩本「自律型致死兵器システム（LAWS）規制議論の動向と今後の展望—2023 年度—」41-43 頁。

45) REAIM 行動要請の翻訳として、注 44「前掲論文」58-62 頁。

用に関する現状認識および国家が採るべき責任ある AI の行動を提示した。

その行動要請には、まず、軍事領域の AI 活用が国際法義務に従っていること（前文 5 項）、そして、軍事領域での AI 活用の潜在的な能力を認めていること（本文 2 項）、軍事領域の AI 活用の利益・リスク・限界の知識やリテラシーを深める必要があること（同 4 項）、人間が軍事領域での AI 利用時の決定の責任を負うこと（同 6 項）、AI 利用の適切な安全措置と人間による監視の確保が重要であること（同 12 項）、AI の軍事利用に関するグッド・プラクティスや教訓の交換が奨励されること（同 14 項）などが規定された。

2024 年 9 月には、第 2 回 REAIM が韓国・ソウルで開催された。その結果、共同主催国の韓国が提案した「REAIM 行動ブループリント（青写真）」（添付翻訳資料①参照）に米・英・日・豪を含む 61 か国が賛同した⁴⁶⁾。当該ブループリントは、前文、本文 20 項（AI が国際の平和と安全に与える影響、軍事領域での責任ある AI の履行、軍事領域での将来の AI ガバナンス構想の 3 項目に区分）、そして、後文で構成される。

まず、AI が軍事作戦、指揮統制、諜報・監視・偵察（ISR）活動、訓練、情報管理、兵站支援を含むあらゆる軍事領域で活用可能であると認識されている（前文 1 項）。AI 活用にリスクが伴うので、国家を含むすべての利害関係者に以下のように行動するように推奨する。

「AI が国際の平和と安全に与える影響」では、軍事領域での AI 活用が、国際の平和・安全・安定化を害しない方法で開発、配備、使用されるべきであることを確認する（本文 1 項）。軍事領域での AI 活用により、事態認識、精度、効率性が向上し、国際人道法の履行が強化され、文民保護を支援する可能性があることを認識する（同 2 項）。他方で、それには、予見不可能なリスクをもたらす可能性があり、特に、設計上の欠陥、データ、アルゴリズムその他のバイアスによる意図せざる結果、技術的な誤用や悪意のある使用によって軍拡競争、誤算、事態の悪化の原因となる可能性があることも認識する（同 3 項）。特に注意すべき軍事領域での AI 活用は、AI 対応兵器、戦闘作戦のための AI 対応意思決定支援システムなどがあることも認識する（同 4 項）。核兵器の使用に関する至高の決定を通知・実行するための決定的なすべての行動に人間の制御と関与を維持することが特に重大であることを強調する（同 5 項）。

「軍事領域での責任ある AI の履行」では、AI 活用は、国際法（国連憲章、国際人道法、国際人権法）に従って開発、配備、使用されなければならないことを確認する（同 7 項）。軍事領域での責任ある AI 活用を確保するために認識すべき点として、網羅的ではないとしても、(a) AI 活用は倫理的で人間中心的であること、(b) 軍事領域での AI 機能は適用可能な国内法や国際法に従って活用されること、(c) 人間が軍事領域での AI 活用の使用と影響について責任と説明責任を有し、それらを機械に

46) REAIM Blueprint for Action, September 9–10, 2024, file:///C:/Users/siwam/Downloads/REAIM_Blueprint%20for%20Action.pdf, accessed 23 December 2024.

移譲できないこと、(d) AI 活用の信頼を確保するために、データ、アルゴリズムその他のバイアスによる誤作動や意図せざる結果のリスクを軽減する安全装置の確立が必要であること、(e) 軍事領域での AI の開発、配備、使用に人間の適切な関与を維持する必要があること、(f) 関係者が軍事領域での AI 機能による成果を理解・説明・信頼できるように、軍事領域での AI の説明可能性や追跡可能性が必要であること（同 9 項）を承認する。効果的な法的審査手続き、信頼醸成措置および適切なリスク軽減措置の開発ならびにグッド・プラクティスや教訓などの情報交換を奨励する（同 11 項）。そして、軍事領域での責任ある AI の取組は、AI 技術の研究・開発・実験・革新と並行して行われ、後者を妨げないことを強調する（同 12 項）。

「軍事領域での将来の AI ガバナンス構想」では、軍事領域での AI ガバナンスを議論する際に、可能な AI 活用とその潜在的影響（利点とリスク）の共通理解を促進すべきであることを認識する（同 13 項）。国家や地域は、軍事領域での AI 機能の統合化の多様な段階にあり、多様な安全保障環境において、そして、安全保障上の懸念事項が多様であることに留意する（同 14 項）。軍事領域での AI 活用ではデータが重要な役割を果たしているため、適切なデータ・ガバナンス・メカニズムを一層議論する必要があることに留意する（同 17 項）。REAIM サミット、AI・自律性の責任ある軍事利用政治宣言、特定通常兵器条約（CCW：Convention on Certain Conventional Weapons）の自律型致死兵器システム（LAWS）に関する政府専門家グループ（GGE：Group of Governmental Experts）、国連総会決議など複数の取組を承認する（同 19 項）。そして、後文で、諸国家にこの「REAIM 行動ブループリント」への参加を勧誘している。

2024 年の「REAIM 行動ブループリント」は、2023 年の「REAIM 行動要請」と同じ認識や方向性で書かれているが、行動の内容がより詳細かつ明確になってきた。90 か国以上が会議に参加した中で豪、仏、独、伊、日、韓、英、米を含む 61 か国が同文書を支持する一方、中、露は支持しなかった。

(b) 米国の AI 軍事利用政治宣言

米国は、上記の REAIM 2023 の国際会議において、REAIM 行動要請とは別に、「AI・自律性の責任ある軍事利用政治宣言」イニシアティブ（構想）⁴⁷⁾ を提唱した。本政治宣言は、REAIM 行動要請と同様に非法文書形式で、AI の責任ある軍事利用に関する国際的合意を形成し、その開発・配備・使用の指針となり、そして、ベスト・プラクティス（好事例）の情報交換の基盤となることを目指した。

その政治宣言では、A 項から J 項までの 10 項目の措置⁴⁸⁾ が推奨された。それは、武力紛争での AI 利用に国際法を遵守させること（序文）、軍事 AI 能力を利用するための適切な措置（法的審査、法違反報告制度、法教育）の実施（B 項）、意図せざるバイアスを最小化させること（D 項）、システムの

47) 注 44「前掲論文」43-47 頁。

48) AI 軍事利用政治宣言の翻訳として、注 44「前掲論文」62-65 頁。

能力と限界を十分理解させる必要があることおよび自動化バイアス（automation bias）⁴⁹⁾を軽減化すること（G 項）、リスクや機能不全を考慮した実験や保証を厳格化すること（I 項）、機能不全リスクを軽減するための安全措置（機能不全の検知・回避解除・不活性化）を取ることに（J 項）である。

2024 年 3 月には、賛同国が 54 か国となった AI 軍事利用政治宣言の第 1 回年次会合が、米国において開催された。そして、軍事領域での AI の責任ある開発、配備および使用を確保するために各国が実施すべき 10 の措置についてより具体的な実施策の検討に向けた作業部会の設置が確認された⁵⁰⁾。

