自爆するドローン IAI「ハロップ」(写真：AP/アフロ)

研究レポート

自律兵器の現状

2021-02-22
佐藤丙午（拓殖大学教授）

「安全保障と新興技術」研究会　第2号
「研究レポート」は、日本国際問題研究所に設置された研究会参加者により執筆され、研究会での発表内容や時事問題等について、タイムリーに発信するものです。「研究レポート」は、執筆者の見解を表明したものです。なお、各研究会は、「研究レポート」とは別途、研究テーマ全般についてとりまとめた「研究報告書」を公表する予定です。

自律兵器について

　自律兵器に関する議論には、幾つかの誤解が存在する。
　まず、自律兵器がAI兵器と呼ばれ、それは操作者の管理を離れ、AIが自律的に判断して攻撃を実施する兵器と理解されることである。この理解は、映画の『ターミネーター』シリーズに登場するロボット兵のような「兵器」の存在を想定する。別の議論では、自動兵器と自律兵器とが混同されて批判される。いわゆる自動兵器は、兵器の自動化(automation)の形で実装されており、珍しいものではない。これに対して、自律化された兵器は、兵器単体を指すのではなく、自律化されたシステムのもとで運用される兵器を意味する。このような兵器システムは、まだ開発途上であり、配備されていないとされる。
　特定通常兵器使用禁止制限条約(CCW)は、2014年より自律型致死無人兵器システム(LAWS)の規制に関する協議を進めており、2017年からは政府専門家会議(GGE)を開催している。CCWの議論でも、兵器が自律化された際の課題を検討している。兵器の自律化をめぐる議論では、システムの死活的機能(critical function)の自律化が、国際人道法などの国際規範に反する結果につながる可能性を考慮し、その技術的及び機能的な特徴を導き出そうとしている。すなわち、CCW-GGEの議論でも、「殺人ロボット」や「Slaughterbots」のような兵器の存在を前提としていない。

自律兵器の定義

　つまり、自律兵器の規制は、自律化された兵器システムの作動全体を俯瞰した上で、何を規制すれば国際人道法に適合した使用に抑えることが可能かという問題に帰着する。それをふまえると、規制方法としては、技術を規制するか、それとも兵器の使用環境を

限定するか(機能的な面からの規制)ということになる。ドイツの国際安全保障研究所(SWP)内に設けられた自律兵器規制に関する国際パネル(IPRAW)は、「デザインの管理(control by design)」と「用法の管理(control in use)」という二分法を提起しているが、国際社会の多くの議論でも、この二分法に基づく規制が検討されている。

まず技術面での定義では、もし自律化された兵器システムの死活的機能を欧州議会のように「照準(target)」と「攻撃(attack)に」限定したとしても、その兵器システムには、システムが兵器の作動環境を事前に特定し、変化に合わせて反映させる能力が必要になる。具体的には、データ収集センサー、収集されたデータを解析し行動を決定するハードウェアとソフトウェア(およびアルゴリズム)、他のエージェント(人間及び機械)との交流を可能とする通信技術(いわゆる「人間と機械のインターフェイス」)、そして行動を実施するアクチュエーター(一般的に攻撃兵器)などである。兵器としてイメージされるのは、最後のアクチュエーターになるが、自律兵器システムではその目的が破壊あるいは無力化であるかは問われない。

自律兵器を機能面から定義すると、「一旦作動した後は、自身で指令を判断して、周辺の情報を収集し、それに基づいて判断し、自律的に移動して敵を見つけ、攻撃を加える」機能を持つ兵器となる。人間と機械の指揮命令関係では、人間が「ループ」の外に位置することが前提となり、機械自体の意思決定は機械学習等で自己進化を遂げる。自律兵器がAI兵器と誤解されるのは、自己進化のために人工知能(AI)を利用することに由来する。現実には、人工知能技術の発展は不透明であり、兵器の使用に耐えうるような人工知能が数年で開発される可能性は低いとされる。

「シナリオ」に基づく自律兵器の定義

CCW-LAWS-GGEなどの議論でも指摘されているが、自律兵器を技術面と機能面のみで定義するのは困難である。

兵器システムは、固有の作動状況が存在し、それに応じて技術及び機能面での特性も変化する。このため、国際社会では、兵器の運用状況ごとに複数のシナリオを用意し、それぞれにおいて自律兵器がもたらす国際人道法上の課題を検討している。各国政府、IPRAWやスウェーデンのストックホルム国際平和研究所(SIPRI)などの世界各地のシンクタンク、および市民社会集団がこの試みに取り組んでいる。ただし、シナリオごとに課題は抽出できたとしても、それを執行力が伴う規制措置へと結びつけることが可能かどうかについて、懐疑論が存在する。

懐疑論が生じる一つの理由は、国際社会で検討されるシナリオは、それぞれの国家が直面する自律兵器の課題を反映したものになるためである。もちろん安全保障問題には価値中立的なものはなく、それぞれの関心や、それまでに各国が直面した安全保障課題を反映したものになるのは避けられない。このため、シナリオに基づく自律兵器の定義問題では、国際社会で議論を主導する諸国に比べ、他の国家の関心が十分に反映されない可能性も指摘されている。

