경북대학교 법학연구원『법학논고』
제93집 (2026.04) 425~474면.
Kyungpook Natl. Univ. Law Journal
Vol.93 (Apr 2026) pp.425~474.
http://doi.org/10.17248/knulaw..93.202604.425

자율무기체계 협약 논의의 쟁점과 전망[*] [**]

임 예 준[***]

〈국문초록〉

자율무기체계(Autonomous Weapon Systems)에 관한 현재 국제적 논의의 핵심은 기존 국제인도법의 적용과 준수 여부를 넘어, '새로운 법적 구속력 있는 문서'의 마련이다. 국제사회는 자율무기체계의 합의된 특성을 바탕으로 범위를 설정하고자 노력하고 있으며, 금지와 규제의 이원적 접근방식, 인간통제 원칙, 기계학습으로 인한 예측불가능성 등 여러 핵심 쟁점에 대해 문제의식을 공유하고 있다. 특히 무기체계의 자율성이 무제한일 수 없으며, 인간의 판단 및 통제, 그리고 책임을 유지해야 한다는 점에 대한 공감대를 형성하고 있다. 다만, 금지 및 규율의 범위와 문서의 형식에 관한 의견 조정이 용이하지 않을 것으로 보이는바, 자율무기체계 협약에 관한 국제사회의 논의가 어느 장(場)에서 전개될 것인지 짚어볼 필요가 있다. 이 논문은 자율무기체계 관련 협약 전망을 위해 현재 CCW 체제 내의 추가의정서 채택의 한계를 살펴보고, 유엔 총회를 통한 대안적 조약 협상 경로 모색의 필요와 협약 체결이 현실화되기 이전 연성법적 문서 채택의 의의를 살펴보았다.

진정한 인간통제는 완전한 자율무기체계가 일단 개발·사용된 이후에는 구현될 수 없다. 설령 충분한 관행이 축적되지 않은 선제적 규제라 하더라도, 조약의 제정은 유사한 이해를 가진 국가들의 협상과 정치적 타협을 통해 가능하며, 국제기구의 권고와 결의는 그 추진력을 제공할 수 있다. 2023년 이후 유엔 총회가 자율무기체계를 공식 의제로 다루기 시작했다는 점은 이러한 맥락에서 주목할 만하다. 포용성을 갖춘 유엔 총회는 공동의 목표를 설정하기 유리하며, 투표 방식의 의사결정을 기반으로 한다는 점에서 총의를 요구하는 CCW 체제에서의 추가의정서 채택보다 유리한 조건을 갖추고 있다. 인류 공동의 이익과 목표를 바탕으로, CCW 논의를 보완하는 수준을 넘어 보다 진전된 논의가 유엔 총회에서 적시에 전개되기를 바란다.

주제어 : 자율무기체계, 국제인도법, 인간통제, 예측가능성, 살상로봇

• 투고일 : 2026.03.28. / 심사일 : 2026.04.26. / 게재확정일 : 2026.04.26.

* 고려대학교 연구비에 의하여 수행되었음.

** 이 논문은 제44회 국제인도법 세미나(2025.11.21.) 발표문을 수정·보완한 것으로, 동일 제목의 발표문 요약본은「인도법논총」제45호(2025), 305-307면에 수록되어 있음을 밝힙니다.

*** 고려대학교 법학전문대학원 부교수(국제법)

Ⅰ. 서론

자율무기체계(Autonomous Weapon Systems)의 금지 및 규제를 위한 협약을 조속히 마련해야 한다는 국제사회의 요구가 높아지고 있다. 2023년 10월 5일 António Guterres 유엔 사무총장과 Mirjana Spoljaric 국제적십자위원회(International Committee of the Red Cross, 이하, 'ICRC') 총재는 자율무기체계에 관한 새로운 국제 규칙의 필요성을 강조하며, 자율무기체계의 금지와 규제에 관한 법적 구속력 있는 문서 채택을 위한 협상을 조속히 개시하여 2026년까지 완성할 것을 촉구하는 성명을 발표하였다.[1] 같은 달 유엔 총회 제1위원회에서는 「치명적 자율무기체계(Lethal Autonomous Weapons Systems)」가 단독 의제로 논의되었고, 이를 바탕으로 12월 28일 유엔 총회 결의 제78/241호가 채택되었다.[2] 동 결의는 자율무기체계로 인해 제기되는 도전과 우려를 시급히 다룰 필요가 있음을 강조하며, 회원국 간의 추가 논의가 이루어질 수 있도록 유엔 사무총장에게 관련 보고서를 마련하도록 하였다.[3] 이에 따라 2024년 7월 제출된 사무총장 보고서는 자율무기체계 발전에 관한 회원국 및 옵서버 국가, 국제기구 및 관련 시민사회단체의 견해를 담고 있다.[4] 사무총장은 인간의 통제가 책임성의 확보, 국제법 준수 및 윤리적 판단 보장을 위해 필수적이라는 광범위한 인식과 시급한 조치의 필요성을 확인하며, "인간의 통제나 감독 없이 작동하며 국제인도법을 준수할 수 없는 치명적 자율무기체계의 금지와 그 밖의 자율무기체계에 대한 규제"를 내용으로 하는 '법적 구속력 있는 문서'를 2026년까지 마련할 것을 촉구하였다.[5]

2024년 12월 2일 유엔 총회는 「치명적 자율무기체계」에 관한 결의 제79/62호를 166개국의 찬성으로 채택하였다.[6] 유엔 총회의 논의는 2014년부터 자율

---

1) UN Secretary-General, President of ICRC Jointly Call for States to Establish New Prohibitions, Restrictions on Autonomous Weapon Systems, SG/2264, Press Release (5 October 2023).

2) GA Res. 78/241, *Lethal Autonomous Weapons Systems*, UN Doc. A/RES/78/241 (28 December 2023), paras 1-3.

3) Ibid.

4) Report of the Secretary-General, Lethal autonomous weapons systems, UN Doc. A/79/88 (1 July 2024), paras 11-50. (이하, 'SG Report 2024')

5) Ibid., paras 89-90.

6) GA Res. 79/62, *Lethal Autonomous Weapons Systems*, UN Doc. A/RES/79/62 (2 December 2024). 166개국 찬성, 3개국 반대(벨라루스, 북한, 러시아), 15개국 기권(중국, 피지, 에스토니아, 라트비아, 리투아니아, 인도, 이스라엘, 이란, 니카라과, 폴란드, 루마니아, 사우디아라비아, 시리아, 터키, 우크라이나).

무기체계 관련 논의를 이끌어 가고 있는 「과도한 상해나 무차별한 영향을 초래하는 특정 재래식 무기의 사용 금지 또는 제한에 관한 협약(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)」(이하, '특정재래식무기협약', 'CCW') 체제 내 정부전문가그룹 논의의 교착상태와 한계를 반영한다. 실제 2024년 결의는 사무총장 보고서에 담긴 다양한 견해와 정부전문가그룹에서 충분히 논의되지 않은 사항의 포괄적 논의를 위하여, 2025년 공개 비공식협의(Open Informal Consultations) 소집을 결정하였다.[7] 비록 이틀이지만, 관련 시민단체와 전문가들은 공개 비공식협의 개최가 자율무기체계에 관한 새로운 조약 체결을 향한 모멘텀이 될 것이라고 보았다.[8] 나아가 자율무기체계에 관한 논의의 장을 제네바에서 뉴욕으로 옮김으로써, 대안적 조약 협상 경로를 모색할 수 있다는 기대를 낳았다.

그러나 2025년 5월 12일과 13일 실제 진행된 논의는 이러한 기대에 부응하지 못한 것으로 보인다.[9] 공개 비공식협의가 진행된 후 대한민국을 포함한 21개국은 공동성명을 통해 정부전문가그룹을 통한 CCW 체제 내에서의 논의가 가장 적합한 논의의 장이며, 유엔 총회를 비롯한 공개 비공식협의 절차는 단지 정부전문가그룹 임무 이행을 뒷받침하는 방식으로 소집되어야 한다고 목소리를 높였다.[10] 이후 2025년 12월 5일 채택된 「치명적 자율무기체계」에 관한 유

7) Ibid., para 7. CCW 체제 밖에서 관련 논의가 이뤄지는 것을 반대하는 국가들의 의견이 반영되어 2일로 제한되었다.

8) Benjamin Perrin, "Lethal Autonomous Weapons Systems & International Law: Growing Momentum Towards a New International Treaty", *ASIL Insights,* Vol.29, Issue 1 (January 24, 2025). Elizabeth Minor, "Opportunities after the UNGA Resolution on Autonomous Weapons: Moving Toward a New Treaty", Article 36 (December 22, 2024).

9) 공개 비공식협의는 뉴욕 유엔본부에서 96개국 이상의 정부 대표단과 다수의 시민단체가 참여하여 개최되었다. 오전 세션에서는 정부전문가그룹 의장의 논의 경과 브리핑과 유엔 사무총장 보고서(A/79/88)가 소개되었으며, 오후 세션에서는 법적, 인도적, 안보적, 기술적, 윤리적 차원의 고려사항을 주제로 한 전문가 브리핑과 질의응답 세션이 진행되었다. Open informal consultations on lethal autonomous weapons systems, held in accordance with GA resolution 79/62, May 12-13, 2025, https://unodaweb-meetings.unoda.org/public/2025-05/Programme%20ver.%202025-05-02.pdf (최종 접속일: 2026년 2월 10일).

10) Informal Consultations on Lethal Autonomous Weapons Systems, Joint Statement by Australia, Bulgaria, Canada, Czechia, Denmark, Estonia, Finland, France, Germany, Greece, Italy, Japan, Latvia, Lithuania, Luxembourg, Poland, Portugal, Sweden, Republic of Korea, Romania, and the United Kingdom, 12 May 2025, https://unny.mission.gov.au/unny/250512_Informal_Consultations_Lethal_Autonomous_Weapons_Systems.html (최종 접속일: 2026년 2월 10일)

엔 총회 결의 제80/57호는 3년 연속으로 채택된 결의이지만 조약 협상을 담지 못했으며, 공개 비공식협의의 추가 개최도 결정하지 않았다.[11] 더욱이 2025년 결의에는 2024년 반대표를 던진 러시아, 벨라루스, 북한에 더해 이스라엘, 미국, 부룬디가 반대표를 행사함으로써 이후의 논의 진전에 대한 부정적 전망을 안겨주었다.[12] 실제로 일부 국가는 기존 국제인도법 규칙으로 충분히 대응이 가능하다거나, 추가적인 규범 형성을 위한 협상이 시기상조라는 입장이다.[13] 나아가 자율무기체계의 규율 문제는 각국의 군사적 이해관계와 긴밀히 연계되어 있어, 필요성을 불문하고 향후 협약 발전의 가능성을 단정하기는 어렵다. 특히 자율무기체계를 실제로 개발하고 운용할 능력이 있는 국가들은 협약 체결을 통한 규제를 선호하지 않는다.[14]

이러한 상황을 종합할 때 자율무기체계에 관한 '법적 구속력 있는 문서'가 상기 요청과 같이 2026년까지 채택될 수 있을지는 의문이다. 그럼에도 불구하고, 모든 국가가 자율무기체계에 관한 규율의 시급성을 공감하고 있고, 다수의 국가가 새로운 법적 구속력 있는 문서 마련을 지지하고 있다는 점에서 가까운 장래에 협약 체결 논의가 본격화될 가능성을 반드시 낮게 볼 것은 아니다.[15] CCW 체제 내의 추가의정서 채택을 위한 체약국 간 논의뿐만 아니라, 자율무기체계가 유엔 총회의 정식 의제가 되었다는 점은 이를 방증한다. 비공식협의를 통한 개방성 및 포용성 차원에서는 비록 한 발 후퇴하는 모습을 보였지만, 전반적인 논의의 흐름은 자율무기체계가 내포하는 인도적, 인권적, 법적, 기술적 및 윤리적 위험이 되돌릴 수 없는 지점에 이르기 전에 규율할 필요가 있다는 공동의 문제의식을 반영하고 있다.[16] '새로운 규칙'의 필요성은 기존 국제인도법의 적용을 전제로 하면서도, 기존 체계만으로는 포섭하기 어려운 자율무기체계의 특성을 인정하는 데에서 비롯된다. 다만, 규율의 범위와 문서의 형식에 관한 의견 조정이 용이하지 않을 것으로 보이는바, 자율무기체계 규율에 관한

11) GA Res. 80/57, Lethal Autonomous Weapons Systems, UN Doc. A/RES/80/57 (5 December 2025).

12) 2025년 12월 5일 '치명적 자율무기체계'에 관한 유엔 총회 결의는 찬성 164표, 반대 6표(벨라루스, 부룬디, 조선민주주의인민공화국, 이스라엘, 러시아, 미국), 기권 7표(아르헨티나, 중국, 이란, 니카라과, 폴란드, 사우디아라비아, 터키)로 채택되었다.

13) SG Report 2024, paras 72-73.

14) Kelley M. Sayler, Defense Primer: U.S. Policy on Lethal Autonomous Weapon Systems, January 2, 2025, CRS In Focus, https://www.congress.gov/crs-product/IF11150. (최종 접속일: 2026년 2월 10일)

15) 유엔 사무총장 보고서에 나타난 바와 같이, 다수의 국가는 자율무기체계를 규율하기 위한 새로운 '법적 구속력 있는 문서'의 마련을 지지하고 있다. SG Report 2024, paras 69-71.

16) SG Report 2024, paras 16-17.

국제사회의 논의가 어떠한 쟁점을 중심으로 어느 장(場)에서 전개될 것인지 짚어볼 필요가 있다. 이를 위하여 이하에서는 자율무기체계에 관한 국제사회의 논의 발전을 전반적으로 살펴보고(II), 자율무기체계에 관한 주요 쟁점을 검토한다(III). 다음으로 자율무기체계 관련 협약 전망을 위해 현재 CCW 체제 내의 추가의정서 채택의 어려움을 살펴보고, 대안적 조약 협상 경로 모색의 필요와 협약 체결이 현실화되기 이전 연성법적 문서 채택의 의의를 살펴보도록 한다(IV).

II. 자율무기체계 관련 국제사회의 논의

1. 시민사회 및 유엔 인권특별보고관의 경고

자율무기체계에 대한 국제사회의 논의는 2010년대 초반 이른바 '킬러로봇(killer robot)'에 대한 우려를 계기로 본격화되었다. 2010년 자의적, 약식, 또는 초법적 처형에 관한 유엔 특별보고관(이하, '자의적 처형에 관한 특별보고관') 인 Philip Alston은 유엔 총회에 제출한 중간보고서에서 로봇 기술의 발전이 생명권과 자의적 처형에 미칠 잠재적 영향을 지적하였다.[17] Alston 특별보고관은 이러한 기술이 초래할 법적, 정치적, 윤리적 문제에 대해 국제사회의 시급한 논의가 필요함을 강조하면서, 안전기준의 수립, 신뢰성과 성능시험의 보장, 그리고 조사 및 책임성 확보의 필요성을 제기하였다.[18] 이어 특별보고관이 된 Christof Heyns는 2013년 유엔 인권이사회에 제출한 보고서를 통해 생명권을 침해하는 대인 자율살상로봇의 개발을 전면적으로 금지해야 한다고 주장하였다.[19] Heyns 특별보고관은 국가 차원의 모라토리엄, 국제인도법과 국제인권법 준수를 보장하기 위한 국내적 조치와 로봇 체계의 법적 검토와 투명성 확보를 강조하고, "법학, 로봇공학, 컴퓨터과학, 군사작전, 외교, 분쟁관리, 윤리, 철학 등 다양한 분야의 전문가"로 구성된 치명적 자율로봇에 관한 고위급 전문가 패널 소집을 권고하였다.[20]

---

17) Interim Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, Philip Alston, UN Doc. A/65/321 (2010), pp.10-12.

18) Ibid., pp.16-19.

19) Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, Christof Heyns, UN Doc. A/HRC/23/47 (2013), paras 109-112.

시민사회 역시 자율무기체계에 대한 국제적 논의를 촉진하는 데 중요한 역할을 하였다. 2012년 Human Rights Watch는 「A Losing Humanity: The Case against Killer Robots」 보고서를 통해 인간의 개입 없이 표적을 선정하고 공격할 수 있는 완전자율무기의 위험성에 대한 국제사회의 관심을 환기하였다. 해당 보고서는 완전자율무기가 국제인도법의 기본원칙을 준수하기 어렵고, 무력충돌 시 민간인 피해를 증가시킬 가능성이 크다고 지적하며, 법적 구속력 있는 국제 문서를 통해 개발, 생산 및 사용을 금지할 것을 촉구하였다.[21] 이어 2013년 Amnesty International은 유엔 인권이사회에 제출한 서면 성명을 통해 '완전자율로봇무기'가 인권에 미치는 영향을 국제사회가 충분히 검토해야 함을 강조하고, 인권 보호를 위해 해당 무기의 개발과 사용에 대한 모라토리엄이 필요하다고 주장하였다.[22] 이와 더불어 관련 전문가들 또한 자율무기체계의 개발과 확산이 군사기술혁명에 있어 화약과 핵무기 이상으로 중대한 영향을 미칠 것이며, 전쟁 수단 및 방식에 있어 혁신적인 변화를 가져오는 동시에, 인류의 존재를 위협할 것이라고 경고하였다.[23]

자율무기체계의 금지와 규제를 주장하는 주요 논거는 다음과 같다. 먼저 자율무기체계의 사용은 국제인도법 준수의 전제가 되는 예측가능성, 설명가능성, 추적가능성을 결여하고 있어 국제인도법의 기본원칙을 준수하기 어려우며, 국제인권법상 생명권을 침해할 우려가 있다.[24] 나아가 마르텐스 조항의 관점에서, 무기체계가 인간의 통제 없이 자율적으로 무력을 사용해 인간을 살상하게 되는 것은 기본적으로 공공양심의 명령에 부합하지 않는다는 점도 지적된다.[25] 또한, 공격에 관한 결정이 기계의 알고리즘에 의해 자율적으로 이루어질 경우, 책임 귀속의 불명확성이라는 법적 문제와 함께, 기계에 의한 살상이 인간 존엄성을 침해한다는 윤리적 우려도 제기된다.[26] 군사안보적 측면

---

20) Ibid., para 114.

