

ロボット兵器の自律性に関する一考察

—LAWS（自律型致死兵器システム）を中心として—

上野 博嗣

はじめに

近年の戦場における無人兵器の使用は、火薬、核弾頭に次ぐ第3の軍事革命と言われており¹、無人兵器、すなわち「ロボット兵器(robotic weapons)」は²、人工知能(Artificial Intelligence: AI)との融合により³、自動化から「自律化」へと進化している⁴。また、ロボット兵器のうち特に自律型兵器システム(Autonomous Weapon Systems: AWS)の導入は、将来の戦争の本質に大きな影響を与えることが予想されている⁵。このように、ロボット兵器とAIの結合が、戦場の様相を変えつつあるのである。

中国は2030年までに「最高のグローバルAIイノベーションセンターになる」という目標を掲げ、その過程でAI技術が米国を上回る可能性が指摘されている⁶。また、現ロシア軍参謀総長ガラシモフ(Valery Gerasimov)将軍は、2013年に「戦争の未来」(future of warfare)という記事の中で、AI研究の重要性を挙げ、「近い将来、完全にロボット化された部隊が作ら

¹ Christof Heyns, *A/HRC/23/47: Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions*, UN Human Rights Council, April 9, 2013, pp. 5-6.

² 本稿で使用する「ロボット兵器」は、致死性を有する全ての無人機・無人システムを指す。

³ AIとは、言語間の翻訳など、通常は人間の知能を必要とするタスクをコンピューターシステムによって実行できるもの。UK Development, Concepts and Doctrine Centre (DCDC), *Joint Concept Note 1/18: Human-Machine Teaming* [hereafter *JCN 1/18*], UK Ministry of Defence, May 2018, p. 60.

⁴ 米国防総省では、ロボット兵器の自律性を、「目標を達成するために異なる行動方針の中から独自に展開し選択するシステムの能力」と定義し、この能力を有する自律システム(autonomous system)は、偏差のない規範的な規則によって管理される自動システム(automated system)と区別されると認識している。US Department of Defense(DoD), *Unmanned Systems Integrated Roadmap 2017-2042* [hereafter *USIR 2017-2042*], August 28, 2018, p. 17.

⁵ Michael N. Schmitt & Jeffrey S. Thurnher, “Out of the Loop”: Autonomous Weapon Systems and Law of Armed Conflict” [hereafter “Out of the Loop”], *Harvard National Security Journal*, Vol. 4, February, 2013, p. 231.

⁶ Elsa B. Kania, *Battlefield Singularity: Artificial Intelligence, Military Revolution and China’s Future Military Power*, Center for a New American Security, November 2017, p. 6.

れ、独立して軍事作戦を遂行することが可能になるであろう」と予測している⁷。

これらに対し、2016年6月に米国国防科学委員会(Defense Science Board: DSB)がまとめた『自律性に関する夏季研究報告書』(Summer Study on Autonomy)では、ロボット兵器に自律性を持たせることが「接近阻止／領域拒否(Anti-Access and Area-Denial: A2／AD)」を強化する主要な例として挙げられており、ジャミング等の敵対的環境下(adversarial environment)でも任務が遂行できる自律性を持ったロボット兵器の重要性を指摘している⁸。

このようななか、自律型兵器による危険な軍拡競争が進行中であると指摘されている⁹。特に、ヒューマン・ライツ・ウォッチ(HRW)とハーバードロースクール国際人権クリニックは、2012年11月に『失われつつある人間性: 殺人口ボットに反対する論拠』(以下、「HRW報告書」という。)において、今後AIを含む科学技術の発展により、20年から30年以内に人間の意思が介入することなく、標的を自ら選択し攻撃できる完全自律型兵器(fully autonomous weapons)、いわゆる「殺人口ボット(killer robots)」が開発されると予想している¹⁰。HRW報告書が問題としている点は、人間ではなく、ロボット兵器(若しくはAI)の意思決定によって、人間の生死が決まるという法的、倫理的問題である¹¹。このHRW報告書が指摘する殺人口ボットは、現在「自律型致死兵器システム(Lethal Autonomous Weapons System: LAWS)」と呼ばれており、特定通常兵器使用禁止制限条約(Certain Conventional Weapons: CCW)の枠組みの下で、2014年から非公式専門家会合で取り扱われるようになり、2017年以降は「LAWSに関する政府専門家会合(Group of Governmental Experts: GGE)」(以下、「LAWS・GGE」という。)において議論されている。しかし、ロボット兵

⁷ Quoted, Paul Scharre, *Army of None: Autonomous Weapons and the Future of War* [hereafter *Army of None*], W. W. Norton & Company, 2018, p. 117.

⁸ Department of Defense, Defense Science Board, *Summer Study on Autonomy*, June 2016, p. 61. A2/ADとは、米国によって示された概念で、接近阻止(A2)は、主に長距離能力により、敵対者がある作戦領域に入ることを阻止するための能力を指す。領域拒否(AD)とは、より短射程の能力により、作戦領域内での敵対者の行動の自由を制限するための能力を指す。防衛省『平成30年版 日本の防衛』、2018年、48頁。

⁹ *Army of None*, p. 117.

¹⁰ Human Rights Watch (HRW), *Losing Humanity: The Case against Killer Robots*, [hereafter *HRW report*], Harvard Law School International Human Rights Clinic (IHRC), November 2012, pp.1, 7-9.

¹¹ *Ibid.*, p. 7.

