

자율무기의 법과 윤리

Law and Ethics of Autonomous Weapon System

이 상 용*
Sang Yong Lee

목 차

- | | |
|---------------------------------|---------------|
| I. 서론 | IV. 윤리적 통제 방안 |
| II. 자율무기체계의 확산과 이에 대한 통제
필요성 | V. 법적 통제 방안 |
| III. 자율무기 개념의 구체화 | VI. 결론 |

AI를 비롯한 기술의 발전에 따라 세계 각국이 자율무기체계 개발에 나서고 있다. 자율무기체계는 그 정확성으로 인해 인명 손실과 재산 피해를 줄일 수도 있지만, 전쟁의 문턱을 낮추고, 인간의 존엄성을 위협하며, 책임의 공백을 가져올 것이라는 우려 때문에 반대하는 견해도 많다. 그러나 단순한 찬반론을 넘어 ‘의미 있는 인간 통제’를 확보하는 것이 중요하다. 그 통제 방안으로는 윤리적 통제와 법적 통제가 있을 수 있다.

그러나 논의의 전제로서 우선 자율무기체계의 개념을 정립할 필요가 있다. 이에 관하여 많은 논쟁이 있어 왔지만, 미 국방부의 입장처럼 기능적·관계적 관점에 따라 ‘일단 작동되면 운용자에 의한 추가적인 개입 없이도 표적을 선택하고 공격할 수 있는 무기체계’ 정도로 정의하는 것이 적절할 것이다. 한편 완전자율무기체계 개념을 둘러싼 소모적인 논란에 휘말리기보다는 의미 있는 인간 통제를 실현하는 방안에 주목해야 할 것이다.

윤리적 통제를 위해서는 윤리이론을 바탕으로 실천적 윤리원칙을 마련하는 방안을 검토할 필요가 있다. 그동안 제안되어 온 AI 윤리원칙들을 바탕으로 군사 분야의 특수성을 고려하여 윤리원칙을 마련할 수 있다. CCW의 자율무기체계 지도원칙이나 미 국방부의 AI 윤리원칙 등이 그 사례이다. 모범 관행의 수립이나 정치적 선언 역시 윤리원칙과 마찬가지로 구속력 없는 연성 규범으로서 어느 정도 효과를 발휘할 수 있다.

<https://doi.org/10.35148/ilsilr.2023..56.351>

투고일: 2023. 10. 11. / 심사완료일: 2023. 12. 26. / 게재확정일: 2023. 12. 27.

* 건국대학교 법학전문대학원 교수

Professor, Law School of Konkuk University

법적 통제는 국제법과 국내법으로 나누어 검토할 수 있다. 다만 그 전에 자율무기체계의 법적 지위가 문제될 수 있는데, 행위성과 책임주체성이 인정되기는 어려울 것이다. 국제법적 측면에서는 국제인도법의 원칙들과 그 위반에 대한 민형사상의 책임이 문제된다. 자율무기체계는 국제인도법상 그 자체로 금지되는 무기라고는 할 수 없지만, 고차원적인 판단이 요구되는 상황에서는 인간의 도움 없이 구별의 원칙과 비례의 원칙을 준수하기는 어려우므로 일정한 경우 사용방법 상의 제한이 이루어질 필요가 있다. 자율무기체계를 사용한 전쟁범죄에 대하여는 로마규정에 따른 형사책임이 인정되는데, 다만 자율무기체계를 부하로 취급하여 지휘관책임을 물을 수는 없다. 자율무기체계의 위법한 사용으로 인한 피해자는 관련자들을 상대로 민사책임을 물을 수도 있다.

국내법적 통제로서는 사후적으로 민형사상의 책임을 묻는 것 외에 자율무기체계의 운용과 통제에 관한 사전적 규율을 마련하는 방안이 검토되어야 한다. 미국은 국방부 훈령을 통해 ‘적절한 수준의 인간 판단의 행사’를 기본원칙으로 하여 자율무기체계를 개발·운용하고 실패의 영향과 확률을 최소화하기 위한 규율을 마련하였다. 우리도 이를 참고하여 자율무기체계에 관한 개념, 전략 및 전술, 절차를 확립하여 행정법규나 지침으로 문서화할 필요가 있다.