（c）国連総会決議

オランダおよび韓国は、共同主催国として前述の REAIM2023 および 2024 を開催し、「REAIM 行動要請」や「REAIM ブループリント」を採択してきたように、軍事領域での責任ある AI 国際規範の策定を主導してきた。両国は、次の段階として、当該国際規範の作成方針を国連総会第 1 委員会に持ち込み、「軍事領域での AI および国際の平和と安全に対するその含意」決議案⁵¹⁾を提案した。当該決議案は、2024 年 11 月 6 日に同委員会において圧倒的多数で採択された⁵²⁾。それを受けた国連総会本会議は、総会決議（A/RES/79/239）⁵³⁾を採択した。

前述の如くすでに米国および中国が主導した民生領域での AI に関する決議が 2 本採択されたが、軍事領域での AI に焦点を当てた国連総会決議は今回が初めてとなった。

民生領域での AI を除く（前文 3 項）本決議では、軍事領域での AI に国際法が適用されることの確認（本文 1 項）、先進国と発展途上国間の軍事領域での AI 格差を是正する必要性（同 6 項）、国連事務総長が軍事領域（特に自律型致死兵器システム以外の領域に焦点を当てて）での AI 活用による機会と課題（利点とリスク）に関する加盟国の見解を纏めた報告書を作成し国連総会第 80 会期に提出すること（同 7 項）、そして、第 80 会期議題に「軍事領域での AI および国際の平和と安全に対するその含意」を挙げることに（同 9 項）が決定された。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。

このように、民生用 AI の場合と同様に、軍事領域での AI に関する議論の主戦場も、有志国限定の

49) それは、人間が自動化された意思決定システムからの提案を優先し、自動化なしで作成された矛盾した情報を、たとえそれが正しい場合でも無視する傾向を指す。

50) 外務省 HP「AI と自律性の責任ある軍事利用に関する政治宣言：第 1 回年次会合の開催（2024 年 3 月）」2024 年 3 月 29 日、https://www.mofa.go.jp/mofaj/gaiko/arms/pagew_000001_00477.html, 2024 年 12 月 21 日閲覧。

51) Artificial intelligence in the military domain and its implications for international peace and security, A/C.1/79/L.43, 16 October 2024.

52) 165 か国賛成、2 か国反対、6 か国棄権。United Nations Meeting Coverage and Press Releases, First Committee, “Fourteen New Drafts, Including on Implications of Artificial Intelligence in Military Domain, Approved in First Committee by 34 Votes,” 6 November 2024, GA/DIS/3757, <https://press.un.org/en/2024/gadis3757.doc.htm>, accessed 28 December 2024.

53) Resolution adopted by General Assembly on 24 December 2024, *Artificial intelligence in the military domain and its implications for international peace and security*, A/RES/79/239, 31 December 2024. 159 か国賛成、2 か国反対、5 か国棄権。

国際会議（REAIM）から最も普遍的フォーラムであり、多数決制を採用する国連総会へと移行しつつある。その傾向が、REAIMで醸成された軍事用AI規制理念に賛同する国家の増加に連動して、軍事領域でのAI利用に関する共通認識が急速に形成されつつある。

4. AI 自律兵器の規制

近年のLAWS規制議論は、【年表3：LAWS/AWS規制動向】が示すように、CCW枠内外で活発に行われている。

(1) CCW 内での AI 自律兵器規制

(a) GGE 第1会期

CCW枠内でのLAWSに関するGGE第1会期が、2024年3月4日から8日まで開催された。そこでの議論の概要は、議長要約⁵⁴⁾によれば、以下の通りである。まず、第1会期は、三つのトピック、①特徴付けと定義、②国際人道法（IHL：International Humanitarian Law）の適用（IHLと人間の制御、判断および/または関与の概念との関係を含む）、③リスク軽減と信頼醸成措置を取り扱った。

①特徴付けと定義に関して、LAWSの正確な定義が必要という意見や、一般的な特徴付けで十分という意見、さらに、禁止や規制を含む将来の措置を議論するために定義も特徴付けも特に必要ないという意見に分かれた（要約6項）。LAWSの「Lethal（致死）」という文言を継続する必要があるか否かでも意見が対立した（同7項）。LAWSの「autonomous（自律型）」に関して、自律性はスペクトル上に存在すること、「完全（full）」自律型兵器システムや「部分的（partial）」自律兵器、それらの中間に位置する「半自律型（semi-autonomous）」兵器システムの理解を深める必要性があるなどの意見が聞かれた。多くの代表は、「完全」自律型兵器システムの禁止と「部分的」自律型兵器システムの使用規制という考え方を推奨した。この考え方の根底には、「完全」自律型兵器システムを使用した場合、人間の関与レベルが不十分となりIHLを遵守できなくなるという前提がある。それに対して、自律性の異なるレベルを明確に区別することが困難であるとの見解も見られた（同8項）。

②IHLの適用に関して、ほとんどの代表は、現行のIHLがLAWSに完全に適用されることを再確認したが、新たな法が必要か否かについて見解が分かれた。LAWSの課題やリスクに現行のIHLで十分に対応可能と考える一部の代表は新たな文書は「不要」とであると考えてるが、他の代表は現行のIHLの適用方法を指定したり新たな法を創設したりするために新たな文書が「有益」と考えた（同

54) Chair's summary-First 2024 session of the GGE on LAWS Submitted by the Chairperson, CCW/GGE.1/2024/WP.6, 3 April 2024. 本要約は、GGE第1会期中の作業を理解するために議長独自の権限で作成され、GGE会合に付託された。

【年表 3：LAWS/AWS 規制動向】（著者作成）

年	月 日	LAWS・AWS 関連の CCW、UN、地域会議
2022	10/21	国連総会第 1 委員会での 70 か国共同声明
2023	2/20	CCW・GGE オンライン非公式協議
	2/23～24	コスタリカ主催の CELAC 地域会議：33 か国の「ベレン共同声明」
	3/6～10	CCW・GGE 第 1 会期（5 つの作業文書の提出）
	4/20	CCW・GGE 第 1 回会期間オンライン非公式協議
	4/25～26	ルクセンブルグ主催国際会議
	5/2	CCW・GGE 第 2 回会期間オンライン非公式協議
	5/15～19	CCW・GGE 第 2 会期（GGE 報告書 CCW/GGE.1/2023/2）
	7/22	事務総長「我々の共通の課題：政策概要 9 平和のための新課題」
	9/5～6	トリニダード・トバゴ主催地域会議（CARICOM 宣言）15 か国・地域
	10/5	国連事務総長と ICRC 総裁の共同要請（2026 年までに法文書）
	11/1	国連総会第 1 委員会で LAWS 決議採択（164:5:8）
	11/15～17	CCW 締約国会議
	12/22	国連総会で LAWS 決議採択（152:4:11）
2024	3/4～8	CCW・GGE 第 1 会期
	4/17～18	ECOWAS「自律兵器システムの平和と安全地域会議」コミュニケ
	4/29～30	AWS ウィーン会議
	5/7	GGE 議長と GGE とのオンライン非公式協議（5/17, 6/7 も実施）
	5/24	国連事務総長の LAWS に関する問合せに対する日本の回答書付託
	6/6	GGE 議長とオブザーバーとのオンライン非公式協議
	8/26～30	CCW・GGE 第 2 会期
	7/1	事務総長の LAWS 報告書を国連総会に提出
	11/5	国連総会第 1 委員会で LAWS 決議採択（161:3:13）
	11/13～15 12/2	CCW 締約国会議 国連総会で LAWS 決議採択（A/RES/79/62, 166:3:15）
2025	3/3～7	CCW・GGE 第 1 会期予定
	9/1～5	CCW・GGE 第 2 会期予定（可能な限り年末までに報告書作成）
	11/12～14	CCW 締約国会議予定
	？	国連総会の LAWS 公開非公式協議開催予定
2026		CCW・GGE の 10 日間開催予定
		CCW 第 7 回再検討会議予定