器の自律性という概念について共通認識が得られていないことや¹²、各国の安全保障上の利害の相違、ロボット兵器や AI に係る技術力の格差などを背景に、議論は収斂しているとは言えず、現在に至るまで LAWS の定義すら定まっていない。

日本では 30 大綱の中で、「自衛隊が保有する知見・人材を活用しつつ、自律型致死兵器システム(LAWS)に関する議論を含む国際連合等による軍備管理・軍縮に係る諸活動に関与する」としている¹³。防衛省・自衛隊として、国際的な議論の場で LAWS をめぐる共通認識作りや将来の規範形成に関与し、我が国にとって望ましい形で LAWS を管理していくことが求められるところ、我が国自身もまた LAWS についてしっかりと整理しておく必要がある。

そこで本稿では、LAWS とは何かを明らかにするため、第 1 項でロボット兵器の特徴について述べ、第 2 項でロボット兵器の分類を行った上で、第 3 項でコンピューター・サイエンスの分野からロボット兵器の自律性と AI の関係について述べる。第 4 項では国際人道法の観点から、完全自律型兵器に対する「法規制推進派」と「慎重派」の主張の対立点を明らかにし、現在の日本の立場がどちらに含まれるのかについて述べる。おわりに、LAWS を含む自律型兵器について日本が今後目指すべき方向性について私見を述べたい。

1 ロボット兵器の特徴

なぜロボット兵器が戦場で使用されるのか、その軍事的利点は、生身の兵士に代わり、Dull (退屈)、Dirty (汚い)、Dangerous (危険)、Deep (深い) の 4 つの「D 任務」を遂行することが指摘されている¹⁴。この他にも、特に人の手を介さない自律型ロボット兵器の場合は、自己保存の欲求がない分、恐怖、疲労、復讐心といった人間の感情や体調から生じる判断ミス

¹² Vincent Boulanin & Maaike Verbruggen, *Mapping The Development of Autonomy in Weapon Systems* [hereafter SIPRI report], Stockholm Int'l Peace research Inst (SIPRI), November 2017, pp. 118-119.

¹³ 30 大綱、16 頁。

¹⁴ Dull (退屈) の例は長時間監視任務等、Dirty (汚い) は化学・生物・放射性物質・核環境下での任務等、Dangerous (危険) は脅威の高い領域での任務等、Deep (深い) は敵領域内での長距離移動を要する任務等とされる。DCDC, *Joint Doctrine Note 2/11: The UK Approach to Unmanned Aircraft Systems* [hereafter JDN 2/11], UK Minisy of Defence, March 30, 2011, paras. 307-311.

を回避できるという指摘がある¹⁵。また、非戦闘員への攻撃などの戦争犯罪は、多くの場合、仲間の死に対して復讐を求める戦闘員によって行われるため、ロボット兵器に感情がなく仲間にに対して愛着を覚えないということは、ロボット兵器の使用が民間人の危険の軽減につながるという主張もある¹⁶。

また、自律型ロボット兵器の場合、遠隔操縦型ロボット兵器に比べて、より省人化を図れるという利点がある。従来の遠隔操縦型ロボット兵器の場合、例えば、MQ-1 プレデター(Predator)や MQ-9 リーパー(Reaper)のような無人航空機(Unmanned Aerial Vehicles: UAV)を遠隔操作により 24 時間運用する場合は、1 機の UAV に一軌道(orbit)あたり、7 から 10 人のパイロットを必要とし、また UAV のセンサーを操作する場合は、さらに 20 人が必要である。加えてセンサーデータを確認するには、多数の情報分析員(intelligence analysts)が必要となる¹⁷。このように、UAV の運用には相当な人員が関与しなければならないため、米空軍はこれらの航空機を「無人(unmanned)」と呼ぶことに反対し、あえて「遠隔操縦型航空機(remotely piloted aircraft)」と呼称している¹⁸。他方、ロボット兵器が人の手を離れて自律的に行動できるなら、それを運用する側の省力化・省人化が図られるのである。

さらにジャミングや通信が途絶された環境では遠隔操縦型ロボット兵器では活動が制限されてしまうため、このような環境でもロボット兵器が人間の制御から独立して（つまり、自律して）効果的に活動できることが求められてきている¹⁹。

このように、現代のロボット兵器は、より自律性(autonomy)の高いものを志向するようになりつつあるのが特徴である。

2 ロボット兵器の分類基準

米国防省によると、ある程度自律性を有するロボット兵器は、人間の介

¹⁵ Philip Alston, *A/65/321: Interim report of the Special Rapporteur on extrajudicial, summary or arbitrary executions*, UN General Assembly, August 23, 2010, p. 16.

¹⁶ Andrew P. Williams & Paul D. Scharre, *Autonomous Systems -Issues for Defence Policymakers* [hereafter *Autonomous Systems*], NATO Headquarters Supreme Allied Commander Transformation, October 2015, P. 70.

¹⁷ *Army of None*, p. 16.

¹⁸ Ibid.

¹⁹ Ibid., pp. 15-16.

入の程度によって、①半自律型兵器システム、②人間監視自律型兵器システム、③完全自律型兵器システムの3つに区分される²⁰。(表1参照)

表1【人間の介入の程度によるロボット兵器の区分】

No	区分	内容
①	「Semi-Autonomous Weapon System」(半自律型兵器システム) =「human <i>in the loop</i> (人間が意思決定過程の輪の中にいる兵器)」	人間の指令(human command)により標的を選択し、武器を使用することができる。
②	「Supervised Autonomous Weapon System」(人間監視自律型兵器システム) =「human <i>on the loop</i> (人間が意思決定過程の輪に関与できる兵器)」	ロボット兵器の行動を停止できる人間のオペレーターが監視できる下で、システムが標的を選択し武器を使用できる。
③	「Fully Autonomous Weapon System」(完全自律型兵器システム) =「human <i>out of the loop</i> (人間が意思決定過程の輪の外にいる兵器)」	人間の入力又は相互作用なく標的を選択し武器を使用できる。

出所: *HRW report*, p. 2 及び *Army of None*, pp. 44-47 から筆者作成

現在のロボット兵器を上記の分類基準に当てはめると、MQ-1 プレデターや MQ-9 リーパーなど、①の半自律型兵器システムが主流である²¹。HRW 報告書では、②の人間監視自律型兵器システムの具体例として、対空脅威（対艦ミサイルや航空機）を自動的に探知・識別・迎撃する米海軍の艦船搭載の MK15 ファランクス(Phalanx)近接防御火器システム(CIWS)や、ファランクスの地上版である米陸軍の対ロケット・野戦砲・迫撃砲システム(C-RAM)等を挙げている²²。ただし、これらのロボット兵器は、目標の検知・評価・対応過程が秒単位であることから、判定過程での人間による適切な監視（システムがオペレーターの意図しない行動をした時の停止決定）が事実上ほぼ不可能なため、③のいわゆる完全自律型兵器システ

²⁰ *HRW report*, p. 2; *Army of None*, pp. 44-47.