[주제어] 자율무기, 자율살상무기, 치명적 자율무기, 자율무기체계, 국제인도법

I. 서론

지난 2023년 7월 미 공군의 차세대 드론 XQ-58A Valkyrie의 첫 비행이 있었다.¹⁾ 위 드론은 가시거리 바깥의 표적을 타격할 수 있는 장거리 미사일을 탑재한 채 음속에 가까운 속도로 거의 4000km를 비행할 수 있다. 무엇보다 AI가 스스로 주변 위협요소를 식별하여 표적을 선택한 뒤 인간의 승인이 떨어지면 공격하는 기능을 갖추고 있다. 먼 미래의 일일 것 같았던 자율무기체계(Autonomous Weapon System: AWS)는 이렇게 우리 곁에 바짝 다가와 있다.

자율무기체계는 인명 손실과 재산 피해를 줄일 수도 있지만, 전쟁의 문턱을 낮추고 인간의 존엄을 위협할 것이라는 우려도 제기된다. 치열한 논쟁에도 불구하고 세계 각국은 이미 자율무기체계 개발을 위한 경쟁에 나서고 있다. 이제는 찬반론을 떠나 실질적이고 효과적인 규율 방안을 마련하는 것이 현실적인 과제가 되었다. 우리 정부 역시 미·중 간의 갈등, 핵·미사일·무인기를 비롯한 북한의 비대칭 전력 증대를

1) New York Times, “A.I. Brings the Robot Wingman to Aerial Combat”, 2023. 8. 27, <<https://www.nytimes.com/2023/08/27/us/politics/ai-air-force.html>>, 검색일 2023. 10. 11.

비롯한 안보환경의 변화²⁾에 대응하기 위하여 2023년 3월 「국방혁신 4.0 기본계획」을 발표하면서 유·무인 복합전투체계를 단계적으로 구축하겠다는 목표를 포함시켰다.³⁾ 또한 같은 해 2월에는 네덜란드와 공동으로 ‘인공지능의 책임 있는 군사적 이용에 관한 고위급회의(Responsible AI in the Military Domain Summit: REAIM)’를 개최함으로써 자율무기체계의 규율에 관한 국제적 논의에서 주도적인 역할을 하려는 모습을 보이기도 했다.⁴⁾ 사실 우리나라는 이미 2006년 비무장지대 경계를 목적으로 자율 모드를 갖춘 센트리 로봇(SGR-1)을 개발·시험해본 경험이 있는 기술적 잠재력을 갖춘 국가이기도 하다.⁵⁾

그러나 안타깝게도 우리나라는 아직 자율무기체계의 운용과 통제를 위한 개념, 전략 및 전술, 절차에 관하여 충분한 연구나 준비가 되어 있지 않은 것이 현실이다. 십여 년 전부터 풍부한 논의가 이루어져 온 외국과 달리 학계의 논의도 빈약한 편이다. 특히 자율무기체계를 둘러싼 문제의 핵심이 규범적 통제임에도 불구하고 법학계의 관심은 매우 적었던 것이 사실이다. 예를 들어 최근 관심을 모으고 있는 EU 인공지능 법안은 군사적 목적을 위한 인공지능 시스템에는 적용되지 않으나⁶⁾ 이런 사실은 많이 알려져 있지 않다. 본고는 그동안 국내외에서 이루어져 온 자율무기체계의 윤리적, 법적 통제에 관한 논의들을 종합하여 장차 실질적이고 효과적인 규율 방안을

2) 북한 인공지능 기술 군사화의 현황에 관하여는 김민혁, “북한 인공지능 기술의 군사화와 우리 군의 대응 무기체계 발전방향 연구”, 한국IT서비스학회지 제20권 제1호, 한국IT서비스학회, 2021 참조.

3) 1단계는 원격통제형 중심, 2단계는 반자율형 체계 시범, 3단계는 반자율형 체계 확산 및 자율형 체계 전환으로 구분하여 추진된다. 국방부 보도자료, 제2창군 수준의 국방 재설계, AI과학기술강군 육성: 윤석열 정부의 「국방혁신 4.0 기본계획」 발표, 국방부, 2023 참조.

4) 국제기구국, 박진 외교장관, 「인공지능의 책임 있는 군사적 이용에 관한 고위급회의(REAIM2023)」 참석, 외교부, 2023 참조.