21) Human Rights Watch & IHRC, *A Losing Humanity: The Case against Killer Robots,* (November 19, 2012), pp.3-5.

22) Amnesty International written statement to the 23rd session of the UN Human Rights Council (27 May – 14 June 2013) ACT 30/038/2013, 22 May 2013.

23) 임예준, "인공지능 시대의 전쟁자동화와 인권에 관한 소고 – 국제법상 자율살상무기의 규제를 중심으로 –", 「고려법학」 제92호 (2019), 267-268면. Maya Brehm, Defending the Boundary: Constraints and Requirements on the Use of Autonomous Weapon Systems under International Humanitarian and Human Rights Law (The Geneva Academy of International Humanitarian Law and Human Rights, May 2017), p.7.

24) 임예준, 위의 주, 268면, 279-285면.

25) 임예준, 위의 주, 285-286면.

26) 임예준, 위의 주, 286-291면.

에서는, 자율무기체계가 전쟁 수행방식을 근본적으로 변화시키고 무력사용의 한계점을 낮춤으로써, 무력충돌의 가능성을 높일 수 있다는 우려도 존재한다.[27] 이러한 우려는 2024년 사무총장 보고서에 수렴된 국가들의 견해에서도 확인된다.[28]

한편, 자율무기체계에 대한 국가들의 '인도적 우려'는 기존의 인권적, 윤리적, 법적 차원을 넘어 환경적 차원으로까지 확장되고 있으며, 자율무기체계의 개발과 운용 과정에서 발생하는 과도한 에너지 소비 및 탄소 배출 문제가 지적되기도 한다.[29] 나아가 최근의 논의는 기계학습 기반 인공지능(Artificial Intelligence, 이하 'AI')이 무기체계의 핵심 기능에 적용될 경우 발생할 수 있는 의도치 않은 결과에 주목하고 있다. 이는 일차적으로 악의적인 사이버 개입, 하드웨어 및 소프트웨어의 오작동, 부정확하거나 잘못 해석된 정보에 기반한 의사결정 등 기술적 위험에 대한 우려이지만, 더 나아가 알고리즘 편향(bias)이 기존의 권력 불균형을 심화시키고 소외된 집단에 불균등한 영향을 미침으로써, 분쟁 지역 여성과 아동 등 민간인에게 부수적 피해를 초래할 수 있다는 우려도 제기된다.[30] 자율무기체계가 인공신경망, 특히 심층신경망(deep neural network) 기반 알고리즘에 따라 작동하는 경우, 그 작동 과정의 투명성 및 예측가능성은 구조적으로 제한될 수밖에 없다. 이는 기존에 덜 조명된 문제이므로, 이하 주요 쟁점을 다루는 절에서 별도로 다루도록 한다.

2. CCW 체제에서의 논의

자율무기체계에 관한 주요 국제적 논의는 CCW 체제 내 「치명적 자율무기체계 분야의 신기술에 관한 정부전문가그룹(Group of Governmental Experts related to emerging technologies in the area of lethal autonomous weapon systems)」을 중심으로 진행되고 있다. 2013년 CCW 고위 당사국회의에서 체약국들은 「치명적 자율무기체계」를 공식 의제로 채택하고, 관련 신기술 문제를

27) Philip Alston, "Lethal Robotic Technologies: The Implications for Human Rights and International Humanitarian Law", *Journal of Law, Information and Science* Vol.21 (2011/2012), p.44. ICRC, Expert Meeting: Autonomous Weapon Systems, Technical, Military, Legal and Humanitarian Aspects, Geneva, Switzerland (26 to 28 March 2014), pp.17-18.
28) SG Report 2024, paras 16-50.
29) SG Report 2024, para 19.
30) SG Report 2024, para 44-46.

논의하기 위한 비공식 전문가 회의(Informal Meeting of Experts) 개최를 결정하였다.[31] 이 결정은 자율무기체계에 대한 국제사회의 우려를 다자적 군축체제 내 공식 의제로 편입시켰다는 점에서 중대한 전환점으로 평가된다. 이후 2014년부터 2016년까지 세 차례 비공식 전문가 회의를 거쳐, 2016년 CCW 검토 회의에서 체약국들은 정부전문가그룹을 구성하고, "협약의 대상과 목적의 맥락에서 치명적 자율무기체계 분야 신기술에 관한 여러 선택사항(options)을 검토하여 가능한 권고 사항을 찾고 합의"하도록 하였다.[32] 2017년 이래로 정부전문가그룹은 치명적 자율무기의 개념 및 특성, 관련 신기술로 인한 국제인도법상 과제, 자율무기체계의 개발, 배치 및 사용 과정에서의 인간-기계 상호작용 문제, 관련 기술의 잠재적 군사적 영향 등을 논의해 왔다.

2019년 정부전문가그룹은 그동안의 논의를 바탕으로 11개 지도원칙(Guiding Principles)을 채택하였다.[33] 이는 2018년에 채택된 '가능한 지도원칙(Possible

---

31) Meeting of the High Contracting Parties to the CCW, Final Report, CCW/MSP/2013/10 (16 December 2013), para 32.

32) CCW Fifth Review Conference, Report of the 2016 informal meeting of experts on lethal autonomous weapons systems, CCW/CONF.V/2 (10 June 2016), annex, para. 3. ("explore and agree on possible recommendations on options related to emerging technologies in the area of LAWS, in the context of the objectives and purposes of the Convention, taking into account all proposals－past, present and future")

33) Report of the 2019 session of the GGE, CCW/GGE.1/2019/3 (25 September 2019), Annex IV, Guiding Principles. (이하, 'GGE Report 2019') 11개의 핵심 내용은 다음과 같다.
(a) 국제인도법은 치명적 자율무기체계의 개발 및 사용 가능성을 포함하여, 모든 무기체계에 전면적으로 적용된다.
(b) 무기체계 사용에 관한 결정에 대한 인간의 책임은 유지되어야 하며, 책임은 기계로 이전될 수 없다. 이러한 점은 무기체계의 전체 수명주기 전반에서 고려되어야 한다.
(c) 인간-기계 상호작용은 다양한 형태로, 무기체계의 수명주기의 여러 단계에서 구현될 수 있다. 이러한 상호작용은 치명적 자율무기체계 분야 신기술을 기반으로 한 무기체계 사용이 적용 가능한 국제법, 특히 국제인도법에 부합하도록 해야 한다. 인간-기계 상호작용의 질과 범위를 결정할 때는 운용 환경과 무기체계 전체의 특성 및 성능을 포함한 다양한 요인을 고려해야 한다.
(d) CCW의 틀 내에서 새로운 무기체계의 개발, 배치 및 사용에 대한 책임은 적용 가능한 국제법에 따라 보장되어야 하며, 여기에는 해당 체계가 인간의 책임 있는 지휘 및 통제 체계 아래에서 운용되는 것을 포함한다.
(e) 국제법상 국가의 의무에 따라, 신무기 및 전쟁 수단 및 방법의 연구, 개발, 획득 혹은 채택할 때에는 그 사용이 일부 또는 모든 상황에서 국제법에 따라 금지되는지를 판단해야 한다.
(f) 치명적 자율무기체계 분야의 신기술을 기반으로 새로운 무기체계를 개발하거나 획득할 때에는 물리적 안전, 적절한 비물리적 안전장치(사이버 보안 등), 테러집단의 획득 가능성 및 확산 위험을 고려해야 한다.
(g) 위험평가와 완화조치는 모든 무기체계에 관한 신기술의 설계, 개발, 실험 및 배치의 전 과정에서 이뤄져야 한다.

Guiding Principles)'[34]에 인간-기계의 상호작용에 관한 항목을 추가한 것이다. 지도원칙의 핵심 내용은 치명적 자율무기체계 분야 신기술의 개발 및 사용에 있어 국제법의 적용 및 준수와 전 과정에서의 인간 책임 유지이다. 이를 위해 신무기에 대한 국제법상 국가의 의무를 강조하고, 인간-기계 상호작용을 통한 국제법 준수의 보장 및 치명적 자율무기체계 분야 신기술의 전 과정에 걸친 위험평가와 완화조치의 이행을 강조하고 있다. 지도원칙은 테러집단으로의 확산 위험을 경고하는 한편, 관련 논의가 인공지능 기술의 평화적 이용을 위한 논의를 저해해서는 아니 됨을 명시하고 있다. 전반적으로 지도원칙은 치명적 자율무기체계 관련 신기술을 전면적으로 금지하는 방향이 아니라, 인간의 통제와 책임 유지, 확산 방지 및 안전 확보를 통한 규제의 방향을 제시하고 있다.

비록 원론적 논의에 그치고 방법론적 구체성이 부족하다는 지적도 있으나,[35] 2019년 지도원칙은 자율무기체계 관련 신기술이 제기하는 위험을 국제법적으로 관리하기 위한 최소한의 합의를 반영한 것으로, 단순한 정치적 선언의 수준을 넘어 일정한 '규범적 합의'로 나아갔다는 점에서 의의가 있다.[36] 국가 간 규제 필요성에 대한 인식 차이와 제도적 접근방식의 이견에도 불구하고, 지도원칙은 자율무기체계가 제기하는 위험과 과제를 이해하고 대응하기 위한 기본적 틀을 제공한다.[37] 실제로 2019년 지도원칙은 프랑스와 독일이 주도한 다자주의 연합(Alliance for Multilateralism)의 지지를 받아 정치선언의 형태로 재구성되었으며[38], 이후 유엔 총회 결의 및 각국 정책 논의에서 주요 참조 기

(h) 치명적 자율무기체계 분야의 신기술 사용은 국제인도법 및 기타 국제법상 의무의 준수를 보장하는 방향으로 고려되어야 한다.
(i) 정책 조치 마련에 있어 치명적 자율무기체계 분야의 신기술을 의인화(anthropomorphize)해서는 안 된다.
(j) CCW 맥락에서의 논의 및 잠재적 정책 조치는 지능형 자율기술의 평화적 이용과 기술 진보를 저해해서는 안 된다.
(k) 치명적 자율무기체계 분야의 신기술 문제를 다루는 데 있어, CCW는 협약의 목적과 취지의 맥락에서 적절한 틀을 제공한다. 협약은 군사적 필요성과 인도적 고려 사이의 균형을 달성하려는 것을 목표로 한다.

34) Report of the 2018 session of the GGE, CCW/GGE.1/2018/3 (23 October 2018). (이하, 'GGE Report 2018')
35) Article 36, Critical Commentary on the "Guiding Principles", Policy Note, November 2019, pp.1-4.
36) Laura Bruun, "The Group of Governmental Experts on Lethal Autonomous Weapon System", in *Conventional Arms Control and Regulation of New Weapon Technologies, Non-Proliferation, Arms Control and Disarmament*, 2020, Chapter 13, p.518.
37) Ibid., pp.518-519.

준으로 활용되고 있다.

2019년 11개 지도원칙을 채택하면서, 정부전문가그룹은 자율무기체계에 관한 향후 논의가 정책적으로 선택 가능한 여러 방향으로 발전할 수 있음을 전제로 하였다.[39] 즉, 법적 구속력을 갖는 문서의 채택, 정치적 선언의 채택, 원칙이나 모범 관행(best practice)과 같은 문서를 통한 기존 국제법상 의무의 명확화 등 다양한 접근이 가능하다는 점을 열어두었다.[40] 나아가 정부전문가그룹은 2019년 보고서를 통해 향후 지도원칙을 추가로 발전시키거나 구체화할 수 있으며, 이를 토대로 '치명적 자율무기체계 분야의 신기술에 관한 규범적 및 운영적 틀의 명확화, 검토 및 발전'을 권고하였다.[41]

2019년 지도원칙을 바탕으로 한 2020년 논의의 핵심 쟁점은 국제법 준수를 어떻게 보장할 것인지, 책임 공백을 어떻게 방지할 것인지, 그리고 자율무기체계를 국제법에 부합하도록 개발 및 운용하기 위해 어떠한 유형과 수준의 인간-기계 상호작용이 필요한지에 관한 것이었다.[42] 자율무기체계를 규제하거나 금지하기 위해 새로운 국제법 제정이 필요한가 하는 문제도 제기되었다.[43] 그러나 이후 일부 국가가 새로운 법적 문서에 대한 논의 자체를 차단하거나 지연시키고, 2021년에는 권고를 담은 보고서 채택이 무산되면서, CCW 체제를 벗어나 별도의 장에서 자율무기체계 관련 협약 논의를 진행해야 한다는 목소리가 높아졌다.[44]

교착상태에 있던 CCW 체제 내 정부전문가그룹의 논의는 2023년 고위당사국회의가 "문서의 법적 성격을 미리 규정하지 않은 채, 치명적 자율무기체계 분야의 문서에 포함될 수 있는 요소를 총의에 따라 검토하고 문안화"하라는

38) Declaration by the Alliance for Multilateralism on Lethal Autonomous Weapons Systems, https://www.diplomatie.gouv.fr/en/french-foreign-policy/france-and-the-united-nations/multilateralism-a-principle-of-action-for-france/alliance-for-multilateralism/ (최종 접속일: 2026년 2월 10일)

39) GGE Report 2019, p.7. 이 부분에 대해서는 합의를 이루지 못했기 때문이다.

40) Alexander Blanchard & Netta Goussac, *Towards Multilateral Policy on Autonomous Weapon Systems,* SIPRI (September 2025), p.5.

41) GGE Report 2019, p.7, para 26.

42) Bruun, *supra* note 36, p.520.

43) Ibid.

44) Ray Acheson, "Editorial: From "Constructive Ambiguity" to Unambiguous Destruction", *CCW Report*, Vol.9, No.9 (2021), p.1. Human Rights Watch & IHRC, *An Agenda for Action: Alternative Processes for Negotiating a Killer Robots Treaty* (2022), pp.5-14. Human Rights Watch, Killer Robots: Negotiate Treaty in New Forum (November 10, 2022).

임무를 부여하면서 다시 활성화되었다.[45] 이에 따라 2024년부터는 진행형 문안인 'Rolling Text' 작성이 정부전문가그룹 논의의 중심이 되었다. Rolling Text는 협상의 경과에 따라 수정 및 갱신되는 작업 문서로, 아직 합의된 최종 문안이나 공식 초안이라고 볼 수 없다.[46] 정부전문가그룹 의장은 Rolling Text를 "잠정적으로 합의에 도달한 요소를 담은 검토 문서"로 정의하면서, 향후 협약이나 기타 조치로 발전할 가능성을 열어두되, 그 법적 성격을 미리 단정하지 않는다고 밝혔다.[47]

정부전문가그룹은 2024년 11월 회기에서 다섯 가지 핵심 영역-[I] 치명적 자율무기체계의 특성; [II] 국제인도법의 적용 및 해석-인간통제; [III] 국제인도법 준수를 보장하기 위한 금지 및 규제 조치; [IV] 자율무기체계의 사용 전·중·후 단계에서 취해야 할 조치; [V] 개발 및 사용에 대한 책임과 책임규명 문제-에 대한 Rolling Text를 작성하였다. Rolling Text는 2025년 3월과 9월의 공식 회의뿐만이 아니라, 5월, 12월에 개최된 비공식 회의를 거치며 수정 및 보완되고 있다.

정부전문가그룹이 작업 중인 Rolling Text는 국제사회의 협상 틀로서 일정한 진전을 이루고 있으나, 시민단체와 전문가들은 현재 논의의 근본적 한계를 지적한다. 첫째, 대인 자율무기체계에 대한 명시적 금지가 논의되지 않고 있다는 점이다. Rolling Text가 다양한 제한 조항을 포함하고 있음에도, 인간을 직접 표적으로 삼는 체계의 금지를 다루지 않는다는 것은 중대한 공백이다.[48] 인명 표적형 자율무기체계는 시스템 설계의 편향으로 인한 차별 위험, 국제인도법 및 국제인권법상의 구체적 법적 쟁점, 그리고 광범위한 윤리적 고려사항을 포괄적으로 검토할 필요가 있다.[49] 둘째, 인간통제 개념이 국제인도법의 틀

45) Meeting of the High Contracting Parties to the CCW, Final Report, CCW/MSP/2023/7 (23 November 2023), para 20.

46) 기본적으로 CCW는 총의를 원칙으로 하므로, 모든 조항에서 만장일치가 이루어지기 전까지는 어떠한 내용도 최종 합의로 간주될 수 없다. 그러나 Rolling Text는 논의의 진전을 위해 마련된 잠정적 합의를 반영함으로써, 합의를 향한 최소한의 공통분모를 도출하고 부분적·잠정적 합의가 형성되는 과정을 가시화한다. 이러한 접근은 CCW 체제의 구조적 한계인 총의 규칙에 따른 교착상태를 완화하고, 논의 과정을 제도화하여 시민사회 등 외부 행위자들이 변화의 흐름을 추적할 수 있게 한다.

47) Blanchard & Goussac, *supra* note 40, p.6.

48) Human Rights Watch, Statement on the CCW GGE Consultation on Lethal Autonomous Weapons Systems (May 28, 2025).

49) Elizabeth Minor & Richard Moyes, Key comments on the current CCW GGE 'rolling text' on autonomous weapons, February 17, 2025, https://article36.org/updates/key-comments-on-the-current-ccw-gge-rolling-text-on-autonomous-weapons/ (최종 접속일: 2026년 2월

에 한정되어 있다는 점이다. 자율무기체계의 운용은 평시나 비전시 상황에서도 발생할 수 있음에도 불구하고, Rolling Text는 국제인권법적 고려, 윤리적 요소, 안보적 맥락을 배제한 채 무력분쟁 상황에서의 국제인도법 준수 여부만으로 인간통제를 평가하고 있다.[50] 셋째, AI 기반 감시나 국경통제 등 비군사 영역에서의 활용을 간과하고 있다는 점도 지적된다. 실제로 이러한 맥락에서 인권 침해 및 자의적 살상행위의 위험이 더 크게 제기될 수 있다.[51] 시민사회와 학계는 논의의 초점을 국제인도법에만 한정하지 않고, 국제인권법, 비전시 상황, 치안 및 국경통제 영역에서의 AI 활용 등으로 확대할 필요가 있다고 주장해 왔다.[52] 이러한 비판은 CCW 체제 내에서 논의가 진행될 경우 자율무기체계에 관한 규범 형성 범위가 제한될 수 있음을 보여준다.