11 項）。また、LAWS はどのような要素や特性があれば IHL と両立し得ないと判断するかという方法に関して、幾つかの代表は、（意義ある =meaningful）人間の制御、判断および/または関与（human control, judgement and/or involvement）が IHL の遵守に不可欠であると発言した。必要な制御レベルは、

運用環境を含む幾つかの考慮事項に依存するという意見があり、人間の制御が「名ばかり (nominal)」であってはならないと明記すべきで、人間の制御の代わりに人間の「行為・営み」(human “agency”)の使用を提案する代表もいた(同13項)。また、IHLの遵守が確保される時点に関して、兵器システムのユーザーの時点とする意見がある一方で、兵器システム自体がIHLに準拠して使用できるように設計される時点とする意見も主張された(同14項)。

③リスク軽減と信頼醸成措置に関して、ほとんどの代表は、どのような種類のリスクがあるのかを明確に特定し、軽減措置を通じてそれらに包括的に対処する必要があると主張した。ある者は、人道的・倫理的、安全保障上、地政学的、技術的、システム・セキュリティ上のリスクなど幾つかの分類を強調した。対処困難なアルゴリズムのバイアスも指摘された(16項)。多くの代表は、法的審査などを介してリスク管理が可能であると述べた。また、リスクを可能な限り最小限に抑制するためにも、兵器システムのライフ・サイクル全般を通じての人間の関与や人間の制御に言及した。他方、兵器システムの精度を高め、人間のオペレーターによる間違いを減らすことで軍隊がIHLの履行や遵守を支援するなどのLAWSの利点も議論すべきという見解も見られた(17項)。幾つかの代表は、当該措置を異なる分類に区分することが有益であり、これらをLAWS関連文書の技術的附属書に含める案が示唆された。また、ほとんどの代表は、リスク軽減や信頼醸成措置は禁止や規制を補完するものと捉えていた(18項)。

以上、三つのトピックについての議論は、従来の論点の繰り返しが散見され、LAWS規制問題が取り上げられた2014年から10年以上たっても十分に各論点が収斂していない事実が明確化された。

(b) GGE 第2会期

GGE 第2会期は8月26日から30日まで開催された。そこでの議論の概要は、議長要約⁵⁵⁾によれば、以下の通りである。まず、第2会期では、議長の提案文書「ローリング・テキスト」⁵⁶⁾を基に、四つのトピック(①特徴付けと定義、②IHLの適用、③リスク軽減と信頼醸成措置、④説明責任と責任)が検討された。

一般討議では、GGEがLAWS議論の適切なフォーラムとして中心的な役割を果たすことに幅広い支持が表明された。また、多くの代表は、CCWの下でLAWSに関する法典化作業の国連事務総長の呼び掛けを歓迎するとともに、LAWSに関するCCW第6議定書交渉を要請した(要約6項)。

ローリング・テキスト上の「特徴付けと定義」に関して、自律兵器システムの機能の理解、「致死(lethal)」の用語の使用および「名ばかりの人間の入力(nominal human input)」について多様な見解が表明された(同9～13項)。特徴付けに関するローリング・テキスト上ですでに規定されている、起

55) *Chair's summary-Second 2024 session of the GGE on LAWS Submitted by the Chairperson*, CCW/GGE.1/2024/WP.11, 2 October 2024. 本要約も、GGE 第2会期中の作業を理解するために議長独自の権限で作成され付託された。

56) GGE on LAWS Rolling text, status date: 26 July 2024, https://reachingcriticalwill.org/images/documents/Disarmament-fora/ccw/2024/gge/documents/Rolling_text_draft.pdf, accessed 28 December 2024.

動後に人間のオペレーターによる更なる介入なく標的を「選択し交戦する（select and engage）」に「識別（identification）」を追加する議論や、自律機能のリストが累積的か否かの議論も見られた。さらに、「オペレーター」を「ユーザー」または「人間のユーザー」に置き換える意見は、人間の関与をより明確にすることから、相当な支持があった（同 14 項）。

「IHL の適用」に関して、意見の大幅な収斂が見られたという。若干の意見の相違は、「制御（control）」の前に「人間の（human）」を入れるか否か、「人間の判断（human judgement）」を「人間の制御（human control）」に置き換えるか否か（置き換える意見が相当の支持を得ている）であった。さらに、マルテンス条項の追加や IHL に並行しての国際人権法の考慮が強調された（同 17 項）。禁止と規制に関して、ローリング・テキストは、LAWS を二つのタイプに区分した「2 層アプローチ（two-tier approach）」を採用している。IHL と両立し得ない LAWS は禁止され、IHL と両立し得る LAWS は規制される。本アプローチは幅広い支持を受けたという（同 18 項）。LAWS における AI 使用が IHL の遵守にもたらす課題に関して、LAWS の予測可能性、信頼性、追跡可能性、説明可能性の諸概念が当該懸念の対処ツールとして役立つという意見がある一方で、当該概念は IHL に固有のものではないので、テキストから削除すべきという意見もある。ある代表は、以前合意された用語として、「標的の種類」、「LAWS の運用期間、地理的範囲、規模」や「交戦回数」を制限するなど中核概念の具体的な制限について決定する必要性が強調された（同 20 項）。また、幾つかの代表は、対人 LAWS⁵⁷⁾ を禁止するより強硬な文言の必要性を主張したという（同 21 項）。

「リスク軽減と信頼醸成措置」に関して、多くの代表が、LAWS に関する国内の法的審査の必要性に関する文言を歓迎するとともに、ジュネーヴ諸条約第 I 追加議定書 36 条下の法的審査義務に沿った具体的な文言の挿入を示唆した（同 22 項）。もっとも、IHL 上の法的審査と能力の技術審査は異なるとの指摘があった（同 23 項）。

「説明責任と責任」に関して、説明責任を確保する幾つかの措置が国内レベルで実施されなければならないことが強調された（同 26 項）。さらに、IHL 違反の単なる「報告」を超えて説明責任を確保するための効果的なフォロー・アップの実施の必要性も指摘された（同 27 項）。

そして、最後に、議長は、今までの議論を基に、第 2 会期の要約とローリング・テキストの最新版を作成し共有すると述べた（同 29 項）。

2024 年 7 月 26 日付のローリング・テキストは GGE 第 1 会期および第 2 会期の議論を参考にしながら、改訂版ローリング・テキスト（2024 年 11 月 8 日付⁵⁸⁾）が公開された（添付翻訳資料②参照）。

57) この発言は、LAWS に対人用と対物用があることを示唆しているのか、LAWS 自体が致死性であるがゆえに対人用ではないのか、「Anti-personnel LAWS」の発言は自己矛盾のように思える。

58) GGE on LAWS Rolling text, status date: 8 November 2024, <https://reachingcriticalwill.org/images/documents/Disarmament-fora/ccw/2024/gge/documents/rolling-text-nov24.pdf>, accessed 29 December 2024.