²¹ “Out of the Loop,” p. 236.

²² *HRW report*, pp. 9-13.

ムに含むことも可能であるとしている²³。この他に③の完全自律型兵器の前駆的(precursor)なものとして、韓国が北朝鮮との非武装地帯(DMZ)に配備しているサムソン社製 SGR-1(静止型武装歩哨ロボット)や、イスラエルがカザ地区を対象として敵対的なレーダー信号を検出して攻撃する徘徊型兵器(loitering munition)といわれるハーピー(Harpy)が挙げられている²⁴。(図1及び2参照)

図1【SGR-1】²⁵出所：*HRW report*

図2【ハーピー発射風景】

出所：*Army of None*

なお、②と③の間にある唯一の実質的な違いは、③の完全自律型兵器はオペレーターの意図しない作動をロボット兵器がした時に、オペレーターがオーバーライドする機能がないというところに違いがあるとされる²⁶。

3 ロボット兵器の自律性とAIの関係

次に、ロボット兵器の自律性を向上させると言われるAIとロボット兵器の自律性の関係について整理する。システムの自律を可能にする技術として、米国防総省はAIのほか、機械学習(machine learning: ML)を挙げている²⁷。

²³ Ibid., pp. 2-3.

²⁴ Ibid., pp. 13-15, 18. このハーピーについては、チリ、中国、インド、韓国、トルコ等が所有しているといわれている。*Army of None*, p. 47.

²⁵ 敵兵が降伏を示していることを SGR-1 が認識できるかどうかのテスト風景とされる。*HRW report*, p. 14.

²⁶ “Out of the Loop,” p. 236.

²⁷ 機械学習とは、コンピューターが大量のデータを処理しながらデータの「分け

(1) AIの種類と、完全自律型兵器の2つの区分

AIは用途に基づき、大きく分けて「特化型(narrow)AI」と「汎用型(general)AI」の2種類ある。「特化型AI」は、特定の決まった作業を遂行するためのもので、自動運転技術や画像・音声認識、将棋、チェスなど、一つの機能に特化して稼働するものである。「汎用型AI」は、特定の作業やタスクに限定せず人間と同様の、あるいは人間以上の汎化能力を持ち合わせており、プログラミングされた特定の機能以外にも自身の能力を応用して対応できるとされる²⁸。また、AIがどれだけ人間の精神（もしくは思考）を模倣するか、またはAIが意識や自我を持っているかどうかという区分により「弱い(weak)AI」と「強い(strong)AI」に分けられる²⁹。「弱いAI」は、あらかじめ決められた枠内のみを思考するAIである。「強いAI」は、人間のように精神を持ったAIである。「弱いAI」は「特化型AI」に、「強いAI」は「汎用型AI」に、それぞれ対応した概念であると言える³⁰。以下、本稿では「特化型AI」を「弱いAI」の概念を含むもの、「汎用型AI」を「強いAI」の概念を含むものとして使用する。

「特化型AI」は、その決められた枠内（例えばチェスの試合など）であれば、既に人間の能力を超えているものも存在している。しかしながら「特化型AI」の能力がいくら高度になったとしても、あらかじめ人間が決めたこと以外は何もできないため、「人間の能力を補佐、拡張するAI」と言われている。

一方、「汎用型AI」は、いわゆる人間の脳と同じ機能をもったAIであるが、現在の技術において「人間のように思考するコンピューター」は存在していない³¹。よって現在のAIは、すべて「特化型AI」であり「弱いAI」である。

では、この「汎用型AI」がいつ作成されるのかについては様々な意見が

方」を自動的に習得し、一旦「分け方」を習得すれば、それを使って未知の画像データ等を判読することができる技術。松尾豊『人工知能は人間を超えるか』角川EPUB選書、2015年、117頁；USIR 2017-2042, pp. 17-18.

²⁸ 「AIを知るための4つの類型」「特化型」と「汎用型」、「強い」と「弱い」 MUFG Innovation Hub、<https://innovation.mufg.jp/detail/id=123>、2018年10月25日アクセス。

²⁹ 哲学者ジョン・サールが1980年に「強いAI」には「精神が宿る」ものだと主張した。鳥海不二夫『強いAI・弱いAI』有斐閣、2017年、2-4頁から再引用。

³⁰ AIは学問分野としてはコンピューター科学(computer science)に属するが、発展途上の分野であり、研究者の中でもAIの定義付けや分類分けをすることが困難であることから、specialized AI (weak AI)、Artificial General Intelligence (strong AI)と記載されることもある。SIPRI report, p. 90.

³¹ 『人工知能は人間を超えるか』38頁。

ある。一部の研究者は、全脳模倣（エミュレーション）が 2040 年代には早くもスーパーコンピューターで可能になると予測しており、またこのため AI の専門家の大多数は「汎用型 AI」が 2040 年までか、今世紀の終わりまでに可能になるかもしれないと考えている³²。他方、ストックホルム国際平和研究所では、汎用型 AI は空想科学(Science Fiction: SF)の領域と評価しており³³、「汎用型 AI」がいつ頃開発されるのかは、誰にも予測できないとも言われている³⁴。「汎用型 AI」については、現在の「特化型 AI」の技術の先にあるものではなく、どういった方向で研究を進めれば「汎用型 AI」を開発することができるのか、その開発方法すらわかつていないというのが実情である³⁵。

では、冒頭で述べた HRW 報告書が批判の対象とする完全自律型兵器は「特化型 AI」と「汎用型 AI」のどちらの AI を搭載していることを想定しているのだろうか。これは完全自律型兵器における共通理解を図る上で重要なポイントである。

HRW 報告書では、サムソン社製 SGR-1 やイスラエル製のハーピーを完全自律型兵器の前駆的なものとして批判している³⁶。しかし、現状で「汎用型 AI」は実在しないのだから、SGR-1 やハーピーに搭載されているのは「特化型 AI」に分類されるものである。「特化型 AI」は特定の決まった作業を遂行することしかできず、兵器の場合ならば、人間が設計したプログラムに従って与えられる標的に対し、そのパラメーターに合致したものだけに反応（攻撃）できるのであり、AI 自ら目標を選定し攻撃することはできない。そのため、「特化型 AI」を搭載したロボット兵器を投入したオペレーターは、事前にその兵器が何を攻撃するのか、相当程度に予測することができる。一方、「汎用型 AI」が人間のように思考することができると思えば、目的（例えば「勝利」など）あるいは包括的な任務（例えば、「敵部隊の殲滅」など）を与えるだけで、周囲の状況を把握し、情勢判断をした上で、目的又は包括的な任務を達成するために合理的な標的を AI 自ら選定し、攻撃することも可能になる。この場合、そのロボット兵器を投入したオペレーターは、事前にロボット兵器が何を攻撃するのか予測することができない。

³² *Army of None*, pp. 231-232.