3. 유엔 총회 결의 및 공개 비공식협의

서론에서 언급한 바와 같이, 유엔 총회는 2023년 12월 28일 '치명적 자율무기체계'에 관한 결의 제78/241호를 최초로 채택하였다. 동 결의는 CCW 체제 내 정부전문가그룹을 통해 관련 사안에 대한 이해를 더욱 심화시켜 나갈 것을 촉구하는 한편, 유엔 사무총장에게 회원국 간의 추가 논의를 위한 보고서 마련을 요청하였다.[53] 이어 2024년 12월 2일 채택된 결의 제79/62호는 정부전문가그룹의 기존 임무 완수를 독려하는 한편, 정부전문가그룹에서 충분히 다루어지지 않은 관련 사항들의 포괄적 논의를 위하여 2025년 공개 비공식협의를 소집하기로 결정하였다.[54] 동 결의는 모든 유엔 회원국 및 옵서버 국가, 국제 및 지역기구, ICRC, 비정부기구, 학계, 산업계의 참여를 허용하도록 명시하였으며, 정부전문가그룹 의장에게도 협의 참여를 요청하였다.

CCW 내 논의의 장기적 정체를 비판해 온 인권단체와 시민사회는 논의의 중심이 유엔 총회로 전환된 것을 환영하였다. 이들은 다양한 이해관계자가 참여할 수 있는 포용적 논의의 장인 공개 비공식협의가 향후 관련 협약 논의의 원동력이 될 것으로 기대하였으며, 공개 비공식협의를 CCW 체제 내 정부전문

10일)

50) Ibid.

51) Ibid.

52) Bonnie Docherty, "Autonomous Weapon Systems and Threats to Human Rights", *WS Diplomacy Repor*t, Vol.2, No.1 (May 7, 2025). p.4.

53) GA Res. 78/241, *supra* note 2, paras 1-3.

54) GA Res. 79/62, *supra* note 6, para 7.

가그룹 논의의 구조적 한계를 보완할 잠재적 경로로 평가하였다.[55] 반면, 협의 직후 대한민국을 포함한 주요 이해국들은 공동성명을 통해 “정부전문가그룹 외부의 절차는 군사적, 기술적, 법적 전문성이 요구되는 복잡한 문제의 해결을 오히려 지연시킬 우려가 있다”고 지적하였다.[56] 이들은 자율무기체계에 관한 논의는 여전히 CCW 체제 내 정부전문가그룹을 중심으로 이루어져야 하며, 유엔 총회 및 비공식협의 절차는 “정부전문가그룹의 임무 이행을 보완하는 수준에서 한정적으로 진행되어야 한다”는 입장을 밝혔다. 즉, 대한민국을 포함한 자율무기체계 주요 이해국들은 관련 논의가 기존과 같이 CCW 체제 내에서 이루어져야 한다는 입장을 견지하고 있다. 실제 2025년 12월 5일 채택된 결의 제80/57호는 공개 비공식협의의 추가 개최를 결정하지 않았다. 이는 향후 협약 협상의 개시 여부뿐만 아니라, 협상이 진행될 제도적 틀의 선택을 둘러싼 국가 간 입장 차이를 보여준다.

4. 새로운 ‘법적 구속력 있는 문서’의 요청

1) 자율무기체계 규율의 법적 근거

치명적 자율무기체계의 개발 및 사용 가능성을 포함하여, 모든 무기체계에 국제인도법이 전면적으로 적용된다는 원칙에 대해서는 이견의 여지가 없다. 실제 공식 논의에서 국제인도법이 자율무기체계 관련 신기술에 적용되지 않는다고 주장한 국가는 전무하다.[57]

현재 자율무기체계를 특정하여 금지하거나 제한하는 조약은 존재하지 않는다. 그러나 국제인도법상 국가들은 전쟁 방식의 적법성 및 신무기의 적법성을 검토해야 할 의무를 진다.[58] 1949년 8월 12일자 제네바 제협약에 대한 추가 및 국제적 무력충돌의 희생자 보호에 관한 의정서(이하, ‘제네바협약 제1추가의정서’) 제35조 제1항은 충돌 당사국의 전투 수단 및 방법 선택권이 무제한적이지 않음을 기본원칙으로 선언하고, 제2항에서 “과도한 상해 및 불필요한 고통을 초래할 성질의 무기, 투사물, 물자, 전투 수단을 사용”이 금지됨을 확인하

55) Human Rights Watch, Killer Robots: UN Vote Should Spur Treaty Negotiations, Urgent Action Needed to Address Autonomous Weapons Systems, December 5, 2024.
56) Informal Consultations on Lethal Autonomous Weapons Systems, Joint Statement, *supra* note 10.
57) Article 36, Critical Commentary on the “Guiding Principles”, Policy Note, November 2019, pp. 1-4.
58) 이하 내용은 임예준, 앞의 주 23), 278-279면 .

고 있다. 동 의정서 제36조는 신무기 · 전투 수단 또는 방법의 연구 · 개발 · 획득 및 채택에 있어 그 사용이 동 의정서 및 적용 가능한 국제법의 다른 규칙에 의하여 금지되는지 여부를 결정할 의무를 체약 당사국에 부과하고 있다. 이 조항은 무기가 개발 · 획득되거나 국가 무기고에 편입되기 이전에 적법성을 판단함으로써 국제법을 위반하는 무기의 사용을 예방하고 제한하는 것을 목적으로 한다.[59] 이러한 적법성 검토는 무기 자체뿐만 아니라 무기가 사용되는 방식에도 적용된다.[60]

특정한 무기나 전쟁 방식이 명시적인 조약이나 관습규칙으로 금지되지 않은 경우에도, 인도원칙(principle of humanity)과 공공양심(public conscience)의 명령에 따라 금지될 수 있다.[61] 이러한 내용을 담은 마르텐스 조항(Martens Clause)은 1899년 헤이그협약(II) 및 1907년 헤이그협약(IV)의 전문(前文)에 언급되었으며, 제네바협약 제1추가의정서 제1조 제2항에도 명시되어 있다. 동 조항에 따르면, 의정서 또는 다른 국제협정의 적용을 받지 아니하는 경우에도 민간인 및 전투원은 확립된 관습, 인도원칙 및 공공양심의 명령으로부터 연원하는 국제법 원칙의 보호와 권한 하에 놓인다. 마르텐스 조항은 국제인도법상 '인도'와 '공공양심'이라는 개념을 명시함으로써 윤리적 문제에 관한 국제인도법의 논의를 연결해주는 고리가 된다.[62]

자율무기는 발포와 같은 전통적인 공격수단을 활용할 수 있으나, 인공지능 기반의 '자율성'을 갖춘다는 점에서 기존 국제법으로 규제되지 않는 신무기에 해당한다. 자율무기가 신무기로서 적법하기 위해서는 모든 무기, 전투 수단 및 방법에 적용되는 국제인도법의 일반적 규칙과 특정 무기 또는 전투 수단을 금지하거나 그 사용방식을 제한하는 국제인도법의 특정 규칙에 부합하여야 한다.[63] 아울러 이러한 신무기는 배치되기 이전에 인도의 원칙과 공공양심의 명령에 부합하는지 여부를 검토하여야 한다.[64] 그러나 인공지능 기반 자율무기 체계는 기존 신무기와는 차원이 다른 문제를 제기한다. SIPRI 보고서는 자율

59) ICRC, 신무기, 전투 수단 및 방법의 적법성 검토에 관한 지침: 1977년 제1추가의정서 제36조 이행조치(2006), 3면.
60) ICRC, 위의 주, 7-8면.
61) 임예준, 앞의 주 23), 278면.
62) Neil Davison, "A legal perspective:Autonomous weapon systems under international humanitarian law", in UN, *UNODA Occasional Papers No. 30, November 2017 Perspectives on Lethal Autonomous Weapon Systems* (January 2018), p.8.
63) ICRC, 신무기, 전투 수단 및 방법의 적법성 검토에 관한 지침, 8면.
64) ICRC., 위의 주, 14면. Robin Geiss, *The International-Law Dimension of Autonomous Weapons Systems* (Friedrich Ebert Stiftung, October 2015), p. 11.

무기체계를 다른 무기수단 및 전투 방법과 구별하는 사회 기술적 특성으로 다음 세 가지를 제시한다. 첫째, 사전에 프로그래밍 된 표적 프로필과 기술적 지표를 기반으로 작동하며 센서와 소프트웨어를 통해 인식기능을 수행한다는 점; 둘째, 사용환경에 의해 부분적으로 작동하므로 무력행사 결정을 기존 무기보다 이전 시점에 설정할 수 있다는 점; 셋째, 인간 운용자의 감독 및 시스템 중단 가능성이 유지될 수 있으나, 기본 작동방식이 인간의 입력 없이도 목표를 식별, 선정하거나 공격할 수 있도록 설계되어 있다는 점이다.[65]

2) 새로운 규칙과 문서의 필요

자율무기체계 규율에 관한 쟁점 중 하나는 새로운 규칙과 법적 구속력 있는 문서를 마련해야 하는가, 아니면 기존 국제법으로 충분히 규율할 수 있는가이다. 이 문제에 대해서는 정부전문가그룹의 초기 논의에서부터 국가 간 견해차가 존재하였다.[66] 일부 국가는 자율무기체계의 국제인도법 준수 메커니즘을 강화하는 것으로 충분하며, 새로운 규칙의 제정이 필요하지 않다는 입장이다. 예컨대, 자율무기체계 논의 초기인 2013년 영국은 기존 국제인도법 규칙으로 자율무기체계의 규제가 가능하다고 보고, 새로운 국제 규칙을 통한 금지보다 효과적 통제에 방점을 두었다.[67] 2022년 정부전문가그룹에 제출된 서면 제안서를 보면, 자율무기체계를 직접 개발할 가능성이 있는 국가군은 주로 원칙 및 모범 관행을 통한 기존 국제법 규칙의 구체화 등 점진적 접근방식을 선호하는 경향을 보인다.[68] 가령 일본은 2024년 사무총장 보고서에 수록된 답변서에서 법적 구속력이 있는 문서 도출에 부정적임을 명시적으로 밝히면서, 실효성 있는 규칙이 중요하다고 강조하였다.[69] 대한민국은 조약의 필요성이나 이에 대한 반대 입장을 명확히 밝히지는 않았으나, 별도의 조약 체결에는 소극적인 입

65) Laura Bruun, *Towards a Two-Tiered Approach to Regulation of Autonomous Weapon Systems: Identifying Pathways and Possible Elements*, SIPRI, August 2024, p.2.

66) Bruun, *supra* note 36, p.518. 정부전문가그룹의 논의는 일관되게 기존 국제인도법이 자율무기체계에도 적용된다는 점을 확인했지만, 기존 법규가 자율무기체계 규제에 있어 충분한지는 합의를 이루지 못하고 있다. 다수 국가는 “새로운 규범적 문서가 필요하다”고 보았으나, 미국, 러시아, 이스라엘 등은 “기존 규칙으로 충분하다”며 새로운 협약 체결에는 반대했다.

67) Article 36, Killer Robots: UK Government Policy on Fully Autonomous Weapons, April 2013, p.5.

68) 호주, 캐나다, 일본, 한국, 영국, 미국은 “Principles and Good Practices on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems”을 제안하였다. https://reachingcriticalwill.org/disarmament-fora/ccw/2022/laws/documents (최종 접속일: 2026년 2월 5일).

69) SG Report 2024, Annex I, p.67.

장으로 분류할 수 있다.[70]

반면, 자율무기체계를 직접 개발 가능성이 낮은 국가군은 법적 구속력 있는 문서, 즉, 의정서의 채택을 지지하는 경향을 보인다.[71] 물론 이는 기존 국제인도법의 일반 원칙만으로는 자율무기체계의 기술적, 운용적 복합성을 충분히 규율하기 어렵다는 인식을 바탕으로 한다. 2024년 사무총장 보고서에 나타난 국가들의 견해를 살펴보면, 공동 답변을 제출한 이베로아메리카 국가군(Ibero-American States)[72]과 CCW 체약국 중 입장을 같이한 16개국[73], 오스트리아[74], 아일랜드[75], 키리바티[76], 말라위[77], 네덜란드[78] 등이 새로운 법적 구속력 있는 문서의 마련을 명시적으로 지지하였다. 이들은 국제법상의 공백을 보완하고, 국가별로 상이한 규제 조치로 인한 규범의 파편화를 방지할 필요가 있다는 점에서 새로운 규칙 마련을 지지한다.[79]

새로운 국제 규칙의 필요성에 관한 논의는 기존 국제법 적용의 한계에서 비롯된다. 자율무기체계에 대해 기존 국제법, 특히 국제인도법이 적용된다는 점

---

70) 대한민국은 자율무기체계 규율에 있어 기존 국제인도법 준수를 핵심 기준으로 삼아, 그 본질상 국제인도법에 부합할 수 없는 체계에 대해서는 금지를 인정하는 한편, 그 외 체계에 대해서는 위험 완화 조치를 통한 규제를 강조하는 실용적 접근을 취한다. 특히 인간의 개입이 국제인도법 준수의 필수 요건은 아니며, 구별, 비례성, 예방조치 원칙의 충족 여부가 판단의 핵심 기준임을 강조한다. 아울러 이러한 논의가 CCW 체제 내 정부전문가그룹을 중심으로 합의에 기반하여 이루어져야 한다는 입장을 견지하고 있다. SG Report 2024, Annex I, pp.90-91.

71) 새로운 의정서에 관한 제안 Proposal: Roadmap Towards New Protocol on Autonomous Weapons Systems을 제출한 국가는 아르헨티나, 코스타리카, 과테말라, 카자흐스탄, 나이지리아, 파나마, 필리핀, 시에라리온, 팔레스타인, 우루과이. 법적 구속력 있는 문서 채택의 요청에 관한 Written Commentary Calling for a Legally-Binding Instrument on Autonomous Weapon Systems은 상기 국가에 더하여 에콰도르, 페루가 참여하였다. https://reachingcriticalwill.org/disarmament-fora/ccw/2022/laws/documents. (최종 접속일: 2026년 2월 5일).

72) 안도라, 아르헨티나, 볼리비아 다민족국, 브라질, 콜롬비아, 코스타리카, 쿠바, 칠레, 도미니카공화국, 에콰도르, 엘살바도르, 과테말라, 온두라스, 멕시코, 니카라과, 파나마, 파라과이, 페루, 포르투갈, 스페인, 우루과이 및 베네수엘라의 공동 답변서. SG Report 2024, Annex I, pp.20-21.

73) 칠레, 콜롬비아, 코스타리카, 도미니카공화국, 에콰도르, 엘살바도르, 과테말라, 카자흐스탄, 나이지리아, 파나마, 페루, 필리핀, 시에라리온 및 팔레스타인의 공동 답변서. SG Report 2024, pp.34-36.

74) SG Report 2024, Annex I, pp.24-28.

75) SG Report 2024, Annex I, pp.57-61.

76) SG Report 2024, Annex I, pp.68-70.

77) SG Report 2024, Annex I, pp.73-76.

78) SG Report 2024, Annex I, p.77.

79) SG Report 2024, paras 69-70.

은 관련 문서에서 이견 없이 확인되고 있다. 이처럼 기존 국제법은 자율무기체계의 개발과 사용을 규율하기 위한 기본적인 원칙과 규칙을 제공하지만, 자율무기체계가 지닌 기술적 복잡성과 예측불가능성을 충분히 규율하기에는 특정성(specificity)과 집행성(enforceability) 측면에서 한계를 지닌다.[80] 특정성의 결여는 자율성의 범위, 예측불가능성의 정도, 인간 개입 및 통제의 수준 등 자율무기체계의 핵심 쟁점에 관한 명확한 기준의 부재에서 비롯된다. 현행 국제인도법 규칙은 인간 행위자의 판단과 책임을 전제로 설계되어 있어, 기계학습 기반의 비결정적 시스템 작동을 충분히 포섭하지 못한다. 집행성의 한계는 제1추가의정서 제36조에 규정된 신무기 검토 절차가 각국의 재량에 전적으로 맡겨져 있다는 점에서 명확히 나타난다. 국가 간 공동 검증이나 정보 공유, 투명성 확보를 위한 제도적 메커니즘이 부재한 상황에서, 자율무기체계의 합법성 판단은 개별 국가의 정책과 군사적 필요성에 종속될 수밖에 없으며, 국제적 수준의 사전 통제나 책임성을 담보하기 어렵다.[81] 나아가 자율무기체계를 규율하는 구체적인 국제조약이 없는 경우, 국가들은 일반 규칙의 적용 방식을 상이하게 해석할 수 있다.[82]

이러한 배경에서, 자율무기체계에 특화된 '보완적이고 구속력 있는 새 조약'의 필요성이 꾸준히 제기되어 왔다. Human Rights Watch는 자율무기체계 논의 초기부터 인도주의 원칙과 공공양심의 명령에 근거하여 완전한 자율무기에 대한 선제적 금지조약의 필요성을 강조해 왔으며,[83] 2025년 보고서에서는 자율무기체계가 생명권과 인간 존엄성 등 기본적 인권에 미치는 위협을 분석하며 새로운 조약의 체결을 재차 촉구하였다.[84] ICRC 또한 2021년 5월 발표한 공식 입장에서 자율무기체계가 초래하는 인도주의적, 법적, 윤리적 위험을 지적하며, 새로운 법적 구속력이 있는 규칙의 채택을 권고하였다.[85] ICRC는 자율무기체계 기술 및 활용의 발전 속도를 고려할 때, 국제적으로 합의된 한계를 시의적절하게 설정하는 것이 매우 중요하다고 강조하면서, 예측 및 설명이 불가능한 자율무기체계와 인간을 표적으로 무력을 행사하는 자율무기체계의 금

80) Perrin, *supra* note 8.
81) Ibid.
82) Ibid.
83) Human Rights Watch, Heed the Call, A Moral and Legal Imperative to Ban Killer Robots, August 21, 2018.
84) Human Rights Watch, A Hazard to Human Rights, Autonomous Weapons Systems and Digital Decision-Making, April 28, 2025.
85) ICRC Position on Autonomous Weapon Systems, May 12, 2021.