(c) CCW 締約国会議

2024 年 11 月に開催された CCW 締約国会議の最終報告書は、2023 年の最終報告書（CCW/MSP/2023/7）と同様に、GGE ができる限り早期に、願わくは 2025 年末までに作業を完了し、その報告書を第 7 回 CCW 再検討会議に付託するよう決定された⁵⁹⁾。

そこでの討議⁶⁰⁾では、幾人かの代表が、次年度の作業を見据えて、前述の改訂版ローリング・テキスト（2024 年 11 月 8 日付）に謝意を表明した。また、その前の週に国連総会第 1 委員会で採択された決議で 2025 年に LAWS に関する非公式協議が決定された点に関して、多くの国家が当該決議を支持するとともに、オーストリアは、その非公式協議が GGE の作業を補完し、それらの努力に弾みや包括性を加えると強調した。他方、豪・中・仏・印・イスラエル・韓・土・英・米などは、国連での LAWS 議論を無条件に支持するのではなく、CCW が AWS 議論の「最も」適切なフォーラムであると牽制的な発言をした。軍事領域での責任ある AI の使用に関して、イスラエルが米国の AI 軍事利用政治宣言を支持したという。

また、パレスチナは、ガザ地区での多数の文民殺害の原因がハブソラやラベンダーといった AI システムであると強調した。他方、イスラエルは、自律的に標的を選択し標的と交戦する AI システムを使用しておらず、標的の識別に使用している情報に基づく手続きは人間によって処理されていると反論した。

このように、CCW 締約国会議での LAWS 議論は、CCW 外（国連総会）での LAWS 議論や LAWS よりも広範な軍事領域での AI 活用問題が取り上げられ、それらから大いに影響を受けている。

(2) 国連内での AI 自律兵器規制

2023 年の国連総会は、LAWS に関する初の決議（A/RES/78/241）を採択した⁶¹⁾。国連事務総長は 2024 年 7 月に、当該決議に従って、加盟国、オブザーバー国家、国際機関、市民社会などに LAWS に関する見解を聴取しまとめた報告書（A/79/88）⁶²⁾を 79 会期の国連総会に付託した。事務総長は、その報告書において、国連総会が LAWS 問題を継続して考察することを奨励した⁶³⁾。というのも、国連総会は、ほぼ普遍的な加盟国数を有し広範な事項を取り上げていることから、包括的な討議の場であり、

59) *Final Report (Advance Version)*, CCW/MSP/2024/7, 2 December 2024, par. 22. 日程に関して、GGE 第 1 会期（2025 年 3 月 3～7 日）、同第 2 期（同年 9 月 1～5 日）、CCW 締約国会議（同年 11 月 12～14 日）が決定された（*Ibid.*, par. 27）。

60) Laura Varella, “CCW Meeting Made Informal Due to Russia’s Objections to Civil Society,” *CCW Report*, Vol. 12 No. 4, 21 November 2024, 2-4, <https://reachingcriticalwill.org/images/documents/Disarmament-fora/ccw/2024/hcp/reports/CCWR12.4.pdf>, accessed 29 December 2024.

61) 注 44「前掲論文」53-54 頁。初の LAWS 国連総会決議の前兆として、決議として採択されなかったが、2022 年 10 月の国連総会第 1 委員会での 70 か国共同声明があった。

62) *Lethal autonomous weapons systems Report of the Secretary-General*, A/79/88, 1 July 2024.

63) *Ibid.*, par. 92.

LAWS と国際の平和および安全の他の問題との関係性を検討する場として相応しいと考えられたからである。

加えて、当該決議が 79 会期の暫定議題に LAWS を挙げることを決定した（本文 4 項）ために、2024 年に LAWS 問題が国連総会第 1 委員会で議論された。そして、同委員会は、2024 年 11 月 5 日に、以下の内容を含む AWS 決議案⁶⁴⁾ を採択した⁶⁵⁾。それを受けた国連総会本会議は、自律型致死兵器システム（LAWS）に関する総会決議⁶⁶⁾ を 2024 年 12 月 2 日に採択した。総会は、国際法が自律兵器システム（AWS）に適用されることを確認し（前文 2 項）、軍事領域での新興技術の活用が深刻な課題と懸念を生じさせていること（同 3 項）や、軍事領域での AI 活用の影響を包括的に取り扱うことおよび LAWS の議論との補完性を確保することの重要性（同 8 項）に留意している。また、事務総長が 2 層アプローチに沿って自律兵器システムの禁止と規制を規定する法文書の交渉を早急に締結するよう繰り返し要請していること（同 13 項）に注目し、責任や説明責任を確保し国家が国際法を遵守するための武力行使におけるヒトの役割の重要性（同 14 項）を強調している。そして、2023 年の CCW 締約国会議で合意されたマンデート（GGE はできる限り早期に、願わくは 2025 年末までにその作業を終了させること）の遂行を要請した（本文 4 項）上で、GGE のマンデートの遂行を完全に補完し支援する形で 2025 年にニューヨークで 2 日間の公開非公式協議（open informal consultations）を開催することを決定した（同 7 項、9 項）。開催時期について、その非公式協議は、補完と議論の充実のために、可能であれば、GGE 会合後に開催されることが要請された（同 11 項）。

本決議は、NGO の Stop Killer Robots⁶⁷⁾ によれば、「AWS に関する条約に向けた小さな一歩」であると評した。というのも、決議は、AWS に関する法文書交渉を緊急に開始すべきという国際社会の大多数の希望が明確に反映されず、一部の軍事大国によってその部分が縮小されて穏健なアプローチになったからである。しかしながら、定例化してきた国連での LAWS 議論は、CCW が他のフォーラムに対して優先的ではあるが独占的機関ではないことを証明し、さらに、補完的とはいえ、10 年以上も成果物を出していない CCW 枠での LAWS 議論に大きな刺激（または圧力）を付与する存在であることは明らかである。

64) 決議案について、A/C.1/79/77, 18 October 2024.