5) 일부 문헌에 소개된 것과 달리 실전 배치되지는 않았다(박문언, “자율무기체계에 대한 국제적 논의와 우리 군의 대비방향”, 국방논단 제1920호, 한국국방연구원, 2022, 7쪽).

6) European Commission, Proposal for a Regulation of the European Parliament and of the Council: Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts (COM/2021/206 final) Article 2. 3. EU 집행위원회가 초안을 제출한 이후 생성형 인공지능이나 대규모 언어모델을 비롯한 새로운 인공지능 기술이 확산함에 따라 기반모델(foundation model)에 관한 규율 등을 추가한 수정안이 2023. 6. 14. 유럽의회에서 가결되었다. 그 후 2023. 12. 9. EU 집행위원회와 EU 의회, 그리고 유럽위원회 등 3개 기관이 잠정적으로 법안에 합의함으로써 역사적인, 그러나 논쟁적인 세계 최초의 AI 규제 입법이 가시화되었다. The European Council The Council of the EU, “Artificial intelligence act: Council and Parliament strike a deal on the first rules for AI in the world”, <<https://www.consilium.europa.eu/en/press/press-releases/2023/12/09/artificial-intelligence-act-council-and-parliament-strike-a-deal-on-the-first-worldwide-rules-for-ai/>>, 검색일: 2023. 12. 27.

마련하는 데 도움을 주기 위하여 작성되었다.

II. 자율무기체계의 확산과 이에 대한 통제 필요성

1. 자율무기체계의 확산

세계의 군사 선진국들은 드론과 같은 원격무기는 물론 자율무기체계의 개발에도 열을 올리고 있다. 여기에는 전투원의 손실과 무력충돌 과정에서 발생하는 인간의 희생을 최소화하려는 동기가 자리잡고 있다. 자율무기체계는 인간보다 월등한 정확성을 지닐 뿐 아니라 인간과 달리 신체적 제한이나 감정적 영향에서 자유롭다. 이러한 자율무기체계의 능력은 딥러닝 등 AI 기술의 발전에 힘입은 바 크다. AI 기술은 판단과 의사결정을 자동화하는 범용 기술로서 군사 분야에서도 무기체계, 전력지원체계, 운영유지 등 전 영역에서 활용될 수 있다.⁷⁾

많은 나라들이 자율무기체계 개발 현황의 공개를 꺼리고 있지만, 그동안 밝혀진 것만으로도 자율무기체계가 급속도로 확산되고 있음을 알 수 있다. 오늘날 활용되고 있는 자율무기시스템은 자동적 방어체계와 자동적 공격체계로 나누어볼 수 있다.⁸⁾ 전자에는 미 해군의 Phalanx CIWS, 미 육군의 C-RAM, 이스라엘의 Iron Dome, 독일의 NBS Mantis, 네덜란드의 Goalkeeper, 러시아의 Kashtan 등과 같은 근접방어무기들이 포함되며, 미 육군의 Patriot, 미 해군의 AEGIS System도 이에 포함될 수 있다.⁹⁾ 후자에 속하는 것으로는 앞에서 보았던 미 공군의 차세대 드론 XQ-58A, 미 해군의 무인공중전투체계인 X-47B, 이와 유사한 영국의 Taranis, 적 레이더를 탐지 공격하는 이스라엘의 Harpy¹⁰⁾ 등이 포함될 수 있을 것이다. 이들은 긴급모드나 자율모드에서 동작할 경우 자율적으로 표적을 획득하고 교전할 수 있는 능력을 갖추고 있다. 우리

7) 이종관/한창희, “미래전과 국방 인공지능 체계”, 한국통신학회논문지 제44권 제4호, 한국통신학회, 2019, 785-787쪽.

8) 한희원, “인공지능(AI) 치명적 자율무기(LAWs)의 법적·윤리적 쟁점에 대한 기초연구”, 중앙법학 제20집 제1호, 중앙법학회, 2018, 328-331쪽.

9) 전통적인 지뢰 역시 자동적 방어체계에 속하지만 이를 자율무기시스템으로 다루지는 않는 것이 일반적이다.

10) 순찰 지역을 자율적으로 운항하다가 적 레이더를 탐지하여 공격할 수 있다. 이보다 길이가 조금 긴 하롭(Harop)에는 공격 중지 기능이 탑재되어 있다.