³³ *SIPRI report*, p. 92.

³⁴ *Army of None*, pp. 231-232.

³⁵ 『強い AI・弱い AI』 18 頁。

³⁶ *HRW report*, pp. 13-18.

こうした違いを踏まえて、完全自律兵器システムをさらに細分化するなら、「特化型完全自律兵器システム」(Fully AWS with narrow AI)と、「汎用型完全自律兵器システム」(Fully AWS with AGI)の2つに区分することができる(表2参照)。ハーピーやSGR-1は前者に属すると言える。また、後者は現存しないため、具体例を挙げることは出来ない。ただし、イメージアップを助けるために敢えてSFの例を挙げるならば、映画『ターミネーター』の登場ロボットがこれに該当すると言えるかもしれない³⁷。

表2【完全自律兵器システムの分類】

特化型完全自律兵器システム (Fully AWS with narrow AI)	人間の入力又は相互作用なく、人間が設計したプログラムに従って標的を搜索、選択し、攻撃できる兵器(兵器が何を攻撃するか、相当程度に予測できる)。
汎用型完全自律兵器システム (Fully AWS with AGI)	人間の入力又は相互作用なく、AI自らが標的を搜索、選択し、攻撃できる兵器(兵器が何を攻撃するか、予測できない)。

作成筆者

これまで述べてきたように「特化型AI」と「汎用型AI」を単に「AI」と呼称したり、あるいは標的を搜索、選択し、攻撃するという一連のサイクル(いわゆるキル・チェーン)を人間が介入することなく実行できる兵器を単に「完全自律型兵器」と呼称したりすると、本質的に性質の異なるはずの兵器を混同することとなり、錯誤を起こしやすい。人間が設計したプログラムに従ってキル・チェーンを行う「特化型完全自律兵器」と、目的を達成するためにAI自らキル・チェーンを行う「汎用型完全自律兵器」は、倫理上の問題や人間に与える危険性という観点から別次元であろう。

よって、こうした相違をあらためて分類し直し、図式化すると、図3のとおりとなる。HRW報告書では、この両者を区別せずに「完全自律型兵器」を批判していると思われるが、完全自律型兵器はレベルの異なる2つのAIに区分して評価すべきである。

³⁷ 映画『ターミネーター』シリーズでは、人類の殲滅を目的とするスカイネット(SkyNet)という自我を持った架空のコンピューターが登場する。また、シャーレ(Paul Scharre)は現実世界において、米国防総省や国連では、自律型兵器に関する重要な討論がされる際、10回中9回は誰かがターミネーターについて言及していると述べている。Army of None, p. 264.

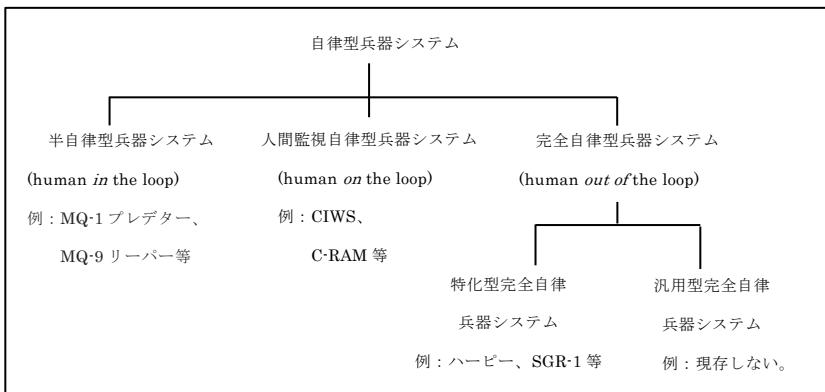


図3【自律型兵器システムの区分】

作成筆者

(2) ロボット兵器の自律性を向上させるAI技術の可能性と課題

次に、ロボット兵器の自律性を向上させるAI技術として、最も重要なディープ・ニューラル・ネットワーク(Deep Neural Network: DNN)について述べる。

DNNは、MLの一種であり、生物学における動物の脳内のニューロンの働きと対比しつつAIにアプローチするものである。DNNは、入力と出力の間に複数の隠れたレイヤー(層)があり、入力データと出力の間にネットワーク内の層を追加すると、ネットワークの繋がりが人間の脳のニューロンの繋がりのように、はるかに複雑になることから、ネットワークがより複雑なタスクを処理できるようになることが知られている³⁸。

2015年に、マイクロソフトの研究者チームは、人間の画像識別能力を超えるDNNを初めて作成したと発表した。150,000画像の標準テストで、人間のエラー率が推定5.1パーセントであるのに対し、マイクロソフトのDNNはこれを下回る4.94パーセントのエラー率であった³⁹。2015年の段階ですでに画像認識の分野では、DNNが人間の能力を超えていたことが分かる。一方、既存のロボット兵器に搭載されている自動標的認識(automatic target recognition: ATR)ソフトウェアは、認識能力や敵対的環境(ジャミング等)での脆弱性が指摘されており、既存のATRのほとんどは、戦車、航空機、潜水艦、レーダーなど、大きくて明確に定義された軍

³⁸ *Army of None*, pp. 86-87.

³⁹ Ibid., p. 87.