65) Isabelle Jones, “161 states vote against the machine at the UN General Assembly,” 05/11/2024, <https://www.stopkillerrobots.org/news/161-states-vote-against-the-machine-at-the-un-general-assembly/>, accessed 28 December 2024 ; “In Nearly 50 Separate Recorded Votes, First Committee Approves 15 Drafts on Conventional Weapons, Divergent Approaches to Outer Space Security,” GA/DIS/3756, 5 November 2024, <https://press.un.org/en/2024/gadis3756.doc.htm>, accessed 28 December 2024. 表決は、161 か国賛成、3 か国反対（ベラルーシ、北朝鮮、ロシア）、13 か国棄権（中国、エストニア、インド、イラン、イスラエル、ラトビア、リトアニア、ポーランド、ルーマニア、サウジアラビア、シリア、トルコ、ウクライナ）であった。

66) *Resolution adopted by the General Assembly on 2 December 2024, Lethal autonomous weapons systems. A/RES/79/62*, 10 December 2024, 166 か国賛成、3 か国反対、15 か国棄権。

67) Note 65, Isabelle Jones, op. cit.

5. まとめにかえて

(1) 三領域の AI 規制の相違点と共通点

民生用領域での AI 規制について、2019 年以降、諸国家および国際機構は、積極的に議論を重ね、非法文書（倫理規範、行動準則などのソフト・ロー）および法文書（国内法や EU 法）を策定してきた。特に、2022 年 11 月のチャット GPT という生成 AI（Generative AI）の登場以降、AI の現実のリスク・脅威に直面した現代社会は、一層、AI 規制プロセスを推進してきた。軍事領域全般での AI 規制について、特に 2023 年の国際会議「REAIM サミット」以降、約 60 か国が賛同する REAIM 行動要請、REAIM ブループリントおよび AI 軍事利用政治宣言といった非法文書が採択され、法文書まで至らないとしても、軍事領域での責任ある AI の利用法について合意形成が水平的（賛同国の増加）にも垂直的（詳細な規定内容）にも着実に進展している。これも、実際に AI による自律技術（飛行・走行・航行・潜航）の開発や指揮統制支援システムの実戦使用が明らかになってきたからであろう。他方、AI 兵器規制に関して、2014 年⁶⁸⁾ からすでに 10 年以上 CCW 枠内で議論を重ねてきたが、その成果は 2019 年の GGE 報告書で採択された 11 の指針原則（guiding principles）⁶⁹⁾ しかなく、現在も CCW の GGE で LAWS 審議が継続中である。

三領域での AI 規制動向の相違点は、どこに由来するのか。第 1 に、民生用や軍事領域全般では、上記の如く、現実の AI 利活用により具体的で現実的な規制議論が可能であった。他方、LAWS 規制議論は、禁止・規制すべき対象の定義さえ合意されず、「事後的・対処的・具体的」な討議ではなく、「事前的・予防的・抽象的」な討議に終始してきたからであろう。もっとも、近年、ウクライナ・ロシア戦争やイスラエル・ハマス戦争の勃発で、LAWS または AWS の研究開発が急速に進展し、当該兵器が実戦投入が間近に迫る様相から、より具体的な LAWS 規制議論が行われるようになった⁷⁰⁾。第 2 に、民生用よりも軍事用全般、軍事用全般よりも兵器用の AI 規制の方が、国家安全保障に直結する死活的課題であるので、諸国家の立場を調整することが一般的に困難となり得る。そこには、軍事大国と軍事小国の対立および軍事大国間（米・中・露）⁷¹⁾ の二重の対立構造が見られる。特に、CCW の手続き規則が、多数決ではなく、コンセンサス方式を採用しているので、より一層、CCW 全当事国

68) 2014 年までの LAWS 規制について、岩本誠吾「国際法における無人兵器の評価とその規制動向」『国際安全保障』42 巻 2 号 2014 年 9 月 21–23 頁。

69) CCW/GGE.1/2019/3, 225 September 2019 and CCW/MSP/2019, 13 December 2019. 岩本誠吾「第 4 章 AI 兵器と国際法と倫理規範—LAWS 規制の現状と展望」弥永・山田編著『AI・DX が変える現代社会と法』商事法務 2021 年 3 月 95–98 頁。

70) LAWS 規制において、2 層（禁止と規制）アプローチが合意されつつあり、合法使用のための具体的な規制要件として、人間の制御や判断、標的の種類、LAWS 運用の期間、地理的範囲、規模、攻撃回数、自己破壊・不活性化・自己無力化装置の装着など列挙されている。LAWS の合法使用に関する規制様式に、技術的事項に関する附属書を持つ CCW 改正第議定書 II が大いに参考となるだろう。

71) 軍事大国間では、AI を巡る覇権争いという積極的な兵器開発競争だけでなく、対立国家の AI 兵器開発に対抗するために軍備競争から一方的に離脱できない消極的な兵器開発競争という 2 側面がある。

の合意を得るにはハードルが高くなっている。

他方、三領域での AI 規制動向の共通点も指摘できる。民生用であれ軍事用であれ、「安全で、安心で、信頼できる」AI システムであるために、バイアスを軽減・回避する公平性、AI 本体に対する信頼性や安全性、不具合の原因究明のための透明性や追跡可能性や説明可能性、AI による意図せざる行動の解除・停止の制御可能性、AI 使用に対する人間の責任性といった規制原則は、AI 利用の三領域に適用できる。前述した米国主導の国連総会決議（A/RES/78/265）の採択に先立ち、米国連大使が発言したように、民生用の包摂的で建設的な対話が、軍事領域での AI 利用の議論のモデルとなり得るのである。

また、民生用、軍事用全般および兵器用の AI 規制の共通した近年の動向として、国連総会が討論の場として重用され、その結果、AI に関する国連総会決議が採択されるようになった。それは、民生 AI 利用に関する 2024 年 3 月・7 月の二つの総会決議（A/RES/78/265 and A/RES/78/311）、軍事領域全般の AI 利用に関する同年 12 月の総会決議（A/RES/79/239）および AI 兵器（LAWS）に関する同年 12 月の総会決議（A/RES/79/62）である。加えて、後者の二つの決議は、当該課題を次年度の総会 80 会期議題に挙げることを決定しており、それらの継続的な討議が予定されている。国連総会は、特に、LAWS 規制の補完的フォーラムであるとしても、あまり生産的でない CCW・GGE を前進させるための大きな政治的推進力となっていることは確かである。ただし、国連総会が、CCW・GGE を見限って多数決で独走すれば、軍事大国を置き去りにする危険性がある。

（2）日本の立場

では、日本は、三領域（民生用、軍事用、兵器用）の AI 規制に関して、どのような法的・政治的立場を取っているのか。民生用 AI 規制に関しては、前述したように、リスク・ベース・アプローチを採用した「広島 AI プロセス包括的政策枠組み」を基に、法規制の方向に展開しようとしている。軍事用および兵器用 AI 規制に関しては、外務省および防衛省が関連文書を公表しているため、当該文書から日本の立場を明らかにする。

まず、外務省は、国連事務総長の要請に従って、LAWS に関する問合せに対する回答書（作業文書）⁷²⁾ を国連に 2024 年 5 月に付託した。それによれば、LAWS を「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（lethal force）をもって交戦することができる」ものと定義する。そして、IHL を遵守して兵器システムを使用するためには、IHL 上の責任を問うことのできる人間の関与（human involvement）が必要であるという。日本は、人間の関与が全く及ばない完全自律型の致死性を有する兵器システムが、必ずしも IHL の遵守を確保できる段階にはない（are not necessarily