나라의 SGR-1이나 이스라엘이 가자지구 경계지역에 배치한 Sentry Tech System 역시 자율모드로 작동할 수 있다. 이 밖에도 각국에서 자율무기시스템의 개발이 한창이다.¹¹⁾

2. 자율무기체계에 관한 논의

2.1 논의의 경과

자율무기체계에 관한 문제 제기는 2000년대부터 몇몇 학자들에 의하여 이루어져 왔지만,¹²⁾ 2010년대 초반 알스톤(P. Alston)과 헤인즈(C. Heyns)가 UN에 자율무기체계의 위험성을 경고하는 보고서들을 제출한 것을 계기로 본격적인 논의가 이루어지기 시작했다.¹³⁾ 이들은 오늘날 자율무기체계와 관련된 대부분의 이슈를 다루면서 이후의 논의를 위한 토대를 마련했다. 비슷한 시기에 국제로봇무기통제위원회(ICRAC), 휴먼라이츠워치(HRW), 국제여성평화·자유연맹(WILPF), 국제적십자위원회(ICRC) 등의 비정부기구들도 자율무기의 제한이나 통제의 목소리를 높였다.¹⁴⁾

그 후의 논의는 주로 특정재래식무기금지협약¹⁵⁾(이하 ‘CCW’라고 한다) 기구를 통해 이루어졌다. CCW는 2014년부터 비공식 정부전문가회의를 통해, 그리고 2017년부터는 매년 개방형 정부전문가그룹¹⁶⁾을 통해 자율무기체계와 관련된 논의를 주도하고 있다. 정부전문가그룹은 2019년 자율살상무기와 관련된 11가지 지도원칙을 제

11) 러시아의 현황에 관하여는 이수훈/유영철, “러시아 안보·국방 분야의 인공지능 기술 도입과 개발 현황: 분석과 한계”, 중소기업 제44권 제3호, 한양대학교 아태지역연구센터, 2020 참조; 중국의 현황에 관하여는 Elsa B. Kania, *Chinese Military Innovation in Artificial Intelligence*, Center for a New American Security, 2019 참조.

12) 대표적으로 Peter Warren Singer, *Wired for War: The Robotics Revolution and Conflict in the 21st century*, Penguin Books, 2009.

13) Philip Alston, *Interim Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions*, UN Doc. A/65/321, 2010; Christof Heyns, *Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions*, UN Doc. A/HRC/23/47, 2013.

14) 이들의 구체적 주장에 관하여는 박문인, “자율무기체계의 국제법적 허용성과 규제방안”, 서울대학교대학원 박사학위 논문, 2019, 68-85쪽 참조.

15) 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(CCW). 2001. 12. 10. 발효되었다.

16) Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons System(GGE).

시하는 성과를 낳기도 했지만, 그 후의 논의는 대체로 지지부진하다. 현재로서는 자율살상무기체계의 통제를 위한 구체적 방안을 놓고 법적 구속력 있는 문서를 마련하지는 제3세계 국가들의 입장과 정치적 합의 및 모범 관행의 형성으로 충분하다는 군사선진국들의 입장이 대립하고 있는 상황이다.¹⁷⁾ 참고로, 2023년의 회의에서는 자율살상무기체계에 국제인도법(International Humanitarian Law: IHL)이 적용된다는 점 및 국제인도법을 준수할 능력이 없는 자율살상무기체계는 금지되어야 한다는 점이 선언되었다.¹⁸⁾

2.2 반대론과 찬성론

학자들과 정부 및 비정부기구들은 나름의 논리를 들어 자율무기체계의 허용 여부를 두고 논쟁을 해 왔다. 아래에서 찬반 양론의 논거들을 살펴보도록 한다.

2.2.1 반대론의 논거

자율무기체계의 활용에 반대하는 주장의 논거로는 i) 전쟁의 문턱을 낮출 것이라는 우려, ii) 국제인도법의 준수 불가능, iii) 책임의 불명확성 내지 흠결, iv) 인간의 존엄성에 대한 위협 등이 제시된다.