지, 그리고 그 밖의 체계에 대한 규제 마련을 제언하였다. 이러한 요청은 서론에서 언급한 2023년 10월 유엔 사무총장과 ICRC 총재의 공동 성명으로 공식화되었다. 동 성명은 자율무기체계에 대한 새로운 국제조약의 조속한 협상 개시와 2026년 말까지의 채택을 촉구하면서, ① 효과를 예측할 수 없는 자율무기체계의 금지, ② 사용의 장소 · 시기 · 기간 · 규모 · 대상에 대한 제한, ③ 인간의 효과적 감독과 적시 개입 및 비활성화 능력의 보장을 포함해야 할 구체적 금지 및 제한 항목으로 제시하였다.[86] 전술한 바와 같이 유엔 사무총장은 2024년 보고서를 통해 법적 구속력 있는 문서 마련의 필요성을 재차 강조했으며[87], 유엔 총회는 결의를 통해 사무총장의 요청에 '주목'하였다.[88]

실제 새로운 조약의 체결보다는 기존 국제인도법의 적용을 강화하고, 원칙의 해석 및 적용 방식을 발전시키는 방식을 제안하는 견해도 있다.[89] 정부전문가그룹의 Rolling Text 작업 과정에서도 기존 국제인도법의 언어를 사용하는 것이 연속성 확보에 필요하며, 동시에 예측가능성, 신뢰성, 추적가능성, 설명가능성과 같이 자율무기체계에 특유한 요소들을 포함하여 이를 발전시켜야 한다는 주장이 제기된다.[90] 또한, 효과를 예측하고 통제할 수 없는 자율무기체계의 개발을 금지하는 것은 기술 발전 전반을 저해할 수 있고, 무기체계 생산 단계에서 지나치게 이른 시점에 과도한 제약을 초래할 수 있다는 반론도 적지 않다.[91] 그러나 인공지능을 통한 자율성이 주요 기능에 도입된 자율무기체계는 본질적으로 예측불가능하며, 인간을 행위자로 전제하는 기존 국제인도법만으로는 충분히 규율될 수 없다. 이는 자율무기체계를 규율하는 새로운 규칙과 기준 정립의 필요성을 뒷받침한다. 기존 국제인도법이 기본적으로 적용되더라도, 자율무기체계를 규율하는 특정 조약의 제정은 해석의 일원화, 명확한 금지 및 제한의 설정, 기술 사용에 대한 책임성과 윤리적 통제의 확보에 유리하다.[92] 이하에서는 새로운 규칙 마련이 필요하다는 관점에서, 자율무기체계 관

86) 앞의 주 1).
87) SG Report 2024, paras 89-90.
88) GA Res. 79/62, *supra* note 6.
89) Laurie R. Blank, "New Treaty Law on Autonomous Weapons? An Opportunity to Reframe the Discourse", *Case Western Reserve Journal of International Law,* Vol.57 (2025), p.248. Craig Martin, "Autonomous Weapons Systems and Proportionality: The Need for Regulation", *Case Western Reserve Journal of International Law,* Vol.57 (2025), pp.293-298.
90) Chair's summary – First 2025 session of the GGE on LAWS, UN Doc. CCW/GGE.1/2025/WP.1 (7 April 2025), p.6, para.22.
91) Ibid., p.6, para.24.
92) Perrin, *supra* note 8.

련 논의의 주요 쟁점을 검토함으로써 협약 체결 가능성과 협약 논의에 필요한 내용을 살펴보기로 한다.

Ⅲ. 자율무기체계 관련 논의의 주요 쟁점

1. 자율무기체계의 개념 및 범위

'자율무기체계'에 대해서는 여러 작업 정의(working definition)가 제시되어 있으나, 보편적으로 합의된 개념 정의는 아직 없다고 보아야 할 것이다.[93] 자율무기체계의 개념 정의는 규율 대상 및 범위의 설정과 직결되므로, 법적 구속력 있는 문서 마련 과정에서 반드시 논의되어야 할 핵심 쟁점이다.

1) '자율무기체계'

자율무기체계는 무인무기체계 발전의 연속선상에 있다.[94] 초기 자율무기체계 논의에서 '자율성'은 기계가 인간의 감독 없이 사전에 프로그래밍된 바에 따라 스스로 작동하는지, 그리고 인간이 이를 사후에 중지할 수 있는지의 여부를 중심으로 논의되었다. 이러한 관점에서 접근한 Human Rights Watch의 2012년 보고서는 인간의 개입 정도에 따라 ① 표적 선정과 무력행사가 오직 인간의 명령에 따라 이루어지는 인간-개입형(Human-*in*-the-Loop) 무기체계, ② 표적 선정과 무력행사가 로봇에 의해 수행되나 인간 운용자가 이를 감독하며 필요할 경우 개입하여 중지시킬 수 있는 인간-감독형(Human-*on*-the-Loop) 무기체계, ③ 표적 선정과 무력행사가 인간의 입력이나 상호작용 없이 이루어질 수 있는 인간-배제형(Human-*out-of*-the-Loop) 무기체계로 구분한다.[95] 이 중 Human Rights Watch는 인간의 개입이 배제된 무기체계와 감독이 허용되더라도 제한적이어서 실질적으로 배제에 준하는 무기체계를 '완전자율무기'로 간주하였다.[96]

---

93) Mariarosaria Taddeo & Alexander Blanchard, "A Comparative Analysis of the Definitions of Autonomous Weapons Systems". *Science & Engineering Ethics,* Vol.28 (2022), pp.1-22. SG Report 2024, para. 5.

94) Christof Heyns, "Human Rights and the Use of Autonomous Weapons Systems (AWS) During Domestic Law Enforcement", *Human Rights Quarterly,* Vol.38 (2016), p.354.

95) Human Rights Watch & IHRC, Losing Humanity, *supra* note 21, p.2.

2012년 처음 도입되어 2023년 개정된 미국 국방부의 지침(DoDD) 3000.09는 '자율 및 반자율(autonomous and semi-autonomous)' 무기체계를 규율 대상으로 한다. 동 지침은 '자율무기체계'를 "활성화된 이후에는 운용자의 추가 개입 없이 표적을 선정하고 공격에 관여할 수 있는 무기체계"로 정의하며, 운용자가 개입을 통해 작동을 중지(override)할 수 있도록 설계되었으나 일단 활성화되면 운용자의 추가 입력 없이도 표적을 선정하고 공격에 관여할 수 있는 운용자 감독형 자율무기체계(operator-supervised autonomous weapon systems)도 이에 포함된다.[97] 이는 동 지침의 규율 범위가 아직 실현되지 않은 것으로 평가되는 완전자율의 Human-*out-of*-the-Loop 무기체계에 한정되지 않고, Human-*on*-the-Loop 무기체계까지 포함함을 의미한다.

한편, '반자율무기체계'는 "활성화된 이후에는 운용자가 선정한 개별 표적 또는 특정 표적군만을 교전 대상으로 하도록 의도된 무기체계"로 정의되며, 교전 관련 기능(표적 획득, 추적, 식별 등)에 자율성이 활용되는 무기체계와 발사 후 망각(fire and forget) 또는 발사 후 표적 포착(lock-on-after-launch) 유도탄이 이에 포함된다.[98] 따라서 운용자가 교전 대상 선정에 대한 통제권을 유지하는 한, 표적 획득, 추적, 식별 등 교전 관련 기능에 자율성을 활용하는 무기체계라 하더라도 '자율무기체계'가 아닌 '반자율무기체계'로 분류된다. 즉, 동 지침은

---

96) Ibid., pp.2-3.

97) 미국 국방부는 2012년 11월 '무기체계의 자율성(Autonomy in Weapon Systems)'에 대한 국방부 지침(DoDD) 3000.09를 발표한 이후, 2023년 1월 업데이트한 지침을 발표하였다. United States, Department of Defence, DoD Directive 3000.09. "Autonomy in Weapon Systems", January 25, 2023, p.21. ("A weapon system that, once activated, can select and engage targets without further intervention by an operator. This includes, but is not limited to, operator-supervised autonomous weapon systems that are designed to allow operators to override operation of the weapon system, but can select and engage targets without further operator input after activation.") 2012년 지침에서는 주체를 'human' 또는 'human operator'로 명시하였으나, 2023년 개정 지침은 'operator'로 변경하였다. 이전 지침의 원문은, 임예준, 앞의 주 23), 266면, 각주 5 참고.

98) Ibid., p.23. ("A weapon system that, once activated, is intended to only engage individual targets or specific target groups that have been selected by an operator. This includes: Weapon systems that employ autonomy for engagement-related functions including, but not limited to, acquiring, tracking, and identifying potential targets; cuing potential targets to operators; prioritizing selected targets; timing of when to fire; or providing terminal guidance to home in on selected targets, provided that operator control is retained over the decision to select individual targets and specific target groups for engagement. "Fire and forget" or lock-on-after-launch homing munitions that rely on TTPs to maximize the probability that the only targets within the seeker's acquisition basket when the seeker activates are those individual targets or specific target groups that have been selected by an operator.")

무력의 행사 단계보다 표적 선정을 인간통제의 핵심으로 본 것이다.

이 지침에 따르면 자율성의 정도가 높은 순서대로 완전자율, 감독 기반 자율, 반자율 무기체계로 분류된다. 감독 기반 자율무기체계까지가 '자율무기'에 해당하며, 이미 상용화되었다고 볼 수 있는 반자율무기체계는 자율무기체계 범위에서 제외된다. 예컨대, 영국의 Taranis 드론, 미 해군의 자율함정 Sea Hunter, 보잉의 무인잠수정 Echo Voyager, 러시아의 무인전차 Uran-9 등은 모두 반자율무기체계로 분류된다.[99] 이스라엘 항공우주산업(IAI)이 제작한 레이더 탐지 선회체 공탄 Harpy 역시 'fire-and-forget' 방식을 취하므로 반자율무기체계로 분류된다.[100] 다만 IAI는 Harpy를 '완전자율'로 설명하고 있어, 적용하는 분류 기준에 따라 해석이 달라질 수 있음을 보여준다.[101]

반면, ICRC는 2014년 전문가 회의를 통해 자율무기체계를 "주요 기능에 자율성이 있는 모든 무기체계. 즉, 인간의 개입 없이 표적을 선정(탐색, 탐지, 식별, 추적, 선정)하고 공격(무력행사, 무력화, 손상, 파괴)할 수 있는 무기체계"라고 정의하였으며, 이 정의는 현재까지 관련 논의 전반에 걸쳐 사용되고 있다.[102] ICRC는 모든 단계에서 완전한 자율성을 갖춘 무기체계가 아니더라도, 표적 선정과 공격이라는 핵심 기능에 자율성이 도입된 경우에는 규범적 검토가 필요하다는 취지에서 광의의 개념 정의를 채택하고 있다.[103]

상기 논의를 종합하면, 자율무기체계 규율에 관한 논의는 아직 실현되지 않

---

99) 조현식, "인공지능, 자율무기체계와 미래 전쟁의 변환",「21세기정치학회보」제28집 제1호 (2018), pp.123-124.

100) Harpy는 발사 후 인간의 추가 개입 없이 표적을 탐색하고 추적하며 공격할 수 있으나, 운용자가 임무 영역과 기준 신호를 설정해야만 작동하므로 완전 자율무기체계로 분류되지는 않는다. 김민혁, 김재오, "자율살상무기체계에 대한 국제적 쟁점과 선제적 대응 방향".「국방연구」제63권 제1호 (2020), 174면,

101) Israel Aerospace Systems HARPY loitering munition, Automated Decision Research, https://automatedresearch.org/weapon/israel-aerospace-systems-harpy-loitering-munition/ (최종 접속일: 2025년 11월 12일). HARPY, Anti Radiation Loitering Munition, https://www.iai.co.il/p/harpy. Harpy의 주요 임무는 레이더 시스템이나 대공 방어망을 찾아 파괴하는 것이다. 발사 후 목표 지역 상공에서 선회하면서 레이더 신호를 탐지하고 사전에 저장된 레이더 신호 프로파일과 비교하여 일치하는 레이더를 식별하면 돌진하여 자폭하는 방식이다.

102) ICRC, Expert Meeting: Autonomous Weapon Systems, Technical, Military, Legal and Humanitarian Aspects, Geneva, Switzerland (March 26-28, 2014), p.7. ("Any weapon system with autonomy in its critical functions. That is, a weapon system that can select (i.e. search for or detect, identify, track, select) and attack (i.e. use force against, neutralize, damage or destroy) targets without human intervention"). 2021년 ICRC의 기본입장에 관한 문서도 동일하다.

103) 임예준, 앞의 주 23), 270면.

은 것으로 평가되는 완전자율무기체계, 즉, 인간-배제형에 한정되지 않고, 인간-감독형 또는 주요 기능에 일정 정도의 자율성이 도입된 무기체계 전반을 포함한다. 자율성의 정도는 인간의 개입 방식, 범위, 수준과 연관되나, 그 세부적인 기준은 '자율무기체계'에 대한 작업 정의에 따라 달라질 수 있다. 한편 자율무기체계의 정의는 규율 범위와 직결되므로, 국가들 간 견해차가 클 수밖에 없다. 관련하여 룩셈부르크는 공통된 정의의 확립이 법적 구속력 있는 문서에 관한 협상 개시의 전제조건이 아니라는 점을 명시적으로 강조하고 있다.[104] 2025년 12월 회의까지 마련된 정부전문가그룹의 Rolling Text에서는 9월 회의 이후 의장의 추가 제안을 반영하여[105], 자율무기체계의 특성을 다음과 같이 개방적으로 기술하고 있다.[106]

> 1. [CCW의 적용 범위 내에서, 그리고 문서의 요소라는 목적에서,] 치명적 자율무기체계는 하나 이상의 무기와 기술적 구성요소가 기능적으로 통합된 결합체로서, 일단 작동되면, 환경 속의 정보를 활용하여, 인간 운용자의 개입 없이 표적을 식별, 선정, 공격할 수 있는 무기체계로 특징지을 수 있다. [...]
> C. 상기 설명은 향후의 새로운 해석이나 이 특징의 잠재적 수정, 그리고 특정 유형의 체계를 제외할 가능성에 영향을 미치지 않는다.

2) '치명적' 자율무기체계

국제사회의 초기 논의는 인간 살상을 목적으로 하는 '킬러로봇'에 대한 우려에서 출발하였으나, 오늘날 시민사회와 관련 전문가들의 논의는 점차 '자율무기체계' 전반으로 확장되면서, '치명적(lethal)'이라는 수식어를 사용하지 않는 경향을 보인다. 반면, 국가 중심의 포럼, 특히 2014년 이후 CCW 체제 내 논의와 2023년 이후 유엔 총회는 '치명적' 자율무기체계를 일관되게 논의의 기본 단위로 삼고 있다.

그러나 '치명적'이라는 수식어가 개념 정의에 포함될 수 있는지, 그리고 치명성을 기준으로 규율 범위를 설정하는 것이 적절한지는 의문이다. 모든 무기는 사용방식과 표적에 따라 치명적일 수도, 비치명적일 수도 있으며, 치명성은 본질적으로 무기 설계의 속성이 아니라 사용 효과에 관한 것이다. Bruun은 치명

104) SG Report 2024, pp.70-71.
105) GGE on LAWS, Additional suggestions from the Chair for consideration on 5 September regarding Box I of the rolling text.
106) GGE on LAWS, Rolling Text, status date: 18 December 2025.

성 개념이 규율 범위를 본질적으로 협소화하며 자율무기체계의 위험을 충분히 포착하지 못한다고 지적한다.[107] 유엔 사무총장의 2024년 보고서에서도 일부 국가들은 '치명적'이라는 표현이 국제인도법상 근거가 없으며, 치명성은 무기의 사용방식에 따라 발생하는 효과에 불과하다고 지적하였다.[108]

Rolling Text 논의 과정에서도 '치명적'이라는 표현의 삭제를 주장하는 의견이 제시되었다. 그 근거로는 치명성이 무기체계의 본질적 특성이 아니며 국제인도법에 명시적 근거가 없다는 점, CCW와 국제인도법 모두 부상 및 민간 시설에 대한 파괴도 규율 대상으로 삼는다는 점, 치명성을 규율 기준으로 삼을 경우 논의 대상의 범위가 지나치게 좁아진다는 점이 제시되었다.[109] 반면, 이미 위임된 임무에서 명시된 사항이므로 그대로 유지해야 한다는 의견도 있었으며, 이 입장에서는 표현 삭제로 인한 논의 범위의 과도한 확대를 우려하였다.[110]

이러한 논의의 간극을 좁히기 위하여 정부전문가그룹 의장은 2025년 3월 '치명적'이라는 표현을 정의하는 방안을 제안하였다. 동 제안에 따르면, "치명적이란 한 명 또는 그 이상의 사람을 살상하거나 부상시킬 능력을 갖춘 무기를 의미하며, 물체의 손상 또는 파괴에 사용될 수 있는 무기도 포함한다."[111] 이는 '치명적'이라는 표현을 인명 살상에 국한하지 않고, 부상과 물적 파괴까지 포함하는 폭넓은 개념으로 재구성한 것으로, 자율무기체계의 규율 범위가 지나치게 협소해지는 것을 방지하려는 취지이다. 2025년 12월 회의까지 마련된 Rolling Text에 따르면, '치명성'은 그 범위가 인명 살상에 국한되지 않고 부상의 초래, 물적 손상 또는 파괴 등 다양한 유형의 유해한 결과를 포괄하는 개념으로 이해된다.[112]

---

107) Bruun, *Towards a Two-Tiered Approach*, *supra* note 65, p.2.