72) Note 62, A/79/88, pp. 65–67. 和文の要約として、「LAWS 分野における新興技術に係る我が国国連提出作業文書要約」<https://www.mofa.go.jp/mofaj/files/100687670.pdf>、2024 年 12 月 23 日閲覧。参照、外務省 HP「自律型致死兵器システム（LAWS）について」2024 年 6 月 24 日、https://www.mofa.go.jp/mofaj/dns/ca/page24_001191.html、2024 年 12 月 30 日閲覧。

reached a stage of ensuring) と考え、当該兵器システムを開発する意図はない旨、表明してきたという。

そして、日本も LAWS に対して 2 層（禁止と規制）アプローチを採用して、①性質上過度の障害または無用の苦痛を与えるもの、本質的に無差別なもの、その他 IHL に従って使用することができない兵器システム、また、②適切なレベルの人間の判断が介在せず、人間による責任ある指揮命令系統の中での運用が確保できないような、人間の関与が及ばない完全自律型致死兵器システムは、その開発および使用は国際的に認められてはならない (must not be permitted) と考える。上記以外の兵器システムについては、IHL を遵守する形での使用を担保するための制限またはリスク軽減策の実行が求められる。具体的な規制措置として、標的の種類の制限、自己破壊・自己不活性化・自己無力化メカニズム等による兵器システムの作動期間、地理的範囲、規模等の制限に言及している。人間の関与の在り方として最も重要な点は、指揮官や操作者が意図した形で兵器システムを運用できるような状態を確保することであるという。

今後の LAWS 規制議論の在り方として、その成果物は法的拘束力のある条項 (legally binding provisions) ではなく、実効的なルール (effective rules) であると繰り返し表明している。特に、人道的考慮と安全保障の観点のバランスを取りつつ、高い技術力を持つ国（軍事大国、著者注）を含むすべての利害関係国 (all stakeholder States) の参加を得ながら、コンセンサスでルールの形成を図ることこそが、ルールの実効性の担保につながると指摘する。その観点から、日本は、CCW 枠での議論の継続を強く支持している。

さらに、2024 年 7 月に防衛省は、軍事領域での AI 活用に関して初めて AI 活用推進の基本方針を策定した。その「防衛省 AI 活用推進基本方針」⁷³⁾ によれば、AI を重点的に活用する分野として、①目標の探知・識別、②情報の収集・分析、③指揮統制、④後方支援業務、⑤無人アセット、⑥サイバー・セキュリティ、⑦事務処理作業の効率化が列举されている（ただし、その 7 つの分野に限定する趣旨ではないという）。その活用には、人間の関与を確保する必要があることに留意すべきと指摘する⁷⁴⁾。そして、AI 利用に伴うリスクに対して、総務省・経済産業省策定の「AI 事業者ガイドライン」⁷⁵⁾ で示された方針を参考にしつつ、リスク軽減措置を行う。AI を使った装備品の研究開発に関して、日本は「人間の関与が及ばない完全自律型の致死性兵器の開発を行う意図がない」⁷⁶⁾ 旨を表明しており、加えて、「国際法や国内法により使用が認められない装備品の研究開発及び導入を行うことはない」⁷⁷⁾ と明

73) 防衛省「防衛省 AI 活用推進基本方針」2024 年 7 月 2 日、https://www.mod.go.jp/j/press/news/2024/07/02a_03.pdf, 2024 年 12 月 27 日閲覧。

74) 「同文書」8 頁。

75) 総務省・経済産業省「AI 事業者ガイドライン（第 1.0 版）」2024 年 4 月 19 日。共通指針として、①人間中心、②安全性、③公平性、④プライバシー保護、⑤セキュリティ保護、⑥透明性、⑦アカウントビリティなどが規定されている。「同文書」12-19 頁。

76) 「同文書」23 頁。

77) 「同文書」25 頁。

言する。

LAWS 規制に関して、国際社会が、追加規制不要派—非法文書派—法文書派の三つ巴状況⁷⁸⁾にある現時点で、日本が、軍事大国を含むすべての国家がコンセンサスでソフト・ロー（行動準則やベスト・プラクティス収集）の形成を図ることは最適解であろう。また、軍事領域での AI 利用についても、民生用 AI ガイドラインを援用することは、軍民両用の AI にとっては合理的である。

科学技術が進めば進むほど、三領域の AI 規制は、法文書形式であれ非法文書形式であれ、ますます必要とされ、これからも相互に影響し合いながら、その規定策定過程がますます奨励されるだろう。特に、LAWS 規制に関して、2026 年は、GGE が同年に開催される第 7 回 CCE 再検討会議に報告書を付託する年であり、国連事務総長が強く要望してきた LAWS 禁止・規制の法文書を交渉し締結し終えるデッドラインでもある⁷⁹⁾。LAWS 規制が 2025 年と 2026 年の 2 年間でどれほど進展するか予測ができないが、国連総会の討議に刺激を受けつつ、軍事大国を含むコンセンサス方式で何らかの具体的な GGE 報告書が採択されることが切望される。

* 添付翻訳資料

①「REAIM 行動ブループリント（REAIM Blueprint for Action）」

人工知能（AI）は、実現化する技術として、軍事作戦、指揮統制、諜報・監視・偵察（ISR）活動、訓練、情報管理、兵站支援を含む軍事のあらゆる側面を変革する並外れた潜在能力を秘めている。

AI の急速な進歩と発展に伴い、軍事領域で AI 技術を強化することに対する各国の関心が高まっている。同時に、軍事領域での AI 活用（AI applications）は、人道的、法的、安全保障、技術的、社会的または倫理的な観点からの様々な課題やリスクにつながる可能性があり、それらを特定し、評価し、対処する必要がある。

関連するリスクと課題に適切に対処しながら AI の利点と機会を活用するためには、AI によって実現されるシステムを含めて、軍事領域での AI 機能は、ライフ・サイクル全般を通じて責任を持って、そして、適用可能な国際法、特に国際人道法に従って適用されるべきである。

REAIM サミット 2023 で規定された行動要請に基づき、我々は、各国、産業界、学界、市民社会、

78) 注 6、岩本「自律型致死兵器システム（LAWS）の国際法政策論」25 頁表 2 参照。Automated Decision Research の「各国の立場（State positions）」は、三つに区分をしている。Automated Decision Research, State positions, https://automatedresearch.org/state-positions/?_state_position_negotiation=yes, accessed 19 October 2024

法文書不支持国 10	立場未宣言国 59	法文書支持国 126
豪、エストニア、印、イスラエル、日、ポーランド、韓、露、英、米	加、仏、ウクライナ、土など	中、独、伊、蘭など