事目標だけしか認識することができない。また、ATRはその認識能力が非常に粗く、標的が人間であることは認識することができるが、その人間が文民か戦闘員であるかを区別する能力はない。また、気象条件に影響されやすい特徴がある⁴⁰。このため、DNNをATRアルゴリズムに転用することにより、敵味方識別や文民と戦闘員を識別する能力が格段に向上し、ロボット兵器の自律性が担保できると考えられるのである。一方で、このDNNは敵対的画像(adversarial images)に対して極めて脆弱である点が指摘されている⁴¹。例えば下の画像を見ていただきたい(図4参照)。人間に左の元画像と右画像の違いがわからないが、右画像には中央のノイズが入れられている。DNNは左画像を正しく認識できたにも関わらず、ノイズが入れられた右画像は全て「ダチョウ(ostrich)」と識別したのである⁴²。このように、仮にロボット兵器がノイズを伴うジャミングを受けた場合、攻撃対象をダチョウと認識して攻撃を止めるならともかく、民間人(民用物)や友軍を攻撃対象と判定するような間違いは、国際人道法の観点から

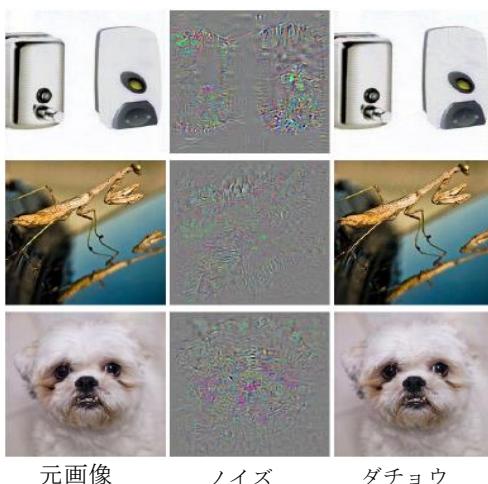


図4【Deep Neural Network 敵対的画像】
出所：*Intriguing properties of neural networks*

⁴⁰ SIPRI report, pp. 24-25.

⁴¹ Army of None, p. 180.

⁴² Christian Szegedy, et al., *Intriguing properties of neural networks*, Cornell University Library arXiv.org, December 2013, p. 6, <https://arxiv.org/pdf/1312.6199.pdf>, last visited June 10, 2019.

は許されないし、DNN の脆弱性を ATR に受容することもできない⁴³。研究者からは、なぜ DNN がこのタイプの操作に影響されやすいのか、ようやく理解され始めたばかりだと指摘されている⁴⁴。従ってこの「特化型AI」に区分される DNN においても、人間であれば絶対にしないような間違い（非人間的出力振動）を起こす間は、その技術をロボット兵器に転用することはできない。人間の介入なしにロボット兵器に攻撃を行わせることができるようになるまでには、相当の時間を要するものと考えられる。

4 完全自律型兵器の法規制の議論

これまでには、主にコンピューター・サイエンスの観点からロボット兵器の自律性について整理した。

次に、こうした整理を基に、国際人道法及び政策の観点から完全自律型兵器に対する法規制推進派と慎重派の主張を考察したい⁴⁵。

(1) 完全自律型兵器に対する法規制推進派の主張

まず、CCW 締約国の中では、ブラジル等の中南米諸国、東西冷戦期にどちらの陣営にも加盟しなかつたいわゆる非同盟(Non-Aligned Movement: NAM)諸国や中国等の 32 か国が、完全自律型兵器を規制するための新たな条約や追加議定書を策定することを望んでいる（2018年5月21日現在）⁴⁶。これらの国々の中には、ロボット兵器に自律性を持たせる AI 技術を自前

⁴³ *Army of None*, p. 183. DNN を含む ML が ATR に使用できないとの同様の指摘として、*SIPRI report*, p. 114.

⁴⁴ *Army of None*, p. 183.

⁴⁵ 法規制推進派と慎重派の議論は、岩本誠吾「ロボット兵器と国際法」弥永真生他編『ロボット・AIと法』有斐閣、2018年、307-308頁に詳しい。

⁴⁶ アルジェリア、アルゼンチン、オーストリア、ボリビア、ブラジル、チリ、中国（中国の立場はあくまで完全自律型兵器の使用を禁止することに賛成するのであり、開発や製造ではないという独自の立場を示している）、コロンビア、コスタリカ、キューバ、ジブチ、エクアドル、エジプト、ガーナ、グアテマラ、バチカン市国、イラク、カザフスタン、メキシコ、ニカラグア、パキスタン、パナマ、ペルー、フィリピン、シオラレオネ、南アフリカ、スリランカ、パレスティナ、ウガンダ、ベネズエラ、ザンビア、ジンバブエ（計 32 か国）。Campaign to Stop Killer Robots, *Report on Archives: Convention on Conventional Weapons Group Governmental Experts Meeting on Lethal autonomous Weapons Systems*, [hereafter CSKR Report], April 9-13, 2018, pp. 7-8, https://www.stopkillerrobots.org/wp-content/uploads/2018/07/KRC_ReportCCWX_Apr2018_UPLOADED.pdf, last visited March 24, 2019.

で開発する能力が十分でないと考えられるものが散見される。

ア ブラジル

ブラジルは、2019年 LAWS・GGE 第1会期において、LAWS を有意な人間の参加(meaningful human participation)なしに、プログラムのパラメーターに従って、合法的と判断した標的を選択し攻撃するものと評価するとともに、国際人道法は、本質的に人間の価値判断に基づいているため、LAWS は国際人道法を遵守できないと主張している⁴⁷。

ブラジルの LAWS に対する主張は、人間の介入なしに単にプログラムのパラメーターに従って攻撃することを問題視しており、本稿でいう「特化型完全自律兵器」を規制の対象にしようとしていると思われる。

イ パキスタン

パキスタンは、2016年第3回の LAWS 非公式専門家会合において、現状で標的殺害(targeted killing)等に使用されている遠隔操縦型の武装ドローンと同様に、LAWS も主権侵害につながるものと評価しており、人の生死を決定する際（攻撃する際）に人間が意思決定過程の輪にいないこと(human out of the loop)を批判の対象にしている⁴⁸。このため、ブラジルと同様に「特化型完全自律兵器」を規制対象にしようとしていると思われる。