자율무기체계가 아군의 희생을 감소시켜 전쟁이 보다 쉽게 발발할 수 있다는 우려는 앞서 본 알스톤과 헤인즈의 보고서 이래 수많은 학자들에 의하여 제기되어 왔다.¹⁹⁾ 2015년에는 스티븐 호킹(S. Hawking)이나 촘스키(N. Chomsky)와 같은 명사들을 포함한 1,000여 명의 전문가들이 이같은 우려를 제기하며 자율무기의 금지를 촉구하는

17) GGE, *Report of the 2022 session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems*, GGE, 2022; 한편 유럽의회는 법적 구속력 있는 문서를 통해 자율살상무기를 규제할 수 있도록 국제협상 개시를 위해 노력할 것을 촉구한 바 있다[European Parliament, *European Parliament Resolution of 12 September 2018 on Autonomous Weapon Systems*(2018/2752(RSP)), European Parliament, 2018, paras L. 3-4].

18) GGE, *Report of the 2023 session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems*, GGE, 2023, para 21.

19) Peter Asaro, "How Just Could a Robot War Be?", *Current Issues In Computing and Philosophy* Vol. 175, IOS Press, 2008, p. 56; Paul W. Kahn, "The Paradox of Riskless Warfare", *Philosophy & Public Policy Quarterly* Vol. 22 No. 3, George Mason University, 2002, pp. 2-8; Yuna Huh Wong, et al., *Deterrence in the Age of Thinking Machines*, Rand Corporation, 2020(유인시스템이 무인시스템보다 더 억제력을 가질 수 있다는 결론을 제시하였다).

결의안에 서명하기도 했다.²⁰⁾ 엘론 머스크(E. Musk)는 “AI 무기가 핵무기보다 싸고 대량생산이 가능”하다고 함으로써 이를 단적으로 표현했다.²¹⁾ 특히 헤인즈는 UN 보고서에서 개별 무력분쟁에서의 사상자는 줄어든더라도 증가된 무력분쟁으로 인해 전체적인 사상자는 늘어날 것이라고 주장하였다.

반대론의 두 번째 논거는 자율무기체계가 국제인도법을 준수할 능력이 없다는 점이다.²²⁾ 국제인도법은 전투원과 민간인을 구별할 것과 부수적 피해가 군사적 이익에 비례하도록 할 것을 요구하는데 자율무기체계는 이러한 능력이 없다는 것이다. 휴먼 라이츠워치는 이를 이유로 자율살상무기의 선제적 금지를 요구하였고,²³⁾ 국제여성평화·자유연맹은 “자율살상무기를 프로그래밍하는 자체가 이성과 법에 대한 근본적인 도전”이라고 주장하기도 했다.²⁴⁾

세 번째 논거는 자율무기체계의 사용으로 인한 책임을 부담할 자가 불명확하거나 흠결될 수 있다는 점이다. 즉, 자율무기체계의 속성상 이를 활용한 작전 중 국제인도법 위반이 발생한 경우 누구에게 책임을 추궁할 것인지 명확하지 않을 수 있는데,²⁵⁾ 이는 국제인도법을 유명무실하게 만들 수 있다는 것이다. 나아가 일부 학자들은 자율무기체계를 사용하여 범죄의 결과가 발생했더라도 이는 자율무기체계가 스스로 판단해서 결정한 것이어서 책임의 흠결이 발생한다고 주장한다.²⁶⁾

반대론의 네 번째 논거는 자율무기체계가 인간의 존엄성을 침해한다는 것이다.

20) International Joint Conference on Artificial Intelligence, “IJCAI-15”, <<https://ijcai-15.org/>>, 검색일: 2023. 10. 11.

21) 포브스 코리아, “북핵보다 치명적인 인공지능 무기”, 2018. 2. 17, <<https://jmagazine.joins.com/forbes/view/319882>>, 검색일: 2023. 10. 11.

22) Dan Saxon, *Fighting Machines: Autonomous Weapons and Human Dignity*, University of Pennsylvania Press, 2022, p. 5; Filippo Santoni de Sio/Jeroen van den Hoven, “Meaningful Human Control over Autonomous Systems: A Philosophical Account”, *Frontiers in Robotics and AI Vol. 5 Article 15*, Frontiers Media S.A., 2018, p. 2.

23) Human Rights Watch, “Killer Robots”, <<http://www.hrw.org/topic/arms/killer-robots>>, 검색일: 2023. 10. 11.

24) WILPF, *Statement to the CCW Meeting of Experts on Lethal Autonomous Weapon Systems*, Reaching Critical Will, 2014.

25) Filippo Santoni de Sio/Jeroen van den Hoven, *op. cit.*, pp. 2-3; Philip Alston, *op. cit.*, para. 34.