108) SG Report 2024, para 6.

109) Chair's summary - First 2025 session of the GGE on LAWS, CCW/GGE.1/2025/WP.1 (7 April 2025), para 12.

110) Ibid.

111) GGE on LAWS, Suggestions Box I and Box II, 05 March 2025. ("For purposes of these elements of an instrument, lethal means weapons with the capacity to kill or injure one or more persons, which includes weapons that can be used to damage or destroy objects.")

112) GGE on LAWS, Rolling Text, status date: 18 December 2025. ("The fact that a LAWS can be used in a way that does not result in loss of life, such as to damage or destroy objects or to cause injury, does not exclude it from this characterization.")

2. 이원적 접근방식: 금지와 규제

자율무기체계 규율 논의는 '금지(prohibition)'와 '제한/규제(restriction/regulation)'로 구분되는 '이원적 접근방식(two-tiered approach)'을 중심으로 전개되고 있다. 이는 모든 자율무기체계를 일률적으로 금지하기보다, 그 위험성과 자율성의 정도를 고려하여 차등적으로 규율하려는 시도이다.[113] 예컨대, 완전히 자율적이어서 인간통제가 결여된 치명적 자율무기체계는 전면적으로 금지하고, 일정 수준의 인간통제 또는 감독이 가능한 자율무기체계는 규제, 즉, 제한적으로 허용한다는 것이다. 이러한 접근방식은 기술 발전의 속도와 군사적 필요성, 그리고 현실적 규제 가능성을 종합적으로 고려한 절충안이라고 볼 수 있다.

이원적 접근방식은 군비통제에서 흔히 활용되는 구조로, 모든 상황에서 불법으로 간주되는 무기체계의 유형과 사용을 규정하고, 그 밖의 무기체계는 개발 및 사용에 있어 제한과 요건을 설정하는 방식이다.[114] 예를 들어, 실명레이저무기에 관한 CCW 제4의정서는 영구적 실명을 야기하도록 특별히 설계된 레이저무기의 개발, 사용, 양도를 전면 금지하는 한편(제1조), 다른 합법적인 군사 목적을 위해 설계된 레이저 시스템은 "영구적 실명을 전투 기능 또는 전투 방식의 하나로 포함하도록 설계되지 아니한 경우" 금지되지 아니한다고 규정한다(제3조). 현재 논의와 같이 인간의 통제 또는 개입을 자율무기체계가 유지해야 할 핵심 요소로 본다면, 인간의 통제 또는 개입 없이 자동으로 표적을 탐지, 선정, 공격하는 완전자율의 인간 배제형 무기체계는 개발, 사용, 양도 모두 전면 금지의 대상이 되고, 표적의 식별, 선정, 공격의 최종 단계에서 인간이 통제하거나 개입하는 반자율 또는 인간 개입형 무기체계는 일반적 규제 범주에 속하게 된다.[115]

ICRC는 2021년 5월 성명서에서 자율무기체계에 관한 '새로운 법적 구속력 있는 규칙'의 필요성을 강조하며 명시적으로 이원적 접근방식을 제시하였다.[116] 이에 따르면, 인간을 대상으로 무력을 행사하도록 설계되거나 사용되는 자율무기체계, 예측불가능하거나 의미 있는 인간통제가 불가능한 자율무기체계는 금지되어야 한다. 그 밖의 자율무기체계는 엄격한 규제하에 허용될 수 있

113) Perrin, *supra* note 8.
114) Blanchard & Goussac, *supra* note 40, p.5. Bruun, *Towards a Two-Tiered Approach, supra* note 65, p.3.
115) Ibid.
116) ICRC Position on Autonomous Weapon Systems, 12 May 2021.

다. 금지 대상이 아닌 자율무기체계의 설계 및 운용은 ① 민간인 및 민간물자의 보호, 국제인도법 규칙의 준수, 인류 수호(safeguard humanity)를 위하여, ② 표적 유형, 사용의 지속기간 및 지리적 범위, 규모, 사용 상황에 대한 제한과 ③ 인간-기계 상호작용에 대한 요건의 결합을 통해 규제되어야 한다. 구체적인 예를 들어보면, 군사적 목표에 해당하는 물체로 표적을 한정하고, 특정 공격에 대한 인간의 통제가 가능한 방식으로 운용하며, 인간의 효과적인 감독과 적시 개입 및 비활성화를 보장하기 위한 요건을 포함하도록 하는 것이다.[117]

앞서 언급한 2023년 유엔 사무총장과 ICRC 총재의 성명서 및 유엔 사무총장의 2024년 보고서도 이원적 접근방식을 취하고 있다. 2024년 채택된 유엔 총회 결의 제79/62호[118]는 유엔 사무총장의 "이원적 접근방식에 따른 자율무기체계의 금지 및 규제에 관한 법적 구속력 있는 문서의 협상 촉구"에 '주목'하였다.[119] 2024년 유엔 사무총장 보고서에 나타난 국가들의 의견도 금지와 규제로 구분될 수 있다. 다수의 국가는 인간의 통제 없이 완전히 자율적인 무기체계와 국제인도법에 따라 운용될 수 없는 무기체계의 금지를 요구하였다.[120] 국제인도법에 따라 운용될 수 없는 자율무기체계의 특성으로는, 본질적으로 무차별적인 무기체계, 전투원과 민간인을 구별할 수 없는 무기체계, 민간인 또는 민간물자에 대해 무력을 가하도록 설계된 무기체계, 공격이 예상되는 군사적 이익보다 과도한 부수적 피해를 초래할 수 있는지 판단할 수 없는 무기체계, 불필요한 고통 또는 과도한 상해를 유발하는 성격을 가진 무기체계, 그 효과를 신뢰할 수 있을 정도로 예측, 예상, 이해 또는 설명할 수 없는 체계, 그 효과를 제한할 수 없는 무기체계 등이 제시되었다.[121]

국가들은 상기 금지에 해당하지 않는 자율무기체계는 규제 대상이라고 보았다. 다만, 자율성은 그 스펙트럼이 넓고 자율무기체계의 특성과 작전 환경, 사용자에 따라 달라질 수 있으므로, 규제 조치가 이뤄진다면 그 내용의 구체화가 필요하다.[122] 규제의 목적은 핵심 기능에 대한 인간통제의 보장, 인간의 지휘 및 책임 체계의 확보, 전 수명주기에 걸친 국제법(특히, 국제인도법) 준수의 보장이다.[123] 국제인도법 준수를 위해 국가들이 제안한 구체적 조치는 ICRC의

117) Ibid.
118) GA Res. 79/62, *supra* note 6.
119) Perrin, *supra* note 8.
120) SG Report 2024, para 75.
121) SG Report 2024, para 76.
122) SG Report 2024, para 79.

제안과 유사하다.[124] 아울러 무기체계에 대한 엄격한 작동시험, 자율무기체계의 등록·추적 및 분석 보장, 위험평가 수행, 결정자와 운용자에 대한 교육·훈련 보장, 위험 완화 및 안전장치 촉진, 환경적 영향평가 등도 추가적으로 논의되었다.[125]

정부전문가그룹의 Rolling Text도 금지-규제의 이원적 구조를 취하고 있으며, 그 내용 또한 상기 서술한 바와 유사하다.[126] 2024년 11월의 초기 Rolling

---

123) SG Report 2024, para 80.

124) SG Report 2024, para 81. 교전할 수 있는 표적의 유형을 통제하거나 제한: 본질적으로 군사적 목표에 한정할 것; 사용의 지속기간, 지리적 범위, 사용 규모를 제한할 것; 무력 사용 결정에 대해 인간의 승인을 보장할 것; 작전 매개변수(지속시간, 범위, 규모 등)의 변경 시 인간의 승인을 보장할 것. 여기에 자폭, 자동 비활성화, 자동 무력화 메커니즘을 통한 조치를 포함할 것. 교전 횟수를 제한할 것; 인간 운용자의 국제인도법 준수에 대한 적절한 주의의무를 보장할 것; 예측가능성 및 신뢰성을 충분히 확보할 것; 새로운 무기, 전쟁 수단 및 방법이 국제인도법을 준수하는지에 대한 법적 검토를 할 것.

125) SG Report 2024, para 82.

126) GGE on LAWS, Rolling Text, status date: 18 December 2025, Box Ⅲ. 해당 내용은 Rolling text, status date: 26 November 2024, Revised on 6 March 2025를 바탕으로 수정된 버전이다. 2025년 12월 회의까지 합의된 Rolling Text의 내용은 다음과 같다.

1. 과도한 상해 또는 불필요한 고통을 초래하는 성질을 가지거나, 본질적으로 무차별적이거나, 또는 그 밖에 국제인도법을 준수하여 사용할 수 없는 자율무기체계의 사용은 모든 상황에서 금지된다.
2. 자율무기체계의 공격 효과가 그 사용 상황에서 국제인도법이 요구하는 바에 따라 예측 및 제한될 수 없는 경우, 그러한 자율무기체계의 사용은 금지된다.
3. 민간 주민 자체를 공격 대상으로 삼는 경우 뿐만이 아니라 개별 민간인 또는 민간 대상물을 자율무기체계의 공격 대상으로 삼는 것은 금지된다.
4. 국가는 국제인도법에 부합하여 사용할 수 없는 자율무기체계를 제조, 취득, 비축 또는 이전하여서는 아니 된다. 다만, 그러한 체계의 보유, 취득 또는 이전은 훈련 및 대응수단 개발을 목적으로 하는 경우에 한하여 최소한의 필요 수량 범위 내에서 허용되며, 해당 목적이 달성되거나 더 이상 필요하지 않게 되는 경우 즉시 폐기되어야 한다.
5. 자율무기체계는 구별, 비례성 및 공격 시 예방조치를 포함한 국제인도법의 원칙과 요건에 따라 사용되어야 한다.
6. 자율무기체계의 사용 및 그 효과가 구별, 비례성 및 공격 시 예방조치의 원칙과 요건을 포함한 국제인도법에 합치되도록 보장하기 위해서는 상황에 적절한(context-appropriate) 인간의 판단과 통제가 필요하다.
7. 국가는 특히 다음과 같은 조치를 통하여 상황에 적절한 인간의 판단과 통제를 보장하여야 한다.
   A. 자율무기체계가 책임 있는 지휘 및 통제 체계 하에서 운용되도록 보장할 것. 이는 특히 표적의 식별, 선정 및 교전 기능의 효과와 관련하여 인간에 의한 법적 의무 평가가 이루어지도록 하고, 윤리적 고려사항이 충분히 반영되도록 하는 것을 포함한다.
   B. 자율무기체계의 효과가 그 사용에 책임 있는 자에 의해 충분히 예측되고 통제될 수 있도록 보장할 것. 이를 위하여 표적 유형, 작전 지속시간, 지리적 범위, 자율적으로 수행할 수 있는 교전 횟수 등 작전 규모를 제한하거나 기타 방식으로 통제할 수 있다.

Text의 첫 번째 금지 항목인 “국제인도법, 특히 구별, 비례 및 공격 시 사전주의의 원칙과 요건을 준수하여 사용될 수 없는 자율무기체계의 사용은 모든 상황에서 금지된다”[127]는 다섯 번째 항목과 같이 국제인도법에 따라서만 사용될 수 있다는 제한의 영역으로 수정되었다. 아울러 2025년 3월 Rolling Text까지 유지되던 ‘모든 상황에서’ 금지된다는 표현은, 첫 번째 항목의 ‘과도한 상해나 불필요한 고통’을 초래하는 성격의 자율무기체계 사용을 제외하고는 삭제되었다. ‘배치’의 경우도 삭제되고, ‘사용’의 경우만 합의가 이루어졌다는 점도 확인된다. 전반적으로 자율무기체계 금지에 관한 체약국 간 합의의 범위가 축소된 것이다.

한편, 인간의 판단과 통제의 상황 적절성에 대한 강조는 유지되고 있으나, 이전의 경우와 같이 “상황에 적절한 인간의 통제와 판단 없이 작동하는 자율무기체계의 사용은 모든 상황에서 금지된다”[128]가 아닌 “상황에 적절한 인간의 판단과 통제를 보장”할 국가의 조치를 강조하는 방향으로 전환된 것을 확인할 수 있다. 다만, 상황에 적절한 인간의 판단과 통제’를 ‘국가’의 보장으로 명시하고 있다는 점은, 2024년 11월 Rolling Text의 ‘교전 당사자’의 보장으로 명시한 항목보다는 진일보했다고 평가할 수 있으며, 특히 마지막 항의 “금지, 요건 및 조치는 자율무기체계의 설계, 개발 및 사용 단계 전반에서 고려”되어야 한다는 점은 유의미한 수정 사항이다.

모든 경우와 상황에 있어 국제인도법의 기본원칙과 양립할 수 없는 무기체계는 금지되어야 한다. 국제인도법과 양립할 수 없는 자율무기체계를 명시적으로 금지할 경우, 피해 원인의 추적을 불가능하게 하는 기술적 특성과 인간과 기계의 상호작용이 포함되어야 할 것이다.[129] 또한, 본질적으로 예측불가능한

C. 실시간 기계학습을 포함하여 인간의 개입 없이 체계에 의해 임무 매개변수가 중대하게 변경되지 않도록 보장할 것.

D. 자율무기체계를 적시에 비활성화할 수 있도록 하거나, 자폭, 자기비활성화 또는 자기중립화 메커니즘을 통합할 것.

E. 민간인에 대한 피해 또는 민간물자에 대한 손상을 최소화하기 위하여 자율무기체계의 운용을 명확히 정의된 범위(perimeter)로 제한하고, 그 사용을 군사목표에 한정할 것.

8. 위와 같이 정립된 금지, 요건 및 조치는 자율무기체계의 설계, 개발 및 사용 단계 전반에서 고려되어야 한다.

127) GGE on LAWS, Rolling Text, status date: 26 November 2024, Box Ⅲ, 1.

128) Ibid. “5. It is prohibited in all circumstances to employ LAWS that operate without context-appropriate human control and judgment.”

129) Milena Sterio, “Autonomous Weapons Systems and the Need to Update International Humanitarian Law?”, *Case Western Reserve Journal of International Law,* Vol.57 (2025), pp.312-313.

자율무기체계도 금지 대상에 포함해야 할 것이다. 무기체계의 작동 및 기능에 대한 예측가능성은 국제인도법 준수에 있어 핵심적 요소이다.[130] 그런데 실제 모든 자율무기체계에는 일정 수준의 예측불가능성이 내재하는데, 이는 사용자가 구체적인 표적이나 정확한 공격 시점 및 위치를 직접 선택하거나 알 수 없기 때문이다.[131] 이러한 예측불가능성은 자율무기체계의 표적 범위가 확대되고, 작전 시간 및 공간이 확장되어 복잡한 환경에서 사용됨에 따라 더욱 증대된다.[132] 나아가 기계학습 소프트웨어에 의해 통제되는 자율무기체계는 인간이 작동 과정 전반을 이해할 수 없게 만들어, 그 과정을 예측하고 설명하는 것 역시 어렵게 만든다.[133] ICRC는 이러한 '블랙박스(black-box)' 문제는 사용 환경과 무관하게 발생한다고 지적한다.[134] 2025년 12월 Rolling Text에서 추가된 인간의 판단과 통제를 보장하기 위하여 "임무 매개변수가 중대하게 변경되지 않도록 보장할 것"이라는 C항목은 기계학습으로 인한 블랙박스 문제를 고려한 것으로 보인다.

향후 논의에서 고려할 사항은 대인 표적 자율무기체계를 명시적으로 금지할지 여부이다. ICRC를 포함한 관련 시민단체들은 인간을 대상으로 무력을 행사하도록 설계되거나 사용되는 자율무기체계를 전면적으로 금지해야 한다고 주장해 왔다.[135] 유엔 사무총장 보고서에 따르면, 다수의 국가도 인간을 직접 표적으로 삼도록 설계된 자율무기체계의 금지를 언급하였다.[136] 이른바 살상로봇은 인간성과 인간 존엄의 원칙에 배치되므로 그 본질적 특성만으로도 금지될 필요가 있다. 이에 따르면 사람을 표적으로 삼고, 의미 있는 인간의 통제가 불가능한 영화 속 '터미네이터' 뿐만 아니라, 완전한 자율무기체계는 아니지만, 자동 탐지 및 추적 기능을 갖추어 인간을 감지하는 즉시 물리력을 행사하도록 설계되거나 운용되는 대인 경계 로봇(Anti-Personnel Sentry Robots)도 금지 대상에 포함된다.[137] 그러나 현재까지 CCW 정부전문가그룹의 논의는 제3항과 같이 국제인도법에 따라 '민간인'을 표적으로 한 자율무기체계 사용을 금지하

130) Davison, *supra* note 62, p.15.
131) ICRC Position on Autonomous Weapon Systems, Background Paper, 2021, p.7.
132) Ibid.
133) Ibid.
134) Ibid.
135) Ibid., pp.8-9.
136) SG Report 2024, para 77.
137) Elizabeth Minor & Richard Moyes, Regulating Autonomy in Weapons Systems, Article 36 (21 October 2020), https://article36.org/updates/treaty-structure-leaflet/ (최종 접속일: 2026년 2월 11일)

는 수준에 머물러 있으며, 이마저도 '모든 상황에서'라는 부분이 삭제됨으로써 대인 표적 자율무기체계의 전면적 금지로는 논의가 진전되지 않고 있음을 확인할 수 있다.