ウ 中国

中国は 2018 年 LAWS・GGE 第 1 会期において、LAWS は既存の兵器と開発中の兵器に密接に関係しているものの、多くの国が現在のところ存在しないと認識していると評価し、以下の 5 つの特性を含むものを LAWS と見なすべきだと主張している⁴⁹。

それは、①十分なペイロードがあり致死性(lethality)を有すること。②任務を遂行する全過程において、人間の介入と制御が行われない自律性(autonomy)を有すること。③一旦起動されれば、システムを停止する方法がないこと(impossibility for termination)。④状況やシナリオに関係なく標的を殺害する無差別性(indiscriminate effect)を有すること。⑤環境との

⁴⁷ 2019年 LAWS・GGE 第1回会期におけるブラジルの声明書より。

[https://www.unog.ch/80256EDD006B8954/\(httpAssets\)/122DF2DAEE334DDBC12583CC003EFD6F/\\$file/Brazil+GGE+LAWS+2019+-+Item+5+a+-+IHL.pdf](https://www.unog.ch/80256EDD006B8954/(httpAssets)/122DF2DAEE334DDBC12583CC003EFD6F/$file/Brazil+GGE+LAWS+2019+-+Item+5+a+-+IHL.pdf), last visited April 24, 2019.

⁴⁸ 2016年第3回の LAWS 非公式専門家会合におけるパキスタンの声明書より。

[https://www.unog.ch/80256EE600585943/\(httpPages\)/37D51189AC4FB6E1C1257F4D004CAF82?OpenDocument](https://www.unog.ch/80256EE600585943/(httpPages)/37D51189AC4FB6E1C1257F4D004CAF82?OpenDocument), last visited April 24, 2019.

⁴⁹ 2018年 LAWS・GGE 第1回会期における中国の声明書より。

[https://www.unog.ch/80256EDD006B8954/\(httpAssets\)/E42AE83BDB3525D0C125826C0040B262/\\$file/CCW_GGE.1_2018_WP.7.pdf](https://www.unog.ch/80256EDD006B8954/(httpAssets)/E42AE83BDB3525D0C125826C0040B262/$file/CCW_GGE.1_2018_WP.7.pdf), last visited Oct 24, 2018.

相互作用を通じて、自律的に学習する進化性(evolution)を有することである。

中国の LAWS に対する主張は、単に人間が設計したプログラムに従って標的を選択する「特化型完全自律兵器」だけではなく、周囲の状況を把握し、情勢判断をした上で攻撃する「汎用型完全自律兵器」を含めて LAWS と認識していると考えられる。

エ 学説等

HRW 報告書は、国際人道法の観点から、今後開発される可能性がある完全自律型兵器は、国際人道法の中核原則（区別原則、比例原則、軍事的必要性、マルテンス条項）を遵守できないと指摘している⁵⁰。

また、前述のとおり、HRW 報告書では、サムソン社製 SGR-1 やイスラエル製のハーピーを完全自律型兵器の前駆的なものとして批判しており、「特化型完全自律兵器」を含めて規制の対象にしようとしていると推察できる。

（2）完全自律型兵器に対する法規制慎重派の主張

これに対し、CCW 締約国の中で完全自律型兵器として批判されているハーピー や SGR-1 を有するイスラエルや韓国のはか、中国を除く英米仏露など安保理常任理事国を含む 12 か国が完全自律型兵器に関する新しい国際法を制定することを支持しないとしている（2018年5月21日現在）⁵¹。

ア 米 国

米国は、HRW 報告書が出された 2 日後（2012年11月2日）に米国防総省から『指令 3000.09（兵器システムにおける自律性）』を公布した。その中で「自律型兵器システム」 や⁵²、「半自律型兵器システム」 は⁵³、適用さ

⁵⁰ 区別原則：軍事目標と文民を区別し、軍事目標のみに攻撃を行うこと。比例原則：予期する文民被害が予期する軍事的利益を上回る場合に攻撃を禁止すること。軍事的必要性：勝利するためにあらゆる必要な軍事的措置を講じること。マルテンス条項：人道の法則と公共良心の要求に従って戦争手段を評価すること。

HRW report, pp. 30-36.

⁵¹ オーストラリア、ベルギー、フランス、ドイツ、イスラエル、韓国、ロシア、スペイン、スウェーデン、トルコ、米国、英国（計 12 か国）。*CSKR Report*, p. 8.

⁵² 米国防省は「自律型兵器システム」とは、「一旦起動すれば、人間のオペレーターによる更なる介入(intervention)がなくとも標的を選択(select)し攻撃(engage)できる兵器システム」と定義している。DoD, *Directive 3000.09: Autonomy in Weapon Systems*, November 21, 2012, p. 13.

⁵³ 米国防省は「半自律型兵器システム」とは、「一旦起動すれば、人間のオペレーターによって選択された個々の標的または特定の標的のみを攻撃する兵器システム」と定義している。この他に「人間監視自律型兵器システム(human-supervised

れる全ての国内法及び国際法、特に戦争法との両立を保証するものでなくてはならないとしている⁵⁴。更に「自律型及び半自律型兵器システムは指揮官やオペレーターが武器を使用する際、適切なレベルの人間の判断を行使できるように設計されていなければならない」として⁵⁵、あえて「完全自律」という表現を避けている。米国が「完全自律」という用語を使用しない理由は、人間の指令から完全に独立して作動する自律システムは軍事的な観点から役に立たないと考えており、兵器の「完全自律」は開発できぬのではなく、そもそも目的ではないとしている⁵⁶。

この他に2015年に米国防総省で作成された『戦争法マニュアル』では、「戦争法は自律型兵器システムの使用を禁止していない」と明文化する一方⁵⁷、「戦争法は兵器自体ではなく（中略）人に義務を課す」と記載している⁵⁸。米国防総省の観点から、兵器がどれだけ知的になり自律化されても、戦争法の行為主体（権利や義務を負うもの）になるわけではなく、兵器は常に人の手の中にある道具でなくてはならないという立場であると言える⁵⁹。

イ 英 国

英国防省は、2011年に発刊した『統合ドクトリン：無人航空機システムへの取り組み』の中で、自律システムについて、「（機械が）自らを認識（self-aware）でき（中略）人間と同じレベルの状況理解を達成することができなければならない」⁶⁰として、「汎用型AI」に類似したレベルを想定している⁶¹。また、2018年に同じく英国防省から発刊された『統合基本概念：ヒューマン・マシン・チーミング』では、ロボット工学とAIが軍事的優位性を提供するものと理解されながらも「機械は戦場で戦わない（Machines don't fight wars）」として、人間の脳のように物事を処理できない機械は自律し

autonomous weapon system）」を「兵器システムが故障した場合を含め、容認できないレベルの損害が発生する前に、人間のオペレーターが交戦に介入して終了させる能力を提供するように設計された兵器システム」と定義している。Ibid., p. 14.