26) Peter Asaro, “On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanization of Lethal Decision-making”, *International Review of the Red Cross Vol. 94 Iss. 886*, Cambridge University Press, 2012, p. 693; Robert Sparrow, “Killer Robots”, *Journal of Applied Philosophy Vol. 24 No. 1*, Wiley, 2007, p. 67; Noel E. Sharkey, “The Evitability of Autonomous Robot Warfare”, *International Review of the Red Cross Vol. 94 No. 886*, Cambridge University Press, 2012, p. 790; Dan Saxon, *op. cit.*, p. 5.

비록 전쟁에서 적의 생명을 빼앗는 행위가 적법행위로 인정된다고는 해도²⁷⁾ 인간의 존엄성까지 침해하는 것은 허용되지 않는다. 인간의 존엄성은 생명권의 기초가 되지만, 생명권보다 더욱 본질적인 가치이다.²⁸⁾ 독일 헌법은 제1조에서 ‘인간의 존엄성은 불가침으로서 어떤 경우에는 박탈당할 수 있는 생명권보다 더 소중한 인간의 본질적인 가치’라고 선언함으로써 이를 분명히 하고 있다.²⁹⁾ 이러한 법적 태도는 인간을 목적이 아닌 수단으로 취급해서는 안 된다는 칸트의 정언명령에 바탕을 두고 있다.³⁰⁾ 인간을 공격대상으로 삼는 자율무기체계가 인간 수준의 도덕적 존중의 역량을 갖추지 못한다면, 그러한 자율무기체계의 활용은 인간을 대상으로 취급함으로써 인간의 존엄성을 침해하는 것이 될 수 있다.³¹⁾ 이러한 문제는 본질적인 것으로서 기술의 실패가 아니라 정언명령적 실패라고 할 수 있다.³²⁾ 즉, 자율살상무기는 적에 대한 인간적인 존중의 원칙을 위반하는 것으로서 그 자체가 본질적 범죄(*mala in se*)라는 것이다.³³⁾

2.2.2 찬성론의 논거

찬성론은 반대론의 논거들이 부적절하다고 반박한다. 우선 자율무기체계의 채택이 전쟁의 문턱을 낮출 것이라는 우려는 과도하다고 한다. 전쟁은 단순한 기술의 문제가 아니고 정치, 경제, 사회적인 문제들이 복잡하게 얽혀 있으므로 자율무기체계가 전쟁 자체를 쉽게 유발하는 것은 아니라는 것이다.³⁴⁾ 베트남 전쟁이나 이라크 전쟁에서

27) 임예준, “인공지능 시대의 전쟁자동화와 인권에 관한 소고-국제법상 자율살상무기의 규제를 중심으로-”, 고려법학 제92호, 고려대학교 법학연구원, 2019, 288쪽.

28) 한희원, 앞의 논문, 342쪽.

29) 독일 헌법재판소는 테러리스트에 의해 납치된 것으로 의심되는 상업용 항공기를 격추시키는 것을 허용한 독일항공운송보안법이 인간을 항공기의 일부와 같이 대상으로 취급하고 숫자 또는 통계로 취급함으로써 독일 헌법 제1조를 위반하였다고 판시한 바 있다[BVerfG, Judgment of the First Senate of 15 February 2006 - 1 BvR 357/05 - paras. (1-156)].

30) 한희원, 정의로의 산책 제2판, 삼영사, 2011, 214-216쪽.

31) Peter Asaro, “On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanization of Lethal Decision-making”, p. 699; 김자희/장신/주성구, “자율 로봇의 잠재적 무기화에 대한 소고-개념 정립을 통한 규제를 중심으로”, 입법과 정책 제9권 제3호, 국회입법조사처, 2017, 147-148쪽.

32) 한희원, 앞의 논문, 344쪽.

33) Robert Sparrow, “Robots and Respect: Assessing the Case Against Autonomous Weapon Systems”, *Ethics & International Affairs Vol. 30 No. 1*, Cambridge University Press, 2016, p. 110.