3. '인간통제'의 의미

자율무기체계에 관한 핵심 쟁점 중 하나는 '인간통제'를 어떻게 규정할 것인가이다. '인간통제' 여부는 앞서 살펴본 금지와 규제 구분의 일차적 기준이기도 하다. ICRC는 2014년 전문가회의와 공식 성명을 통해 무기체계의 핵심적 기능에 대한 '인간통제'가 국제인도법 적용의 필수적 전제임을 강조하며, 무기체계의 설계 및 운용 단계에서 인간의 판단과 통제를 유지해야 한다는 원칙을 제시하였다.[138] 그러나 '인간통제'가 어떠한 기능에 대한, 어떠한 방식의 통제를 의미하는지는 여전히 논쟁의 대상이다.[139] 이와 관련하여 2019년 정부전문가그룹은 인간통제와 판단을 구성하는 요소 및 요구되는 인간-기계 상호작용의 유형과 정도에 관한 보다 명확한 기준의 마련이 필요하다고 하였다.[140]

'인간통제'의 정도 또는 양상을 수식하는 대표적인 표현으로는 '의미 있는 인간통제(meaningful human control)'가 있다. '의미 있는 인간통제'는 단순한 기술적 감독에 그치는 것이 아니라, 인간의 판단이 무기체계의 결정 과정에 실질적으로 개입하고 그 결과에 대한 책임을 보장하는 통합적 개념으로 이해된다. 이 표현은 2021년부터 진행된 CCW 체제 내 정부전문가그룹 논의에서 반복적으로 사용되었으며, 2023년 유엔 총회 결의 제78/241호도 "의미 있는 인간통제 또는 감독(meaningful human control or human oversight)"을 명시하였다. 이

---

138) ICRC는 자율무기체계의 설계와 사용은 언제나 인간의 통제와 판단하에 이루어져야 하며, 인간의 개입 없이 무기체계가 표적을 선택하고 공격을 수행하는 것은 무력사용 결정이 인간의 법적 및 도덕적 판단으로부터 이탈하는 결과를 초래한다고 지적하였다. ICRC, *Autonomous Weapon Systems: Technical, Military, Legal and Humanitarian Aspects* (Expert Meeting Report, Geneva, 26 - 28 March 2014), pp.1-3. ICRC, Statement on Autonomous Weapon Systems (13 May 2014), www.icrc.org/eng/resources/documents/statement/2014/05-13-autonomous-weapons-statement.htm. (최종 접속일: 2026년 2월 6일)

139) Vincent Boulanin et al., *Limits on Autonomy in Weapon Systems: Identifying Practical Elements of Human Control*, SIPRI & ICRC (June 2020), pp.1-2. 실제로 '인간통제'는 CCW의 자율무기체계 논의 초기부터 주요 쟁점이었다. 2015년부터 2018년까지 CCW 회의에서 '인간통제'에 관한 의견 대립은, 박문언, 「자율무기체계의 국제법적 허용성과 규제방안」(박사학위논문, 서울대학교, 2019), 105-107면 참조.

140) GGE Report 2019, para. 22(a) - (b).

표현은 영국 군축 관련 시민사회단체인 Article 36이 2013년 영국 국방부가 발간한 「The UK Approach to Unmanned Systems」에 대응하여, 무기체계의 운용이 항상 인간의 통제하에 남아 있을 것이라는 영국 정부 원칙의 구체화를 요청하는 맥락에서 처음 사용되었다.[141] Article 36이 제시한 '의미 있는 인간통제'는 개별 공격과 직접 결부된 정보, 행위, 책임성을 요소로 한다.[142] 먼저, 정보 측면에서 공격 결정자는 표적 지역의 충분한 맥락 정보, 표적 선정의 근거, 공격 목표 및 그로 인한 단·장기적 효과에 관한 정보를 보유해야 한다.[143] 행위 측면에서 공격 개시는 반드시 인간의 명시적 판단과 행동을 통해 이루어져야 하며, 단순히 기계적으로 발사 버튼을 누르는 행위만으로는 통제가 확보되었다고 볼 수 없다.[144] 책임성 측면에서 공격 수행에 관여한 자는 그 결과에 대해 법적, 도덕적 책임을 져야 하므로, 결정 과정의 기록과 추적 및 검증 절차가 전제되어야 한다.[145]

한편, 정부전문가그룹은 2024년 11월 마련한 Rolling Text에서 국제인도법의 기술 중립성과 전면적 적용을 재확인하고, 무기 사용과 그 효과에 대한 인간통제 개념을 '상황에 적절한 인간의 판단과 통제(context-appropriate human control and judgment)'로 공식화하였다.[146] Rolling Text의 '상황에 적절한 인간의 판단과 통제'라는 표현은 미국 국방부 지침에서 사용된 용어인 '적절한 인간 판단(appropriate human judgment)'과 유사하다. '적절한'이란 용어는 모든 상황에 일률적으로 적용되는 고정된 기준이 아니라, 무기체계의 특성, 전쟁의 양상, 작전 상황, 교전 규칙, 기술 수준 등 다양한 요소에 따라 전쟁법 준수를 위해 유연하게 적용되는 개념이다.[147] 이는 자율무기체계에 대한 인간의 통

141) Article 36은 킬러 로봇 금지 캠페인(Campaign to Stop Killer Robots) 창립단체이자 2012년부터 완전자율살상로봇금지를 촉구한 영국 시민단체이다. Article 36은 '의미 있는 인간통제' 없이 공격을 수행할 수 있는 완전자율무기체계를 개발하지 않겠다는 영국 정부의 약속을 강화하고, 완전자율무기의 법적 규제를 명확하게 하기 위한 국제조약의 필요성을 인정할 것을 요청하였다. Article 36, Killer Robots: UK Government Policy on Fully Autonomous Weapons, April 2013, p.1, pp.3-5, www.article36.org/wp-content/uploads/2013/04/Policy_Paper1.pdf. (최종 접속일: 2026년 2월 6일)

142) Ibid., p.3.

143) Ibid., p.4.

144) Ibid.

145) Ibid.

146) 원문: "Context-appropriate human judgement and control is needed to ensure the use and effects of LAWS are in compliance with international law, in particular IHL, including the principles and requirements of distinction, proportionality and precautions in attack." 이후 2025년 3월 및 9월 회의까지는 수정되지 않다가, 12월 회의를 통해 '판단 및 통제(judgment and control)'로 병기 순서가 변경되었다.

제는 단일한 고정 기준에 의해 결정되는 것이 아니라, 무기체계의 특성이나 작전 환경, 기술 수준 등 다양한 요소에 따라 맥락적으로 구성되는 적정 수준의 인간 개입을 의미하는 것으로 이해할 수 있다.

'인간통제'는 자율무기체계에 어떠한 '한계'를 설정할 것인가의 문제와도 직결된다. ICRC와 스톡홀름국제평화연구소(SIPRI)가 2020년 공동으로 발간한 보고서는 '인간통제 및 인간판단(human control and human judgment)'을 구성하는 실천적 요소로서, 세 가지 통제의 축, 즉, ① 무기체계 매개변수(parameter) 통제, ② 운용 환경 통제, ③ 인간-기계 상호작용 통제를 제시하였다.[148]

첫째, 무기체계 운용 매개변수에 대한 통제는 자율무기체계가 수행할 수 있는 임무의 성격과 범위를 명확히 규정하고, 표적의 유형, 사용 시기, 지리적 범위 및 작전 규모를 제한하며, 비활성화, 자폭, 무력화 기능을 내장하도록 설계하는 조치를 포함한다.[149] 이러한 매개변수 통제는 자율무기체계의 작동 범위를 기술적으로 한정함으로써 그 효과가 인간의 예측과 판단의 범위 내에서 이루어지도록 하는 장치로 기능한다.

둘째, 운용 환경에 대한 통제는 자율무기체계의 사용이 허용될 수 있는 작전 환경을 사전에 설정하거나, 민간인 또는 민간대상이 존재하지 않는 공간으로 운용을 제한하는 조치를 의미한다.[150] 예컨대 전투원이 명확히 구분되는 개활지에서의 사용으로 한정하거나, 민간 시설이 있는 지역을 배제하는 작전 계획을 수립하는 방식이다. 이러한 환경 통제는 자율무기체계 운용 단계에서 발생할 수 있는 구별의 원칙 위반 위험을 최소화하기 위한 조치이다.

셋째, 인간-기계 상호작용을 통한 통제는 인간 운용자가 자율무기체계의 작동을 지속적으로 감독하고, 필요한 경우 즉각적으로 개입하거나 작동을 중단할 수 있는 기술적·절차적 메커니즘을 의미한다.[151] 즉, 인간이 단순히 시스템의 초기 설정을 담당하는 데 그치지 않고, 실시간으로 무기체계의 판단 과정과 작동 결과를 검토 및 수정할 수 있도록 설계되어야 한다는 것이다. 보고서는 이러한 인간-기계 상호작용을 "의미 있는 인간통제를 실질적으로 구현하기 위한 실천적 전제"로 규정하면서, 세 가지 통제 축의 결합을 통해 예측불가능성과 책임성 결여의 위험을 완화할 수 있다고 보았다.[152]

---

147) 김현중, "넥스트 오펜하이머 시대: 자율살상무기 발전에 따른 예상쟁점 및 대응방안", 「INSS 전략보고」 No.284 (September 2024).
148) Boulanin et al,, *supra* note 139, p.8.
149) Ibid.
150) Ibid., p.9.
151) Ibid.

자율무기에 관한 국제사회의 논의는 인공지능 기술의 개발을 저해하지는 않되, 무력사용에 관해서는 최소한 일정 정도의 '인간통제'가 필요하다는 방향으로 수렴되고 있다. 모든 무기 개발에 있어 인간의 통제는 마지막까지 고수되어야 할 요소이다.[153] 인간통제는 무력행사의 인간적 성격을 유지하기 위한 최소한의 조건이기 때문이다. 국제인도법은 구별의 원칙, 비례의 원칙, 공격에 있어 예방의 원칙을 적용하는 데에 있어 인간의 판단 행위를 전제로 한다. 따라서 인간이 공격 효과를 예견할 수 있을 정도의 '의미 있는' 통제를 행사하지 못한다면, 법적 책임 추궁이 어려울 수 있다. 이러한 맥락에서 인간통제는 국제인도법 준수의 핵심 요소인 예측가능성을 제도적으로 확보하기 위한 장치이기도 하다. 의미 있는 인간통제는 단순히 '인간이 개입한다' 또는 '안한다'의 이분법이 아니라, 책임성과 예측가능성을 확보하기 위한 연속선에서 이해되어야 한다.[154] 인간통제는 단순히 인간이 어느 지점에서 개입하는 여부의 문제가 아니라, 누가, 무엇을, 어떤 단계에서, 어떠한 방식으로 통제해야 하는지에 관한 문제를 포괄한다. 이는 '인간통제'가 단순한 기술적 관리가 아니라, 적법성과 윤리적 정당성을 확보하기 위한 통합적 기준임을 의미한다. 그러나 자율무기의 운용은 본질적으로 복잡한 기술적 과정을 수반하므로 실제 전장에서 인간통제를 구현하기는 점점 더 어려워지고 있다.[155] 특히 스스로 학습하는 인공지능 자율무기체계는 인간통제의 가능성과 범위에 대한 근본적 의문을 제기한다. 가령 표적 프로파일이 기계학습을 통해 생성되거나, 인간의 승인 없이 운용 과정 중에 표적 프로파일이 변경될 수 있는 체계는 특정 사용 맥락에서 발생할 수 있는 결과를 충분히 통제할 수 없으므로 '의미 있는 인간통제'를 충족한다고 보기 어렵다. 이에 관해서는 이하에서 보다 자세히 살펴본다.

4. 기계학습과 '블랙박스'

2025년 9월 24일 유엔 안전보장이사회의 인공지능과 국제 평화 및 안보 의제에 관한 공개 토론에서 유엔 사무총장은 "인류의 운명을 알고리즘에 맡겨서는 안 된다"고 강조하였다.[156] 이 발언은 인공지능이 인간의 통제와 판단을 넘

152) Ibid., pp.7-9.
153) 임예준, 앞의 주 23), 292-293면.
154) Boulanin et al., *supra* note 139, p.8.
155) Sterio, *supra* note 129, p.304.

어선 결정을 내릴 가능성에 대한 국제사회의 근본적 우려를 반영한다. 이러한 맥락에서 인공지능 기반 자율무기체계의 의사결정 과정을 인간이 이해하거나 설명하기 어렵다는 점은 새로운 형태의 도전이 될 수 있다. 기계학습 과정을 완전히 설명할 수 없다는 사실은, 실제 전장에서 운용되는 자율무기체계의 작동 결과를 인간이 충분히 예측하거나 통제하기 어렵게 만드는 근본적 요인이다.[157]

자율무기체계의 '자율성'이 기계학습을 기반으로 한다는 것은, 앞서 살펴본 '인간통제'가 실질적으로 가능하지 않을 수 있음을 의미한다. 전통적 프로그램이 인간이 설계한 알고리즘을 통해 입력값과 출력값의 관계를 사전에 정의하는 것과 달리, 기계학습은 시스템이 스스로 데이터를 분석하여 규칙을 발견하고 모델을 개선하는 구조이다.[158] 즉, 기계는 수많은 실험과 오류를 통해 데이터 속에서 스스로 규칙과 패턴을 발견하고 개선해 나가는데, 이러한 학습 과정은 인간이 설계한 범위를 넘어 독자적으로 진화할 수 있어 그 판단의 경로를 인간이 완전히 추적하거나 예측하기 어렵게 만든다.

자율무기체계에 관한 초기 논의에서 제기된 우려는 주로 인간에 필적하는 판단 능력을 갖추지 못한 기계가 국제인도법상 요구되는 정교한 판단을 수행할 수 있는가에 관한 것이었다.[159] 예측이 어려운 전쟁 상황에서 인간 지휘관이 경험과 직관을 통해 수행해 온 복합적 판단을, 데이터와 알고리즘에 의존하는 자율무기체계가 동일한 수준으로 대체할 수 없다는 점이 핵심 논거였다. 그러나 2020년대 중반 이후의 우려는 기계의 판단 능력 자체보다는, 오히려 그 판단 과정을 인간이 파악하고 설명할 수 없는 '비가시성(opacity)'에 있다. 인공지능의 심층학습 과정이 심층신경망을 기반으로 점점 더 복잡해짐에 따라, 데이터가 각 계층을 거치면서 어떻게 변환되어 최종 결과에 이르는지 추적하기가 어려워졌기 때문이다. 자율무기의 판단 결과를 검증하거나 법적으로 평가할 수 있는 설명가능성의 결여는 문제 발생 시 원인 파악을 어렵게 하여 책임규명의 난제가 될 수 있으며, 나아가 해당 체계에 대한 신뢰성을 저하시킨다. 이

156) 'Humanity's Fate Cannot Be Left to Algorithm,' Warns Secretary-General in Security Council Debate on Artificial Intelligence, UN Press Release, SG/SM/22830, 24 September 2025.

157) Martin, *supra* note 89, pp.291-293.

158) 헨리 키신저, 에릭 슈밋, 크레이그 먼디(이현 옮김), 「새로운 질서: AI 이후의 생존 전략」 (서울: 월북, 2025) 68면.

159) 유준구, "자율살상무기체계의 논의 동향과 쟁점", 정책연구시리즈 2019-18 (국립외교원 외교안보연구소, 2019), 4면.

러한 비가시성은 군사 영역에서의 인공지능 도입이 기존 국가안보 의사결정 구조의 '블랙박스'에 기술적 '블랙박스'가 더해져 이른바 '이중 블랙박스(double black box)'를 형성할 수 있다는 우려를 낳고 있다.[160]

인공지능의 블랙박스 문제는 기계학습 과정에서 편향이 결합될 때 더욱 심각해진다. 학습 데이터에 내재한 편향은 알고리즘을 통해 증폭되며, 비가시성으로 인해 이러한 편향을 사전에 인지하거나 사후에 수정하기가 어렵다. 무기체계의 인식 기능에 사용되는 인공지능은 통상 통계적 알고리즘에 기반하며, 딥러닝의 정도가 높아질수록 데이터의 품질이 결과에 미치는 영향은 커진다.[161] 따라서 기계학습에서 발생하는 편향의 결과는 군사적 맥락에서 더욱 증폭될 수 있다.[162] 유엔 사무총장은 2025년 군사 영역에서의 인공지능이 국제 평화와 안보에 미치는 영향에 관한 보고서에서, 편향된 학습 데이터가 인공지능으로 하여금 국제인도법의 핵심 원칙인 구별의 원칙 준수를 어렵게 만들 수 있다고 경고하였다.[163] CCW 정부전문가그룹에 제출된 문서 역시, 자율무기체계의 맥락에서 성별 · 연령 · 인종 · 신체적 능력 등 다양한 요인이 표적 선정과 공격 대상 식별의 기준으로 사용될 수 있으며, 편향된 데이터 세트나 부정확한 알고리즘이 특정 집단을 더 높은 비율로 오인식하거나 전통적 전쟁 역할을 이유로 민간인 남성을 전투원으로 잘못 분류할 위험이 있음을 지적하였다.[164]

한편, SIPRI 보고서는 군사 영역에서 인공지능의 편향이 사회적 편향, 개발 편향, 사용 편향이라는 세 단계에서 발생하며, 이는 표적 오인식, 민간인 피해 예측의 왜곡, 차별적 감시 등으로 이어질 수 있음을 지적한다.[165] 이 세 단계의 편향은 서로 연결되어 순환적 영향을 미치며, 데이터 생애주기 전반에 걸쳐

160) Ashley S. Deeks, *The Double Black Box* (Oxford University Press, 2025). 저자는 인공지능 기술을 사용한 국가안보의 의사결정은 '이중의 블랙박스'로 결정 과정의 민주성을 확보하지 못한다고 지적한다.

161) 인공지능 시스템이 채택하는 모델의 알고리즘은 통계적 알고리즘과 규칙기반 알고리즘으로 분리된다. 관련 내용은, 신홍균, "인공지능을 이용한 무기체계의 국제인도법상 무차별성에 관한 연구"「법학논총」제35권 제3호 (2023), 105-138면.

162) Chandler, Katherine, *Does Military AI Have Gender? Understanding bias and promoting ethical approaches in military applications of AI,* UNIDIR, Geneva, 2021.