⁵⁴ Ibid., p. 11.

⁵⁵ Ibid., p. 2.

⁵⁶ *SIPRI report*, p. 64.

⁵⁷ Office of General Counsel DoD, *Department of Defense Law of War Manual*, DoD, June 2015 (updated December 2016), p. 353.

⁵⁸ Ibid., pp. 353-354.

⁵⁹ *Army of None*, p. 245.

⁶⁰ *JDN 2/11*, para. 206.

⁶¹ *Army of None*, p. 110.

ないという立場を示している⁶²。

また、英国は2016年第3回のLAWS非公式専門家会合において、LAWSは存在せず、今後も存在しないと言及しており⁶³、(プログラムの規則に従い)ロボット兵器自身で標的を検索、選択し、攻撃する兵器を自国で開発する可能性について言及し、これは単に「自動兵器(automated weapons)」であり、「自律型兵器(autonomous weapons)」ではないと主張している⁶⁴。すなわち、本稿でいう「特化型完全自律兵器」は「自動兵器」の範疇であり、規制されるべきではないという立場を示したことになる。

ウ NATO

2015年にNATOで作成された『自律システム：防衛政策立案者の課題』では、「武力紛争法及び国際人権法は軍事機能(military functions)を自律システム(autonomous systems)に委譲することを禁止していない」としている⁶⁵。他方、NATOも米国防総省と同様に、「ロボットは戦闘員ではなく(中略)戦闘員の手の中にある単なる道具」と評価している⁶⁶。

このように米国、英国やNATOの立場からすると、自律型兵器システムが規制されるのは、兵器に道具性がなくなったとき、すなわち人間の意思を介することなく兵器自らが標的の選択を行うことができるようになった時であり、本稿でいう「汎用型完全自律兵器」の場合と推察できる。

以上をまとめると、表3のように整理できる。

表3【完全自律兵器システムの分類と法規制に係る立場の関係】

区分	完全自律型兵器の規制対象
法規制推進派	特化型完全自律兵器及び汎用型完全自律兵器
法規制慎重派	汎用型完全自律兵器

作成筆者

エ 学説等

この他に、米海軍大学教授(国際法)のシュミット(Michael N. Schmitt)

⁶² *JCN 1/18*, p. iii.

⁶³ 2016年第3回のLAWS非公式専門家会合における英国の声明書より。

[https://www.unog.ch/80256EDD006B8954/\(httpAssets\)/49456EB7B5AC3769C1257F920057D1FE/\\$file/2016_LAWS+MX_GeneralExchange_Statements_United+Kingdom.pdf](https://www.unog.ch/80256EDD006B8954/(httpAssets)/49456EB7B5AC3769C1257F920057D1FE/$file/2016_LAWS+MX_GeneralExchange_Statements_United+Kingdom.pdf), last visited April 26, 2019.

⁶⁴ *Army of None*, pp. 110-111; *JCN 1/18*, p. 60.

⁶⁵ *Autonomous Systems*, p. v.

⁶⁶ *Ibid.*, p. 3.

及び NATO 緊急展開軍団(Rapid Deployable Corps)法律顧問のトナーー (Jeffrey Thurnher)は、次のように主張する。「戦争の目的は、敵の弱点につけこみ目的を達成することにあり、武力紛争法は決して『公平な戦い(fair fight)』を保証するものではない」。「自律型兵器システムの使用が禁止されるとすれば、武力紛争法の規定に反して使用された場合に限り、違法となることを示唆している」。例えば、「慣習国際法の下で生物化学兵器はそれ自体、違法であり、それが合法的な軍事目標に使用されたとしても違法である。反対にライフルは合法であるが、それが民間人や捕虜に使用されれば、違法となり得る」。「自律型兵器システムの完全なる禁止は、法律、政策、および良識的観点から不可能であり、存在しない武器を全面禁止することは時期尚早」と結論づけている。更にシュミットらは、「自律型兵器システムを規制することは、自律的ではない兵器と比較して、民間人や民用品の被害を局限する可能性があるにも関わらず、その可能性を阻害する」として、HRW らの法規制推進派に反論する⁶⁷。

(3) 完全自律型兵器に対する日本の立場

日本は、2016年の第3回 LAWS 非公式専門家会合において、LAWS は「現在のところ存在しないが、戦場を選択し、人間の介入なしに人間を標的として殺害することができる兵器の能力」と理解されているとした上で、「日本政府は無差別な致死性を有するであろう *human out of the loop* のロボットを開発する計画はない」と言及した⁶⁸。この考えは 2019 年 LAWS・GGE 第 1 会期においても踏襲されており、日本は致死性を有する完全自律型兵器を開発する計画はないと言及している⁶⁹。

現在、法規制推進派と慎重派は、共に LAWS の根幹機能(critical functions)について、ある程度の人間の関与が必要であることには同意しており、現在の LAWS・GGE での中心課題は、ロボット兵器に対する「有為な人間の関与(Meaningful Human Control: MHC)」とは何かである⁷⁰。日本はこの「MHC が確保された自律型兵器システムについては、ヒ

⁶⁷ “Out of the Loop,” pp. 232-234, 243.

⁶⁸ 2016年第3回の LAWS 非公式専門家会合における日本の声明書より。

[https://www.unog.ch/80256EDD006B8954/\(httpAssets\)/B367B41929F206A4C1257F9200573ADC/\\$file/2016_LAWS+MX_GeneralExchange_Statements_Japan.pdf](https://www.unog.ch/80256EDD006B8954/(httpAssets)/B367B41929F206A4C1257F9200573ADC/$file/2016_LAWS+MX_GeneralExchange_Statements_Japan.pdf), last visited April 26, 2019.