34) Kenneth Anderson/Matthew Waxman, “Law and Ethics for Robot Soldiers”, *Policy Review Dec. 2012 & Jan. 2013 Iss. 176*, Hoover Institution Press, 2013, p. 43; Marco Sassòli, *Potential Advantages for the Respect of*

미국이 상대보다 우월한 군사 기술을 지녔음에도 불구하고 결정적으로 승리하지는 못한 것이 좋은 예이다. 설령 자율무기체제로 인해 전쟁의 결정이 쉬워졌다고 하더라도 이는 다른 최첨단 장거리 무기의 경우에도 마찬가지라고 할 것이다.³⁵⁾

찬성론은 국제인도법 준수 가능성과 관련된 우려에 대하여도 반론을 제기한다. 국제인도법상 자율무기체제를 직접적으로 금지하는 규정이 없고 자율무기체제의 정확성을 통해 오히려 국제인도법의 준수가 가능해진다는 것이다.³⁶⁾

책임의 불명확 내지 흠결과 관련된 주장에 대하여도 반론이 존재한다. 이에 따르면 자율무기체제의 경우 누가 공격 명령을 했는지 더 쉽게 기록할 수 있기 때문에 책임자를 더 명확하게 가려낼 수 있다.³⁷⁾ 또한 자율성에도 불구하고 대부분의 경우 법률 규정에 따라 지휘관이나 관련자가 법적 책임을 지게 될 뿐만 아니라, 설령 법적 책임이 성립되지 않더라도 그것이 무기의 적법성을 평가하는 요소는 아니다.³⁸⁾ 법적 책임은 국제인도법의 원칙을 구현하는 수단에 불과하다는 것이다.³⁹⁾

이들은 인권 및 생명권의 침해 가능성에 대한 우려도 과도하다고 주장한다. 오히려 자율무기체제의 채택으로 인해 무력충돌로 인한 인명과 재산의 손실을 줄일 수 있으므로, 보다 인도주의적이고 윤리적이라는 것이다.⁴⁰⁾ 바로 이러한 점이 자율무기체제 개발의 원동력이 되고 있기도 하다.

International Humanitarian Law, Professionals in Humanitarian Assistance and Protection(PHAP), 2013, p. 1; Christopher P. Toscano, "Friend of Humans: An Argument for Developing Autonomous Weapon Systems", *Journal of National Security Law & Policy* Vol. 8 No. 1, Georgetown University Law Center, 2015, p. 221.

35) Toscano, *ibid.*, p. 223.

36) Michael Schmitt/Jeffrey Thurnher, "Out of the Loop: Autonomous Weapon Systems and the Law of Armed Conflict", *Harvard National Security Journal* Vol. 4, Harvard Law School Journals, 2013, pp. 243-265; Marco Sassóli, "Autonomous Weapons and International Humanitarian Law: Advantages, Open Technical Questions and Legal Issues to Be Clarified", *International Law Studies* Vol. 90, US Naval War College Press, 2014, pp. 308-310.

37) Thomas W. Simpson/Vincent C. Müller, "Just war and robots' killings", *The Philosophical Quarterly* Vol. 66. No. 263, Oxford University Press, 2016, p. 302.

38) Charles J. Dunlap Jr., "Accountability and Autonomous Weapons: Much Ado about Nothing?", *Temple International & Comparative Law Journal* No. 30, Temple University School of Law, 2016, p. 66

39) Anderson/Waxman, *op. cit.*, p. 17.

40) Martin Lark, "The future of Killing: Ethical and Legal Implications of Fully Autonomous Weapon Systems", *Salus Journal* Vol. 5 No. 1, Charles Sturt University, 2017, pp. 69-70.

3. 자율무기체계에 대한 통제

3.1 통제의 필요성

앞서 자율무기체계에 관한 반대론과 찬성론의 논거들을 살펴보았지만, 모든 주장들이 찬반 양론으로 명확히 구분되는 것은 아니다. 예컨대 국제적십자위원회(ICRC)는 자율무기체계의 개발 자체에 반대하지는 않으면서도 이에 대한 법적, 윤리적 통제를 주장한다.⁴¹⁾ 각각의 주장들이 전제하는 자율무기체계의 개념도 동일하지 않다. 예컨대 책임의 흠결이나 인간의 존엄성 침해는 인간의 통제가 전혀 불가능한 이른바 완전자율무기체계를 전제로 하는 경우에 문제될 뿐이다.⁴²⁾

자율무기체계의 활용을 통해 인명 손실과 재산 피해를 줄일 수 있고, 이 때문에 세계 각국이 자율무기체계의 개발에 나서고 있다는 현실은 무시할 수 없다. 역사적으로 항공기 공습이나 잠수함 공격이 무차별적 성격으로 인해 금지되어야 한다는 주장이 있었지만 현실적 필요성으로 인해 관철되지 못한 것을 상기할 필요가 있다.⁴³⁾ 그러나 아무리 첨단 기술이 적용된 자율무기체계라도 오작동하거나 해킹될 위험은 상존하며, 상황에 따라 국제인도법상의 원칙을 준수하기 어려운 경우도 있을 수 있다.