【図 2：各国の立場】（著者作成）

79) 注 2 岩本「自律型致死兵器システム（LAWS）規制議論の動向と今後の展望—2023 年度—」57–58 頁。

地域的および国際的な組織を含むすべてのステークホルダー（利害関係者）に以下のことを勧誘する。

（AI が国際の平和と安全に与える影響）

1. 軍事領域での AI 活用は、国際の平和、安全および安定化を維持し、害しない方法で開発、配備、使用されるべきであることを確認する。
2. 軍事領域での AI は、事態の認識と理解、精度、精確性および効率性の向上などの利点をもたらす可能性があり、これらにより国際人道法の履行が強化され、武力紛争での文民および民用物の保護の取組が支援されること、そして、軍事領域での AI 活用は、平和構築や平和維持の諸活動の実効性と支援を高め、軍備管理やその他の遵守体制の検証と監視機能を強化する可能性があることを認識する。
3. AI 活用は、軍事領域の様々な側面において予見可能なリスクと予見不可能なリスクの両方をもたらす可能性があることも認識する。これらのリスクは、とりわけ、設計上の欠陥、データ・アルゴリズム・その他のバイアス（偏見）などに由来する意図しない結果、技術の潜在的な誤用または悪意のある使用、AI 活用と世界的および地域的な紛争や安定の複雑なダイナミクスとの相互作用（軍拡競争、誤算、エスカレーション、紛争の閾値の低下のリスクを含む）の原因となる可能性がある。
4. さらに、軍事領域で特に政策上の注意を払うべき大きな影響力を持つ可能性のある活用には、AI 対応（AI-enabled）兵器、戦闘作戦のための AI 対応意思決定支援システム、サイバー作戦での AI、電子戦での AI、情報作戦での AI が含まれ得ることを認識する。
5. AI 技術が国家やテロリスト集団を含む非国家主体による大量破壊兵器（WMD）の拡散に寄与するために使用されるのを防ぐ必要性を強調し、そして、AI 技術は軍縮、軍備管理、不拡散の取組を支援し妨げるものではないことを強調する。そして、核兵器のない世界という究極の目標を損なうことなく、核兵器の使用に関する至高の決定を通知および実行するための決定的なすべての行動について、人間の制御と関与（human control and involvement）を維持することが特に重大であることを強調する。
6. AI によって可能となるシステムを含む軍事領域において無責任な行為体が潜在的に有害な AI 機能を取得し悪用することを防ぐための強固な管理およびセキュリティ措置を確立することの重要性を強調する。ただし、当該措置が他の非軍事領域での AI 機能の利益への公平なアクセスを損なうものであってはならないことに留意する。

（軍事領域での責任ある AI の履行）

7. AI 活用は、適用可能な場合、国連憲章、国際人道法、国際人権法を含む国際法、および適切な場

合、地域文書を含むその他の関連する法的枠組みに従って、開発、配備、使用しなければならないことを確認する。

8. 軍事領域での責任ある AI 活用を確保するために、適切な場合、国家の戦略、原則、標準・規範、政策・枠組みおよび法律を策定することの重要性を強調する。
9. 軍事領域での責任ある AI を確保するために、網羅的ではないが、以下のことを認識する。
 - (a) AI 活用は、倫理的で人間中心的とすべきである。
 - (b) 軍事領域での AI 機能は、適用可能な国内法や国際法に従って適用されなければならない。
 - (c) 人間は、軍事領域での AI の使用と AI 活用の影響について引き続き責任（responsibility）と説明責任（accountability）を有し、そして、責任と説明責任を機械に移譲することはできない。
 - (d) AI 活用の信頼性と信用性は、データ、アルゴリズム、その他の偏向（biases）からを含めて、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安全措置を確立することにより、確保される必要がある。
 - (e) 軍事領域での AI の開発、配備、使用には、武力行使に対する人間の判断と制御（human judgement and control）に関連する適切な措置を含め、適切な人間の関与（appropriate human involvement）を維持する必要がある。
 - (f) 関係者は、軍事領域での AI 機能によって生成された成果（output）（AI により実施可能となるシステムを含む）を適切に理解し、説明し、追跡し、信頼できるべきである。軍事領域での AI の説明可能性（explainability）と追跡可能性（traceability）を改善する取組を引き続き行う必要がある。
10. 国際的な規範枠組み、厳格な試験・評価（T&E）プロトコル、包括的な検証・妥当性確認・認定（VV&A）プロセス、強固な国家監視メカニズム、継続的なモニタリング・プロセス、包括的な訓練プログラム、演習、強化されたサイバー・セキュリティおよび明確な説明責任の枠組みを通じて、国家的、地域的および国際的なレベルで軍事領域での責任ある AI を確保するための措置の策定について、更なる議論に参加し、対話を促進することを約束する（commit）。
11. 効果的な法的審査手続き、信頼醸成措置および適切なリスク軽減措置の開発、ならびに各国間のグッド・プラクティスや教訓に関する情報交換や協議を奨励する。また、産業界、学界、市民社会、地域的および国際的な組織などの他のステークホルダーに、適切な場合、定期的なマルチ・ステークホルダーの交流、事例研究やその他の関連文書の普及および共同発議への積極的な参加を含め、これらの取組に積極的に参加するよう奨励する。
12. 軍事領域での責任ある AI の取組は、AI 技術の研究、開発、実験、革新への取組と並行して実施することができ、それらの取組を妨げないことを強調する。

(軍事領域での AI に関する将来のガバナンス構想)

13. 軍事領域での AI のガバナンスに関する議論に、AI 技術とその能力と限界に関する共通の理解および軍事領域での可能な AI 活用とその潜在的な影響（利点とリスクの両方を含む）に関する共有された理解の促進が含まれるべきであると認識する。
14. 異なる国や地域が、軍事領域で AI 機能を統合化する様々な段階にあり、異なる安全保障環境にあり、そして、多様な安全保障上の懸念事項があることに留意して、幅広い見解を十分に反映するために、このような議論はオープンで包括的な方法で行われるべきであると強調する。
15. 軍事領域での AI のガバナンスに関する議論への発展途上国の完全な参加を促進し、そして、軍事 AI 機能の開発、配備、使用における責任あるアプローチを促進するために、特に、当該国家における能力構築の重要性を強調する。
16. 軍事領域での AI の責任ある開発、配備、使用に関する知識格差の縮小を目的とした能力構築に関する国際協力を強化することを約束する。
17. 軍事領域での AI 活用においてデータが重要な役割を果たしていることに留意し、そして、国家や他の関連ステークホルダーが、データの収集、保管、処理、交換、削除およびデータ保護に関する明確なポリシーと手続を含む適切なデータ・ガバナンス・メカニズムをさらに議論する必要があると認識する。
18. 技術の急速な発展と進歩に対応するために、軍事領域における AI ガバナンスに対する柔軟で均衡のとれた現実的なアプローチの必要性を認識する。
19. REAIM サミットや関連する地域行事、REAM グローバル委員会（REAM Global Commission）の設立、「AI・自律性の責任ある軍事利用に関する政治宣言」、特定通常兵器使用禁止制限条約（CCW）に基づき設置された自律型致死兵器システム分野の新興技術に関する政府専門家グループ（LAWS・GGE）、国連軍縮委員会および軍縮会議での議論を含めて、軍事領域での AI 活用に関する複数の取組における進展を認識する。自律型致死兵器システムに関する国連総会決議 78/241 や関連の地域会議・国際会議にも留意する。そして、これらの取組は、他のフォーラムで進行中の関連テーマの議論に影響を与えることなく、相乗的かつ補完的であるべきであると強調する。
20. 適切な場合、軍事領域での責任ある AI に関する取組は若者との意義ある関わりを必要とする世代の課題であると認識しつつ、軍事領域での責任ある AI に関する世界規模的および地域的な対話を、ステークホルダーによる積極的な関与と当該関係者間の情報交換を伴う開放的かつ包括的な方法で継続することを約束する。