⁶⁹ 2019 年 LAWS・GGE 第 1 回会期における日本の作業文書（以下、“Working Paper”）より。<https://www.mofa.go.jp/mofaj/files/000459707.pdf>, last visited March 25, 2019.

⁷⁰ Elke Schwarz, “The (im)possibility of meaningful human control for lethal

ューマンエラーの減少や、省力化・省人化といった安全保障上の意義があり（中略）現在運用されている高度に自動化された兵器システムは、MHC を確実にするように設計されているため、規制されるべきではない」と主張している⁷¹。

この主張には 3 つのポイントがあると考えられる。1 つ目は、LAWS を「戦場を選択（中略）することができる兵器の能力」と評価しており、兵器が自ら戦場を選択するには人間と同じレベルの状況理解を必要とするため、LAWS の解釈が英国と同様に「汎用型 AI」に類似したレベルを想定していること。2 つ目は、現存兵器システムを自動兵器システムと評価し、将来の自律型兵器システムの使用を留保する立場をとったこと。3 つ目は、将来保有するであろう MHC が確保された自律型兵器システムは、完全自律ではないとして、米国と同様の立場をとったことである。このことから、LAWS に対する日本の立場は、英国と米国の両主張を包含したものと言えよう。

また、自律型兵器システムに必要とされる、MHC の解釈であるが、日本は「使用される兵器に関する情報を十分に掌握した人間による関与を確保すること」等をその一例として挙げている⁷²。

さらに、今後の LAWS・GGE の成果として、主要国を含む国際社会で広く共通認識を確保した上でルールについて合意するのが望ましいが、「自律（autonomous）」、「致死性（lethal）」及び「人間の関与の形態（the form of human control）」について意見の相違があるため、法的拘束力のある文書を直ちに実効的なルール枠組みとすることは困難であり、現状においては LAWS・GGE における議論を踏まえた成果文書が適切なオプションであるとしている⁷³。これらのことから、日本はどちらかと言えば法規制慎重派の立場を示しており、「特化型完全自律兵器」は規制の対象ではないという立場と思われる。

autonomous weapon systems,” *Humanitarian Law & Policy*, International Committee of the Red Cross (ICRC), August 2018, <http://blogs.icrc.org/law-and-policy/2018/08/29/im-possibility-meaningful-human-control-lethal-autonomous-weapon-systems/>, last visited October 5, 2018.

⁷¹ “Working Paper,” pp. 3-4.

⁷² Ibid., p. 4.

⁷³ Ibid., p. 7.

おわりに

LAWS とは何か。本稿では LAWS を明らかにするために、HRW らが指摘する「完全自律型兵器」いわゆる「殺人ロボット」を、AI の種類の観点から「特化型完全自律兵器」と「汎用型完全自律兵器」の 2 つに区分した。

そして、HRW などの「法規制推進派」は、完全自律型兵器の規制対象について、AI 自らキル・チェーンを実行することできる「汎用型完全自律兵器」と、人間が設計したプログラムに従ってキル・チェーンを実行する「特化型完全自律兵器」の両方を規制対象に含めようとしていることが明らかになった。反対に「法規制慎重派」は、より規制対象を縮小する方向に解釈し、「特化型完全自律兵器」を「自動兵器」又は単に「道具」と評価し、自律型兵器が規制対象とされるのは、兵器に道具性がなくなったとき、すなわち「汎用型完全自律兵器」に至った段階であるとして、両者の対立構造が明らかになった。

「特化型完全自律兵器」は、中国が指摘するとおり、既存の兵器や開発中の兵器と密接に関係しており、現在の特化型 AI（主に DNN）技術の脆弱性を克服することにより、開発することが可能と考えられるため、その技術を有さない国が規制に賛成することは当然の帰結であると言える（前掲注 46 参照）。このような対立構造の中で、自律型兵器に規制をかけることは、結局は日本を含めたロボット兵器先進国ばかりを規制することになりかねない。

また、いわゆる「ならず者国家」や「テロリスト」が自律型兵器を入手してしまう場合の危険性も無視できない。なぜなら AI 技術は主に民生技術であり、例えば飛行性能等の低いものであれば、ある程度自律性を有するドローン等は比較的安価な値段で購入できるからである⁷⁴。そして、自律型兵器にどのような規制がかけられようと、彼らはその使用に躊躇しない恐れがある。

したがって、テロリスト等の非国家主体を含め、自律型兵器に係る国際的な軍備管理や各国の開発動向について、我が国は常に注視する必要がある。

⁷⁴ NATO では、1,000 ドル以下で購入できる趣味用のドローンは、MQ-9 リーパーよりも自律していると評価している。Autonomous Systems, p. 6. この他に 2017 年 8 月現在、499 ドルで販売されている中国 DJI 社製 Spark は移動する物体を自律的に追跡し、障害物を避け、電池残量が少なくなると帰宅することができるところされている。Army of None, between p. 224 and p. 225.

30大綱にあるように、急速な少子高齢化や厳しい財政事情を抱える日本にとって、過去にとらわれない徹底した合理化なくしてかかる防衛力の強化は実現できない⁷⁵。そのような情勢のなかで、日本の平和を守り抜くためには省力化・省人化といった意義がある自律型兵器システムは必要である。MHCを十分に確保することを前提に、自律型兵器システムの研究開発を促進することは急務であると言える。また一方で、本稿で論じたように、自律型兵器の規制の在り方について様々な対立構造がある中で、我が国が不利益を被ることのないよう、我が国にとって望ましい国際的なルール作りが求められている。そのためには日本がイニシアティブを取り、国際社会の中で、特に自律型兵器システムに必要とされるMHCに対する共通理解の醸成を図っていく必要がある。

30大綱で述べられているとおり、「自衛隊が保有する知見・人材を活用しつつ、自律型致死兵器システム(LAWS)に関する議論を含む国際連合等による軍備管理・軍縮に係る諸活動に関与する」ことが、今までに求められているのである。

⁷⁵ 30大綱、2頁。