ただし、シナリオ中心の自律兵器の定義では、この兵器システムの何が問題かが明確になる。国際社会で議論されているシナリオでは、兵器システムの作動環境(都市戦闘、公海・海底、領空防衛等がシナリオとして登場するケースが多い)、人間と機械のコミュニケーションの程度・内容、そして作動環境における法的適合性の担保などが変数として用いられる。

同様のプロセスを経て規定された米国の自律兵器に対する政策では、①自律化機能を持つ兵器が、特定の標的や集団を攻撃することに関する司令官や操作員の意図を効率的かつ正確に、そして信頼できる形で反映させることができる場合、②自律化機能を有する兵器が、指揮官や操作員の意思決定に際して、彼らが意図する攻撃対象に関する情報を伝えることができる場合、そして③兵器システムを起動させた際に、司令官や操作員が知らなかった特定の標的を選択し攻撃する場合、に使用が許容されるとしている[1]。

自律兵器システムの実情

これらの定義論争とは無関係に、自律兵器の開発及び各国による採用は拡大している。自律兵器については、新規の兵器システムとしてではなく、既存の兵器システムに自律機能を付与する形で発展しており、レガシーシステムの置き換えもしくは補強という形態で進む。自律化機能という意味では、駆動(mobility)、標的(targeting)、インテリジェンス、インター・オペラビリティ(通信・交流)、そして自己管理(health management)などで進展が見られる。

それぞれの分野では、いくつかの焦点分野が存在する。駆動では、追跡、自動航法、離発着などが注目され、標的では、自動標的認識システムが重要であり、そのシステムはアルゴリズムの訓練とテスト、機械学習の進展などを管理する。さらに、インテリジェンスでは、もの、行動、場所に加え、意図まで解析可能な探知能力と、収集する情報の質のコントロールが重要である。インター・オペラビリティでは、行動調整、広域の監視行動、分散攻撃、そして人間と機械のコラボレーションが重要である。さらに、自己管理では、自己充電や自己修繕などが重視されている。

米国は、数年おきに『無人システムの統合ロードマップ』を発表しており、自律兵器システムに必要な技術条件を提示すると共に、兵器の構想を示している。2017～2042会計年度(FY2017-2042)のロードマップでは、それらに加え、4つの原則(相互運用性、自律化、ネットワークの安全、人間と機械のコラボレーション)と、必要な技術要件を規定している。それらは、機械認識・推論とインテリジェンス(MPRI)、自律システムのチーミング(STAS)、人間と自律システムのインターアクションとコラボレーション(HASIC)、そして試験・評価・検証・査証(TEVV)としている。米軍では、これら条件を短期、中期、長期と分けて、それぞれに必要な措置を規定している。

ただし、すでに実用化された自律化された兵器及びその関連システムは存在する。SIPRIによると、死活的な機能に自律機能がある無人兵器、死活的機能は自律化されていないものの、他の領域に自律化機能が組み込まれている無人兵器、さらには無人化された非兵器の軍事システム(インテリジェンス、ISR、兵站など)を含め、2017年のデータベースで381種類の自律兵器が存在するという。

その中で実用化されているものとしては、まず防空システムとして、米国のPhalanx、ロシアのS-400 Triumfなどが実戦配備されている。また、レーダー、赤外線(IR)、仮想現実(UV)などを利用し、ごく近接でハードキル(角度を変える・貫通力を削減・直前で爆発・撃墜)、もしくはソフトキル(無効化)で防御するアクティブ防護システム(APS)は、フランス、ドイツ、イスラエル、イタリア、韓国、ロシア、南アフリカ、スウェーデン、米国が採用している。さらに、韓国とイスラエルは、ロボットSentryシステムを実用化している。これらに加え、LAWSの実例と批判されることが多い、誘導兵器と無人戦闘システムの組み合わせから発生した滞空兵器(Loitering Weapon)は、ドイツとイスラエルが開発・実用化している。

自律兵器の展望

以上のように、兵器の自律性は進展しており、今後もこの傾向は続く。しかし、人工知能を活用する機械学習の軍事転用の可能性には疑問がもたれており、韓国のDODAAM's Super aEgis IIが、自律走行をして攻撃を加えるような未来は当面到来しないだろう。

自律兵器のシステムには、データの入手が死活的に重要な意味を持つ。そして、状況把握能力が飛躍的に向上する可能性を考慮すると、兵器システムの開発が急速に進む可能性を排除することは賢明ではない。ただし、国際人道法に対する規範は強固であり、戦場の無人化が進むかどうかは疑問である。

CCW-LAWS-GGEでは、2019年8月の会合で11の指導原則に合意している[2]。この指導原則は、各国及び市民社会のコンセンサスを結晶化したものであり、これまで自律兵器をめぐる問題において必要と主張とされてきたものが、ほぼすべて含まれている。この指導原則は、実効力ある規制へと発展させるうえで、操作化する必要が指摘されている。残念ながら、国際社会では2020年の議論は低調であったが、2021年以降の議論では一定の成果を出すことが期待されている。

以上

---

[1] 2019年3月のCCW-LAWS-GGEに提出された米国の作業文書(CCW/GGE.1/2019/WP.5)より。Available at https://undocs.org/en/CCW/GGE.1/2019/WP.5

[2] Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons System, "Report of the 2019 session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems," 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, 25 Sep. 2019, CCW/GGE.1/2019/3. Available at https://undocs.org/en/CCW/GGE.1/2019/3