163) UN Secretary-General, Artificial intelligence in the military domain and its implications for international peace and security, UN Doc. A/80/78 (2025), paras 19-20.

164) Addressing Bias in Autonomous Weapons, UN Doc. CCW/GGE.1/2024/WP.5 (8 March 2024), para 7.

165) Alexander Blanchard & Laura Bruun, *Bias in Military Artificial Intelligence*, SIPRI Background Paper (December 2024), pp.4-7. Sources of Bias in Military AI.

누적되는데, 결과적으로 전투원과 민간인의 구별을 흐리게 하고, 특정 집단에 대한 불리한 구별(adverse distinction)을 초래함으로써, 제네바협약 공통 제3조 및 제1추가의정서 제48조, 제51조 등에서 요구하는 인도적 보호 의무를 위반할 위험을 낳는다.[166] 따라서 군사 영역에 인공지능을 도입하는 경우, 개발과 사용 단계에서 편향성을 사전에 검증하고, 제36조 신무기 검토 절차를 통해 인간의 판단과 통제를 유지하기 위한 제도적 장치를 마련하는 것은 국제인도법 준수의 관점에서 필수적으로 요구된다.[167]

이처럼 인공지능이 자율무기체계의 어느 단계에서 도입되느냐에 따라 우려의 양상이 다르게 나타날 수 있다. 가령 표적 선정 단계에서는 알고리즘의 편향성이 곧바로 구별 원칙 위반 가능성으로 연결될 수 있고, 공격 실행 단계에서는 예측불가능성이 비례성과 예방의 원칙 준수를 어렵게 할 수 있다. 따라서 새로운 법적 구속력 있는 문서가 만들어진다면, 무기체계에 도입된 인공지능이 편향된 데이터로 학습하지 않도록 사전적 예방조치에 관한 논의를 포함해야 할 것이다.

Ⅳ. 자율무기체계 관련 협약 전망

1. CCW 추가의정서로의 채택

자율무기체계에 관한 법적 구속력 있는 문서를 마련하는 방안 중 하나는 CCW의 추가의정서 형태로 채택하는 것이다. 1980년 채택된 CCW은 전투원에게 불필요하거나 정당화될 수 없는 고통을 유발하거나 민간인에게 무차별적으로 영향을 미치는 것으로 간주되는 특정 유형의 무기의 사용을 금지하거나 제한하는 것을 목적으로 한다.[168] 동 협약은 2001년 제1조 개정을 거쳐 비국제적 무력충돌 상황에도 적용되며, 기본협약과 다섯 개의 추가 의정서로 구성되어 있다.[169] 동 협약 제8조 제2항(a)는 기존 의정서에 포함되지 않는 재래식 무기

166) Ibid., pp.8-9.
167) Ibid., pp.10-11.
168) 1980년 10월 10일 채택 (당사국 128개국), 2001년 12월 21일 개정(당사국 90개국). 동 협약에 관한 내용은, https://treaties.unoda.org/t/ccw 참조.
169) CCW의 추가 의정서는 다음과 같다.
• 제1의정서(Protocol I on Non-Detectable Fragments, 1980년)는 인체 내에서 X선으로 탐지할 수 없는 파편을 사용하여 부상을 입히도록 설계된 무기의 사용을 금지한다.

에 관한 추가의정서 협상과 채택을 허용함으로써, 무기 기술의 발전에 유연하게 대응할 수 있는 제도적 장치를 마련하고 있다.[170] 이러한 구조하에 자율무기체계에 관한 논의는 지난 10년간 CCW 체제 내에서 진행되어왔으며, 새로운 법적 구속력 있는 문서 채택을 지지하는 국가들 역시 일차적으로 CCW 추가의정서로의 채택을 지지한다.[171]

그러나 전술한 바와 같이, CCW가 새로운 법적 구속력 있는 문서를 채택하기에 적합한 장인지, 나아가 실제 의정서의 채택으로 이어질 수 있는지에 대해서는 회의적 시각이 대두하고 있다. 가장 큰 이유는 CCW의 총의(consensus) 기반 의사결정 구조에 있다. CCW에서는 모든 체약국의 동의가 있어야 협상을 개시할 수 있으므로, 일부 국가들이 이를 사실상 거부권처럼 행사하여 다수의 의사를 가로막을 수 있다.[172] 실제로 자율무기체계 관련 법적 구속력 있는 문서의 체결에 부정적인 인도, 러시아 등은 새로운 협약으로의 진전을 반복적으로 차단해 왔다.[173] 이러한 구조적 한계는 관련 시민단체와 다수 국가로부터 비판을 받아 왔으며, 그 결과 CCW 외부의 별도 포럼에서 협상을 개시해야 한다는 요구가 점차 높아지고 있다.

• 제2의정서(Protocol II on Prohibitions or Restrictions on the Use of Mines, Booby-Traps and Other Devices, 1980년, 1996년 5월 3일 개정)는 지뢰, 부비트랩 및 기타 장치의 사용을 규제한다. 개정된 의정서는 비자폭 및 비자기 비활성화, 지뢰의 울타리, 감시, 표시 구역 외부 사용을 금지하고, 비탐지성 대인지뢰의 사용 및 이전을 금지한다. 또한 지뢰로 인한 무차별적 피해를 방지하기 위해, 당사국이 이러한 무기를 사용할 때 민간인 보호를 위한 모든 실행 가능한 예방조치를 취할 것을 요구한다.

• 제3의정서(Protocol III on Prohibitions or Restrictions on the Use of Incendiary Weapons, 1980년)는 소이무기의 사용을 다루며, 주로 물체에 불을 지르거나 민간인에게 화상을 입히기 위해 고안된 무기의 사용을 금지한다.

• 제4의정서(Protocol IV on Blinding Laser Weapons, 1995년)는 영구적인 실명을 유발하도록 설계된 레이저 무기의 사용 및 이전을 금지한다. 제4의정서의 채택은 무기체계가 실전 배치되기 전에 사전 예방적으로 금지된 사례이다.

• 제5의정서(Protocol V on Explosive Remnants of War, 2003년)는 전쟁잔여폭발물의 인도적 영향을 예방하고 최소화하기 위한 규정을 담고 있다. 이 의정서는 불발탄 및 버려진 폭발물의 제거·폐기, 민간인 보호조치, 폭발물 사용기록 유지, 국제협력 및 지원, 피해자 지원에 관한 구체적인 조항을 두고 있다.

170) The Convention on Certain Conventional Weapons, UN Office for Disarmament Affairs, https://disarmament.unoda.org/en/our-work/conventional-arms/convention-certain-conventional-weapons. (최종 접속일: 2026년 2월 6일)

171) 가령 스위스는 법적 구속력 있는 문서의 가능성을 열어두되, CCW 체제 내 협상을 우선시하는 입장이다. SG Report 2024, Annex I, pp.109-110.

172) Laura Varella, “Editorial: To Live Long and Prosper, We Must Ban Autonomous Weapons”, AWS Diplomacy Report, Vol.2, No.1 (May 7, 2025). p.2.

173) Human Rights Watch, Killer Robots: Negotiate Treaty in New Forum, November 10, 2022.

아울러 CCW 논의의 범위가 제한적이라는 문제도 지적된다. CCW 체제 내 자율무기체계 논의는 동 협약 제1조 제2항에 따라 무력충돌 상황과 전쟁 무기로 논의 범위가 한정된다. 체약국들은 2001년 협약 적용 검토회의를 통해 동 협약이 비국제적 무력충돌에도 적용된다는 점에 합의하였으나, 국내적 소요나 긴장 상황에는 적용되지 않는다고 보았다. 이러한 점에서 현재 진행 중인 CCW 체제 내의 논의가 국내 법집행 과정에서의 자율무기체계 사용을 포괄하지 못한다는 비판이 있다. 더불어 일부 국가들이 논의의 범위를 국제인도법에만 한정하고 국제인권법 및 형사법 또는 윤리적 논의를 배제하려 하였으며, '치명적'이라는 용어를 고수함으로써 물체 파괴나 비살상적 효과를 지닌 자율무기를 논의 대상에서 제외하려 했다는 비판도 제기된다.[174]

자율무기체계를 단순한 '무기'로 취급하여 그 금지나 제한 여부에만 초점을 맞춘다는 점도 한계로 지적된다. 자율무기체계가 인공지능 기반의 군사적·인도적 기능 체계 전체에 걸쳐 작동하는 기술이라는 점에서, 이를 협소한 무기조약 틀에서만 이해하는 것은 불충분하다.[175] 무기조약은 인간이 무기 또는 무기체계를 사용하는 방식을 규율하는 데 목적을 두고 있으므로, 무력충돌에서 민간인을 보호하는 데 필요한 다른 여러 측면을 다루지 못한다. 실제로 자율무기체계는 비전투원 식별, 민간인 피해 경감, 구호 필요 분석 등 광범위한 인도적, 군사적 기능을 수행할 수 있으므로, 무기 규제 틀로 접근하면 그 기능적 현실을 지나치게 협소화할 수 있다. 따라서 인공지능 기반 시스템이 무기체계를 넘어 훨씬 넓은 범위와 기능을 갖춘 경우, 해당 시스템에 대한 조약 논의 역시 단순히 무기에 관한 규칙을 넘어서는 보다 폭넓은 규제 틀을 갖추어야 한다는 것이다.[176] 즉, 자율무기체계는 단순한 살상 수단이 아니라 인공지능이 적용되는 광범위한 군사적·인도적 기능 체계의 일부로 이해되어야 한다.

2024년부터 CCW 체제 내 정부전문가그룹은 Rolling Text를 통해 향후 협상을 염두에 두고 문서의 구성요소를 정리하는 작업을 지속하고 있으나, 앞서 살펴본 구조적, 실질적 한계로 인해 추가의정서를 마련할 수 있을지, 그리고 그것이 과연 내용상으로도 가장 적절한 방식인지는 향후 논의의 전개를 더 지켜볼 필요가 있다.

---

174) Varella, *supra* note 172, p.2.
175) Blank, *supra* note 89, pp.247-248.
176) Ibid., p.248.

2. 대안적 조약 협상 경로의 모색

유엔 총회 및 CCW 회의에서의 국가별 발언을 종합한 자료에 따르면, 법적 구속력 있는 문서의 채택을 지지하는 국가는 130개국에 달하며, 입장을 밝히지 않은 53개국을 고려하더라도 반대 국가는 12개국에 불과하다.[177] 이처럼 다수의 국가가 자율무기체계 관련 법적 구속력 있는 문서 채택을 지지하고 있음에도 불구하고, CCW의 의사결정 구조와 체약국인 주요 군사 강대국들의 반대 입장을 고려할 때, CCW 체제 내 의정서 채택은 쉽지 않을 것으로 전망된다. 이러한 CCW 체제 내의 교착상태는 대안적 조약 협상 경로 모색을 요청한다.

이미 2022년 Human Rights Watch와 하버드 IHRC가 공동으로 발표한 보고서는 CCW 체제에서의 자율무기체계 논의가 구조적으로 진전될 수 없으므로, 그 한계를 인정하고 외부에서 새로운 협상 프로세스를 개시해야 한다고 주장하였다.[178] 동 보고서는 인도적 군축의 논리와 동력을 회복할 수 있는 현실적 선례로서 대인지뢰금지조약[179], 집속탄금지협약[180], 핵무기금지조약[181]의 협상 과정을 검토하고, 이러한 대안적 절차의 특징으로 공동의 목표 설정, 총의가 아닌 투표 방식의 의사결정, 명확한 시한 설정을 통한 효율적 절차 진행을 제시하였다.[182] 아울러 동 보고서는 시민사회와 국제기구, 그리고 관련 전문가의 법적, 기술적 지원과 참여를 통해 논의를 전개함으로써, 전통적인 국가 합의 기반 포럼에 의존하지 않는 보다 유연한 규범 형성 경로가 가능하다는 점을 강조하였다.[183]

이러한 고찰은 자율무기체계에 관한 협상 경로를 구상하는 데에도 중요한 시사점을 제공한다. 먼저 유엔 총회를 생각할 수 있다. 유엔 총회는 포괄적 회원 구성을 갖추고 있다는 점에서 자율무기체계 관련 협약 검토에 있어 포용성이 가장 높은 장이다.[184] 나아가 2023년 이후 치명적 자율무기체계가 의제로

177) 반대 국가에는 대한민국, 호주, 에스토니아, 일본, 러시아, 벨라루스, 인도, 폴란드, 영국, 조선민주주의인민공화국, 이스라엘, 미국이 포함된다. State positions, Automated Decision Research, https://automatedresearch.org/state-positions/?_state_position_negotiation=no (최종 접속일: 2026년 2월 19일).

178) Human Rights Watch & IHRC, *An Agenda for Action: Alternative Processes for Negotiating a Killer Robots Treaty* (November 10, 2022).

179) Ibid., pp.15-18.

180) Ibid., pp.18-21.

181) Ibid., pp.22-25.

182) Ibid., pp.26-29.

183) Ibid., pp.29-31.

상정되어 연속적으로 결의가 채택되고 있다는 점은, 총회가 자율무기체계에 대한 국제적 우려를 확인하는 것을 넘어 향후 외교 회의의 개최 기반을 마련하거나 작업반 또는 협의체를 구성하여 보다 구조화된 논의를 진전시킬 수 있는 가능성을 보여준다. 다만, 총회 차원의 논의가 곧바로 정식 협상 절차로 전환되는 것은 아니며, 후속 절차가 자동적으로 보장되는 것도 아니다.

다음으로, 자율무기체계 규율의 필요성과 인도주의 원칙을 공유하는 국가군이 규범 형성 선도그룹을 형성하여 독립적인 다자 협상을 개시하는 방안을 생각해볼 수 있다. 이러한 경로에서는 ICRC나 민간 전문가그룹의 법적·기술적 지원 아래 초안을 마련하는 방식으로 구체화될 수 있으며, 이후 유엔 총회의 장에서 공식화되거나 권위를 부여받을 수 있다. 이러한 맥락에서 서론에서 언급한 공개 비공식협의는 자율무기체계와 관련된 인도적, 법적, 윤리적, 안보적 우려의 전 범위를 논의하고, 필요성에 관한 공감대를 확장할 수 있다는 점에서 의의가 있다. 그러나 2025년 유엔 총회 결의가 공개 비공식협의의 추가 개최를 결정하지 않은 상황은 대안적 협상 경로가 순탄하지 않을 수 있음을 시사한다.

3. 연성법적 성격 문서 채택의 의의

자율무기체계에 관한 국제적 규범 형성의 필요성은 여전하며, 단일한 경로에 의존하기보다는 복수의 경로를 상호 보완적으로 결합하는 방향이 보다 현실적일 수 있다. 실제로 새로운 법적 구속력 있는 문서 마련을 지지하는 국가군 중 일부는 CCW 체제 내 논의를 우선시하면서도, 이와 병행하여 진행되는 구상이 정부전문가그룹의 작업을 보완하고 보다 포용적인 접근방식을 촉진한다는 점을 강조한다.[185] 나아가 핀란드와 같이 정치적 또는 법적 문서의 형태로 선택지를 열어두는 경우도 있다.[186]

이러한 맥락에서 자율무기체계 규제와 관련하여 법적 구속력을 갖춘 협약 이외에도, 정치적 선언, 행동강령, 모범 관행 등 연성법적(soft law) 형태 문서의 의의도 살펴볼 필요가 있다.[187] 이러한 문서는 전통적 연원론의 관점에서

184) SG Report 2024, para 62.
185) SG Report 2024, p.61.
186) SG Report 2024, Annex I, p.46.
187) 국제법 담론에서 연성법은 “확정적 개념 범주 및 일관된 법리를 가진 단일 개념이 아니라 다양하고 상이한 다수의 문서형태를 포섭하고 이를 단일명칭으로 명명하기 위해 사용된 개념적 도구”이다. 정경수, “국제법상 연성법의 재인식”, 「안암법학」 제34권 (2011),

법적 구속력은 없지만, 법원칙을 제공하는 방식으로 규범적 역할을 수행하거나 조약 체결 과정의 일부로 기능할 수 있고, 나아가 관습국제법의 형성요소로도 작용할 수 있다.[188] 연성법적 접근은 구속력 있는 협약 체결이 지연되는 현실을 고려할 때, 국제적 기대 수준을 형성하고 해석의 기준을 마련하며 국가 관행을 축적하는 측면에서 실효적인 경로가 될 수 있다. 비구속적인 연성법적 문서는 합의에 도달하기 비교적 용이하다는 장점도 지닌다. 따라서 신기술이나 새로운 이슈에 대해 "'실험적 규범'을 설계하고 조율하는 데 유리"하게 작용할 수 있다.[189] 앞서 살펴본 바와 같이 2019년 정부전문가그룹이 채택한 11개 지도원칙은 자율무기체계에 관한 국제적 합의의 최소 기준을 제시하며, 이후 여러 국제 문서에 인용될 뿐만 아니라 개별 국가의 정책 및 군사 지침에도 반영되고 있다.[190] 유엔 총회의 결의 역시 자율무기체계에 관한 국제적 논의의 방향과 기준을 제시한다는 점에서 그 의의를 찾을 수 있다.

실제로 ICRC는 새로운 법적 구속력 있는 문서와 병행하여 공동의 정책 기준 및 모범사례 지침을 마련함으로써 자율무기체계를 규율하는 방안을 제안하였다.[191] 이는 자율무기체계가 본질적으로 다양한 기술적 구성요소를 포함한다는 점을 고려할 때, 구속력 있는 조약보다 신속하고 유연하게 대응할 수 있는 수단으로 기능할 수 있다. 아울러 자율무기체계에 국제인도법을 적용할 때에는 그 고유한 특성을 고려한 해석 지침이 필요할 수 있는데, 국제인도법 규칙의 적용에 관한 명확화 및 구체화는 연성법적 문서를 통해 선행될 수 있다. 특히 예측가능성, 추적가능성, 설명가능성과 같이 기존 국제인도법에서 다루지 않은 기술적 개념을 지침을 통해 구체화할 수 있다는 장점이 있다. 궁극적으로 지침

944면.