我々は、諸国家にこの「行動のための青写真（Blueprint for Action）」への参加を勧誘するとともに、人類の未来のために責任ある AI を確立するための取組を継続することから、産業界、学界、市民社会、地域的・国際的機構を含むその他の関連するステークホルダーが「行動のための青写真」を支持しそ

れと連携することを歓迎する。

② LAWS・GGE「ローリング・テキスト 状況報告日：2024年11月8日」

GGEの作業のために、そして、すべての事項がコンセンサスに達するまで如何なる事項もコンセンサスに達していないとの理解の下で、部会は、部会のマンデートに従って、その本質やその他の可能な措置を害することなく、文書の一連の要素に関するその作業を推進するために、以下の定式化に暫定的にコンセンサスを見出した。

（LAWSの作業用特徴付け—著者注）

1. 自律型致死兵器システム（LAWS）は、一つ以上の兵器と技術的構成部分の統合された組合せとして特徴付けられ、そのことによって、当該システムが標的を識別しおよび/または選択し、そして、攻撃するという任務の執行において人間のユーザーが介入することなくそれらを実施することができる。
2. 上記の説明は、将来の理解やこの特徴付けの潜在的な変更および特定の種類のシステムを除外する可能性を害するものではない。

（LAWSに対するIHLの適用—著者注）

1. 戦闘の方法または手段を選択する武力紛争当事者の権利は、無制限ではない。
2. 国際人道法（IHL）は、国際的か非国際的かを問わず、武力紛争に適用され、過去、現在、そして、将来のすべての兵器、戦闘手段および方法の使用を規制し、従って、使用される軍事技術に関係なく適用される。
3. それゆえ、IHLは、LAWSの開発および使用に引き続き全面的に適用される。
4. 特定通常兵器条約（CCW）およびその付属議定書またはその他の国際協定で取り扱われていないLAWSに関連する事例の場合、文民と戦闘員は、常に、確立された慣習、人道の諸原則および公共の良心の命令に由来する国際法の諸原則の保護ならびに支配の下に置かれる。
5. LAWSの使用と効果に関する文脈に応じた（context-appropriate）人間の制御と判断は、その使用が国際法、特に攻撃における区別、比例性、予防措置の原則と要件を含むIHLに準拠していることを確保するために不可欠である。

（禁止と制限—著者注）

1. 如何なる状況においても、攻撃における区別、比例性および予防措置の原則と要件を含むIHLに準拠して使用できないLAWSを使用することは、禁止される。

2. 如何なる状況においても、余分な傷害や不必要な苦痛を引き起こす性質を有する LAWS または本質的に無差別的な LAWS の使用は、禁止される。
3. 如何なる状況においても、文民たる住民、個々の文民または民用物を LAWS による攻撃の対象とすることは、禁止される。
4. 如何なる状況においても、LAWS の使用の状況において IHL が要求するように攻撃におけるその効果が予測できず制御できない場合、それを配備または使用することは、禁止される。
5. 如何なる状況においても、文脈に応じた適切な人間の制御や判断なく機能する LAWS を使用することは、禁止される。
6. この目的のために、紛争当事者は、以下のことを行う必要がある。
 - A. LAWS およびその効果が IHL に準拠して使用されるように、十分に予測可能で、信頼性があり、追跡可能で、説明可能であることを確保すること、
 - B. LAWS は、文脈に応じた人間の制御や判断によって運用されることを確保すること、
 - C. 以下のものを含めて、標的の種類、LAWS 運用の期間、地理的範囲および規模を制限すること、
 - i. LAWS が起動後に人間のユーザーによって不活性化（deactivated）できることを確保すること、
 - ii. LAWS が自己破壊（self-destruct）、自己不活性化（self-deactivation）または自己無力化（self-neutralization）装置を組み込むことを確保すること、
 - iii. LAWS が着手できる交戦回数を制限すること、
 - iv. 文民または民用物の集中をよけること、
 - v. たとえば、本質的に軍事目標である標的に対する LAWS の使用を限定することを考慮することで、標的の種類を制限すること、
7. LAWS の任務変数（mission parameters）が、文脈に応じた人間の制御や判断なしにはシステムにより変更できないように確保する。

（IHL の遵守を確保するためのその他の措置—著者注）

潜在的な使用の先立ち、そして、LAWS のライフ・サイクル全般を通じて適用可能な場合、諸国家は、以下のことを行うべきである。

1. 国際法下の自国の義務に従って、新たな兵器、戦闘の手段または方法の研究、開発、取得、または採用に当たり、その使用が国際法により一定の場合またはすべての場合に禁止されるか否かを決定する。この文脈では、国家安全保障上の考慮事項または専有情報に関する商業上の制限を念頭に置きつつ、諸国家間で関連するベスト・プラクティスを自発的に交換することを推奨すること、

2. LAWS が、兵器システムのライフ・サイクル全般を通じて、文脈に応じた人間の制御と判断によって運用されることを確保すること、
3. 想定される使用状況での LAWS の能力と限界および異なるシナリオでの予想される効果が適切に理解されることを確保すること、
4. 現行の兵器システムの使用を深刻に変える開発中または改良中の兵器システムが IHL の遵守を確保するために審査されることを確保すること、
5. 兵器システムが予想されるその使用状況でどのように機能するのかについて人間のユーザーが信頼可能な予測をすることができるように、現実的に再現された運用環境で厳密な実験と評価を実施すること、
6. データ・セット内の有害なバイアス（偏向）を可能な限り検出し修正することができるように定期的に評価を実施し、そして、自動化バイアスを軽減する措置を履行すること、
7. LAWS が依拠する AI 能力内のあり得る有害なバイアスに対処するための措置を履行すること、
8. 国家は、LAWS の人間のユーザーに対して適切な案内、訓練および指示を確実に行うべきである。

（説明責任—著者注）

1. 国家は、人間が常に LAWS に関する決定について、適用可能な国際法に従って、責任と説明責任を負っていることを確保しなければならない。
2. 国家は、説明責任を機械に移譲することができないので、LAWS の使用における決定に対する人間の責任が保持されることを確保しなければならない。
3. 国家は、追跡可能で責任のある人間の指揮統制の連鎖内での LAWS の運用を含めて、適用可能な国際法や国内法に従って、当該システムのライフ・サイクル全般を通じて説明責任を確保しなければならない。
4. 国家は、IHL 下の自国の義務に従って、LAWS の潜在的使用を含む戦闘の手段または方法の使用に対して個人の責任を確保しなければならない。
5. 国家は、IHL 違反を伴うかもしれない事案の報告および適切な文書化のための効果的な内部の監察、調査および救済のメカニズムを確保しなければならない。内部メカニズムは、LAWS の開発、配備および運用に関連する事案にも対処するべきである。
6. 国家は、自国の管轄権下もしくは管理下にある人間によるかまたはその領域上での国際法違反を防止し抑制するために、立法措置およびその他の措置ならびに刑事上または行政上の制裁の適用を含むすべての適切な措置を取らなければならない。