188) 전자는 정태적 측면의 일부이고 후자는 동태적 측면의 일부이다. 관련 내용은, 위의 논문, 946-952면.

189) 관련 내용은, 이주형, "인공지능(AI) 관련 규범과 연성법의 역할에 관한 소고 - 디지털 경제협정을 중심으로", 「사법」 제72호 (2025), 11면.

190) 예컨대, 다자주의 연합(Alliance for Multilateralism) 소속 약 40개 유엔 회원국은 11개 지도원칙을 지지하며 규범적·운용적 틀의 명확화와 발전에 기여할 것을 촉구하였다. 앞의 주, 38) 참조. Matthew Breay Bolton et al., Addressing the Threat of Autonomous Weapons: Maintaining Meaningful Human Control (Friedrich-Ebert-Stiftung New York Office, January 2021), p.1. 2024년 사무총장 보고서에서도 다수의 국가가 2019년 지도원칙을 주요한 성과로 평가하였으며, 동 원칙에 기초하거나 부합하는 각국의 입장을 소개하였다. 예를 들어, 독일은 자국의 자율무기체계 관련 입장이 2019년 지도원칙 및 2021년 NATO의 책임 있는 AI 사용 원칙을 토대로 한다고 밝혔다. SG Report 2024, Annex I, p.50.

191) ICRC Position on Autonomous Weapon Systems, 12 May 2021.

은 국가를 법적으로 구속하지 않더라도 특정 분야에서 반복적이고 일관된 정책 기준의 형성에 기여함으로써, 그러한 관행을 통해 점차 관습국제법으로 발전하여 모든 국가에 의무로 작용할 여지도 있다.[192] 물론 군사 분야에서 연성법적 문서에 실질적 규율 효과를 과대하게 기대하기는 어렵지만, 위에서 살펴본 바와 같이 지침으로서의 역할과 함께 향후 조약 협상의 출발점이 될 수 있다는 점에서 그 의의를 찾을 수 있다.

V. 결론

인공지능 기반 전쟁과 자율무기체계가 '오펜하이머의 순간'에 도달했다는 표현이 자주 사용된다.[193] 현재 자율무기체계는 공식적, 비공식적으로 사용되며, 전쟁 수행방식의 패러다임을 바꿔나가고 있다. SF 영화 속 터미네이터와 같이 완전한 자율성을 갖춘 무기체계가 아니더라도, 인간의 의미 있는 통제와 예측 가능성을 벗어난 무기체계는 이미 존재한다고도 볼 수 있다. 자율성이 발전할수록 무기체계에 대한 인간의 역할은 적극적 조종자에서 수동적 감시자로 전환되고 있다.[194]

국제사회에서 자율무기체계에 대한 논의는 인권 및 군축 관련 시민단체, 학계, 그리고 관련 전문가 집단의 우려에서 시작되어 유엔 차원의 논의로 발전해 왔다. 2014년부터 진행된 CCW 체제 내 논의는 2019년 11개의 지도원칙을 채택하였으며, 2024년부터는 핵심 요소에 관한 합의를 담아내는 Rolling Text 작업 단계로 나아가고 있다. 그러나 2025년 12월까지 마련된 Rolling Text를 보면, 초기 논의에 비해 금지의 범위가 오히려 축소되는 흐름이 확인된다. '모든 상황에서'의 금지 표현이 삭제되고, 대인 표적 자율무기체계의 명시적 금지로는 논의가 나아가지 못하고 있다는 점은 CCW 체제 내 논의의 실질적 한계를 보여준다. 유엔 총회는 2023년 「치명적 자율무기체계」를 공식 의제로 설정하여 최초의 결의를 채택하였고, 2024년 결의에서는 모든 유엔 회원국 및 관련 이해관계자들과의 포괄적 논의를 위한 공개 비공식협의 개최를 결정하였다. 2025년 5월 첫 공개 비공식협의는 CCW 체제 내의 논의를 보완하는 한편, 궁극적으로

192) Sterio, *supra* note 129, p.313.
193) Nick Robins-Early, "AI's 'Oppenheimer moment': autonomous weapons enter the battlefield", *The Guardian,* 14 July 2024.
194) Sterio, *supra* note 129, pp.303-304.

CCW 체제 밖에서의 논의 가능성을 열어두는 계기가 되었다. 대한민국을 포함한 일부 국가들이 명시적으로 공개 비공식협의의 장을 반대하고, 2025년 유엔 총회는 추가 회의 개최를 결의하지 않았지만, 이는 오히려 CCW의 추가의정서 이외의 대안적 협상 경로를 모색해야 할 필요성을 부각시켰다.

이러한 논의의 전개 속에서 자율무기체계에 관한 현재 국제적 논의의 핵심은 기존 국제인도법의 적용과 준수 여부를 넘어, '새로운 법적 구속력 있는 문서'의 마련이다. 국제사회는 자율무기체계의 합의된 특성을 바탕으로 범위를 설정하고자 노력하고 있으며, 금지와 규제의 이원적 접근방식, 인간통제 원칙, 기계학습으로 인한 예측불가능성 등 여러 핵심 쟁점에 대해 문제의식을 공유하고 있다. 다만 각 쟁점에 대한 견해차는 협약 마련을 어렵게 하는 실질적 장벽으로 남아 있다. 그럼에도 불구하고 무기체계의 자율성이 무제한일 수 없으며, 인간의 판단 및 통제, 그리고 책임을 유지해야 한다는 점에 대해서는 이미 공감대가 형성되어 있다.[195] 2025년까지 전개된 논의는 이전의 막연한 '자율성' 논의를 넘어, 기계학습 알고리즘에 의해 발생하는 자율무기체계의 구체적 작동방식과 그 위험을 분석하는 단계로 나아가고 있다. 규제의 필요성과 그 형태에 관해서는 국가 간 견해차가 존재하지만, 핵심 쟁점에 관한 국제적 합의를 바탕으로 법적 구속력 있는 문서의 마련을 시도하는 흐름이 분명히 나타나고 있다.

진정한 인간통제는 완전한 자율무기체계가 일단 개발, 사용된 이후에는 구현될 수 없다. 이 점에서 선제적 규율의 필요성은 아무리 강조해도 지나치지 않다. 대안적 협상 경로와 관련하여, 유엔 총회를 통한 외교회의 개최나 작업반 구성뿐만 아니라, 규율의 필요성과 인도주의 원칙을 공유하는 선도국 그룹이 별도의 다자 협상을 개시하는 방안도 현실적인 경로로 검토될 수 있다. 아울러 법적 구속력 있는 협약 체결이 지연되는 현실을 고려할 때, 정치적 선언, 행동강령, 모범 관행과 같은 연성법적 문서의 역할도 과소평가해서는 안 된다. 이는 국제적 기대 수준을 형성하고, 해석의 기준을 마련하며, 국가 관행을 축적하게 할 수 있다. 설령 충분한 관행이 축적되지 않은 선제적 규제라 하더라도, 조약의 제정은 유사한 이해를 가진 국가들의 협상과 정치적 타협을 통해 가능하며, 국제기구의 권고와 결의는 그 추진력을 제공할 수 있다.[196] 이러한 측면에서 2023년 이후 유엔 총회가 자율무기체계를 공식 의제로 다루기 시작

195) Boulanin et al., *supra* note 139, pp.1-2.

196) 박기갑, "국제적 권고, 지침 분석에 바탕을 둔 "인공지능(AI) 윤리" 관련 국제조약안의 모색", 「국제법학회논총」, 제67권 제4호 (2022), 123면.

했다는 점은 주목할 만하다. 포용성을 갖춘 유엔 총회는 공동의 목표를 설정하기 유리하며, 투표 방식의 의사결정을 기반으로 한다는 점에서 CCW 체제에서의 의정서 채택보다 유리한 조건을 갖추고 있다. 인류 공동의 이익과 목표를 바탕으로, CCW 논의를 보완하는 수준을 넘어 보다 진전된 논의가 유엔 총회에서 적시에 전개되기를 바란다.

[참고문헌]

1. 단행본

- 국내문헌

박문언, 「자율무기체계의 국제법적 허용성과 규제방안」 (박사학위논문, 서울대학교, 2019)

헨리 키신저, 에릭 슈밋, 크레이그 먼디(이현 옮김), 「새로운 질서: AI 이후의 생존 전략」 (서울: 월북, 2025)

- 해외문헌

Blanchard, Alexander & Goussac, Netta, Towards Multilateral Policy on Autonomous Weapon Systems (SIPRI, September 2025)

Blanchard, Alexander & Bruun, Laura, Bias in Military Artificial Intelligence, SIPRI Background Paper (December 2024)

Bolton, Matthew Breay et al., Addressing the Threat of Autonomous Weapons: Maintaining Meaningful Human Control (Friedrich-Ebert-Stiftung New York Office, January 2021)

Boulanin, Vincent et al., *Limits on Autonomy in Weapon Systems: Identifying Practical Elements of Human Control,* SIPRI & ICRC (June 2020)

Brehm, Maya, *Defending the Boundary: Constraints and Requirements on the Use of Autonomous Weapon Systems under International Humanitarian and Human Rights Law* (The Geneva Academy of International Humanitarian Law and Human Rights, May 2017)

Bruun, Laura, *Towards a Two-Tiered Approach to Regulation of Autonomous Weapon Systems: Identifying Pathways and Possible Elements* (SIPRI, August 2024)

Chandler, Katherine, *Does Military AI Have Gender? Understanding bias and promoting ethical approaches in military applications of AI* (UNIDIR, Geneva, 2021)

Deeks, Ashley S., *The Double Black Box* (Oxford University Press, 2025)

Geiss, Robin, *The International-Law Dimension of Autonomous Weapons Systems* (Friedrich Ebert Stiftung, October 2015)

Human Rights Watch & IHRC, Losing Humanity: The Case against Killer Robots (November 2012)

Human Rights Watch & IHRC, An Agenda for Action: Alternative Processes for Negotiating a Killer Robots Treaty (November 2022)

2. 논문

- 국내문헌

김민혁, 김재오, “자율살상무기체계에 대한 국제적 쟁점과 선제적 대응 방향”, 「국방연구」 제63권 제1호 (2020)

김현중, “넥스트 오펜하이머 시대: 자율살상무기 발전에 따른 예상쟁점 및 대응방안”, 「INSS 전략보고」 No.284 (September 2024)

박기갑, “국제적 권고, 지침 분석에 바탕을 둔 “인공지능(AI) 윤리” 관련 국제조약안의 모색”, 「국제법학회논총」 제67권 제4호 (2022)

신홍균, “인공지능을 이용한 무기체계의 국제인도법상 무차별성에 관한 연구”, 「법학논총」 제35권 제3호 (2023)

유준구, “자율살상무기체계의 논의 동향과 쟁점”, 정책연구시리즈 2019-18 (국립외교원 외교안보연구소, 2019)

이주형, “인공지능(AI) 관련 규범과 연성법의 역할에 관한 소고 - 디지털경제협정을 중심으로”, 「사법」 제72호 (2025)

임예준, “인공지능 시대의 전쟁자동화와 인권에 관한 소고 - 국제법상 자율살상무기의 규제를 중심으로 -”, 「고려법학」 제92호 (2019)

정경수, “국제법상 연성법의 재인식”, 「안암법학」 제34권 (2011)

조현식, “인공지능, 자율무기체계와 미래 전쟁의 변환”, 「21세기정치학회보」 제28집 제1호 (2018)

- 해외문헌

Acheson, Ray, “Editorial: From “Constructive Ambiguity” to Unambiguous Destruction”, CCW Report, Vol.9, No.9 (2021)

Alston, Philip, “Lethal Robotic Technologies: The Implications for Human Rights and International Humanitarian Law”, *Journal of Law, Information and Science*, Vol.21 (2011/2012)

Article 36, Critical Commentary on the “Guiding Principles”, Policy Note,

November 2019

Blank, Laurie R., “New Treaty Law on Autonomous Weapons? An Opportunity to Reframe the Discourse”, *Case Western Reserve Journal of International Law,* Vol.57 (2025)

Bruun, Laura, “The Group of Governmental Experts on Lethal Autonomous Weapon System”, in *Conventional Arms Control and Regulation of New Weapon Technologies, Non-Proliferation, Arms Control and Disarmament,* 2020

Davison, Neil, “A legal perspective: Autonomous weapon systems under international humanitarian law”, in *UN, UNODA Occasional Papers No.30, Perspectives on Lethal Autonomous Weapon Systems* (UN, January 2018)

Docherty, Bonnie, “Autonomous Weapon Systems and Threats to Human Rights”, *WS Diplomacy Report*, Vol.2, No.1 (May 7, 2025)

Heyns, Christof, “Human Rights and the Use of Autonomous Weapons Systems (AWS) During Domestic Law Enforcement”, *Human Rights Quarterly*, Vol.38 (2016)

Martin, Craig, “Autonomous Weapons Systems and Proportionality: The Need for Regulation”, *Case Western Reserve Journal of International Law,* Vol.57 (2025)

Minor, Elizabeth & Moyes, Richard, “Regulating Autonomy in Weapons Systems”, Article 36 (21 October 2020)

Minor, Elizabeth, “Opportunities after the UNGA Resolution on Autonomous Weapons: Moving Toward a New Treaty”, Article 36 (December 22, 2024)

Perrin, Benjamin, “Lethal Autonomous Weapons Systems & International Law: Growing Momentum Towards a New International Treaty”, *ASIL Insights,* Vol.29, Issue 1 (January 24, 2025)

Sayler, Kelley M., Defense Primer: U.S. Policy on Lethal Autonomous Weapon Systems, *CRS In Focus*, January 2, 2025

Sterio, Milena, “Autonomous Weapons Systems and the Need to Update International Humanitarian Law?”, *Case Western Reserve Journal of International Law,* Vol.57 (2025)

Taddeo, Mariarosaria & Alexander Blanchard, “A Comparative Analysis of the Definitions of Autonomous Weapons Systems”, *Science & Engineering Ethics,* Vol.28 (2022)

Varella, Laura, “Editorial: To Live Long and Prosper, We Must Ban Autonomous Weapons”, *AWS Diplomacy Report*, Vol.2, No.1 (May 7, 2025)

3. 국제기구 및 ICRC 문서

GA Res. 78/241, Lethal Autonomous Weapons Systems, UN Doc. A/RES/78/241 (28 December 2023)

GA Res. 79/62, Lethal Autonomous Weapons Systems, UN Doc. A/RES/79/62 (2 December 2024)

GA Res. 80/57, Lethal Autonomous Weapons Systems, UN Doc. A/RES/80/57 (5 December 2025)

Interim Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, Philip Alston, UN Doc. A/65/321 (2010)

Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, Christof Heyns, UN Doc. A/HRC/23/47 (2013)

UN Secretary-General, President of ICRC Jointly Call for States to Establish New Prohibitions, Restrictions on Autonomous Weapon Systems, SG/2264, Press Release (5 October 2023)

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

UN Secretary-General, Artificial intelligence in the military domain and its implications for international peace and security, UN Doc. A/80/78 (2025)

Meeting of the High Contracting Parties to the CCW, Final Report, CCW/MSP/2013/10 (16 December 2013)

CCW Fifth Review Conference, Report of the 2016 informal meeting of experts on lethal autonomous weapons systems, CCW/CONF.V/2 (10 June 2016)

Report of the 2018 session of the GGE, CCW/GGE.1/2018/3 (23 October 2018)

Report of the 2019 session of the GGE, CCW/GGE.1/2019/3 (25 September 2019)

Meeting of the High Contracting Parties to the CCW, Final Report, CCW/MSP/2023/7 (23 November 2023)

Addressing Bias in Autonomous Weapons, UN Doc. CCW/GGE.1/2024/WP.5 (8 March 2024)

Chair's summary – First 2025 session of the GGE on LAWS, UN Doc. CCW/GGE.1/2025/WP.1 (7 April 2025)

GGE on LAWS, Rolling Text, status date: 18 December 2025

ICRC, Expert Meeting: Autonomous Weapon Systems, Technical, Military, Legal and Humanitarian Aspects, Geneva, Switzerland (26 to 28 March 2014)

ICRC Position on Autonomous Weapon Systems (12 May 2021)

ICRC Position on Autonomous Weapon Systems, Background Paper (2021)

[Abstract]

Autonomous Weapon Systems and the Prospects for an International Treaty: Key Issues and Negotiating Pathways

Yejoon Rim*

The central question in current international discussions on Autonomous Weapon Systems (AWS) has shifted from whether existing international humanitarian law (IHL) applies, to whether a new legally binding instrument is needed. The international community is working to define the scope of AWS on the basis of agreed characteristics, and shares a common understanding of key issues, including a two-tiered approach of prohibition and regulation, the principle of human control, and the unpredictability arising from machine learning. There is broad consensus that the autonomy of weapon systems cannot be unlimited and that human judgment, control, and accountability must be preserved. Given that reaching agreement on the scope of prohibition and regulation and the form of a legal instrument will not be straightforward, it is necessary to examine in which forum international discussions on the regulation of AWS will unfold. This article examines the difficulties of adopting an additional protocol within the CCW framework, and considers the need to explore alternative negotiating pathways through the UN General Assembly, as well as the significance of adopting soft law instruments prior to the conclusion of a binding treaty.

Meaningful human control cannot be assured once fully autonomous weapon systems have been developed and deployed. Even in the absence of sufficient state practice, a treaty remains achievable through negotiation and political compromise among like-minded states, with the recommendations and resolutions of international organizations providing essential momentum. In this regard, the UN General Assembly's decision to place AWS on its formal agenda

* Associate Professor Korea University School of Law

since 2023 is significant. As an inclusive forum operating on the basis of majority voting, the General Assembly offers more favorable conditions for advancing a legally binding instrument than the consensus-based CCW process. It is expected that the General Assembly will move beyond merely supplementing CCW discussions and facilitate meaningful and timely progress toward a treaty that reflects the common interests and values of humanity.

Keywords : Autonomous Weapon Systems, International Humanitarian Law, Human Control, Predictability, Killer Robots