그렇다면 단순히 자율무기체계의 허용 여부에 대한 찬성 또는 반대의 주장을 펼치기 보다는 자율무기체계에 대한 실효적이고 적절한 통제 방안을 마련하는 것이 보다 건설적인 접근방법일 것이다.⁴⁴⁾ 이와 관련하여 ‘의미 있는 인간 통제’(Meaningful Human Control)에 관하여 좀 더 깊이 살펴볼 필요가 있다.

41) ICRC, *Ethics and autonomous weapon systems: An Ethical for human control?*, CCW/GGE. 1/2018/WP. 5, 2018, paras. 21-26.

42) 유준구, “자율살상무기체계의 논의 동향과 쟁점”, 정책연구시리즈 2019-18, 국립외교원 외교안보연구소, 2019, 4쪽 참조.

43) Paul Scharre, *Army of None: Autonomous Weapons and the Future of War*, W. W. Norton & Company, 2018, p. 348.

44) 류병운, “드론과 로봇 등 자율무기의 국제법적 적법성”, *홍익법학* 제17권 제2호, 홍익대학교 법학연구소, 2016, 75쪽.

3.2 의미 있는 인간 통제

2013년 영국의 민간단체인 ‘Article 36’에 의하여 사용되기 시작한 ‘의미 있는 인간 통제’라는 용어는 오늘날 자율무기체계 규제를 위한 핵심적 개념 중 하나가 되었다.⁴⁵⁾ 의미 있는 인간 통제의 개념은 앞서 자율무기체계에 대한 찬반 양론을 검토하면서 보았던 윤리적, 법적 문제를 해소하는 기능을 하며,⁴⁶⁾ 이 때문에 일부 학자들은 이것이 자율무기체계 개념의 필수 요소가 되어야 한다고 주장하기도 한다.⁴⁷⁾

의미 있는 인간 통제가 구체적으로 무엇을 내용으로 하는지에 관하여는 다양한 견해들이 제시되어 왔지만,⁴⁸⁾ 아직 일반적으로 합의된 것은 없고 여전히 논의가 진행 중이다. 중요한 것은 인간의 통제는 단순한 반응이 아니라 ‘의미 있는’것이어야 한다는 점이다. 다만 그 구체적 내용은 맥락에 따라 유연하게 이해될 필요가 있다. 예컨대 유엔군축연구기구(UNIDIR)는 인간이 개입할 수 있는 충분한 시간적 여유, 판단의 행사, 공격의 종료나 철회 등의 다양한 방식으로 이해될 수 있다고 밝힌 바 있다.⁴⁹⁾

한편 의미 있는 인간 통제를 인간과 기계의 상호작용 방식과 관련지어 설명하는 견해⁵⁰⁾가 있다. 이에 따르면 의미 있는 인간 통제는 i) 무기체계의 작동에 있어 인간의 승인이 있어야 인정된다는 입장(human in the loop), ii) 필요한 경우 인간이 거부하거나 중지시킬 수 있으면 인정된다는 입장(human on the loop), iii) 무기체계의 작동이 인간에 의해 사전에 프로그램된 경우에도 인정될 수 있다는 입장(human out of the loop)⁵¹⁾이 있을 수 있다. 참고로 미국은 무기체계에 대한 통제는 그 자체가 목적이 아니라 국제인도법상의 원칙을 이행하는 유용한 수단에 불과하다는 입장에서 ‘인간

45) International Committee for Robot Arms Control, “Killer Robots: UK Government Policy on Fully Autonomous Weapons”, Apr. 2013, <http://www.article36.org/wp-content/uploads/2013/04/Policy_Paper1.pdf>, 검색일: 2023. 10. 11.

46) Netherlands, *Examination of Various Dimensions of Emerging Technologies in the Area of Lethal Autonomous Weapons Systems, in the Context of the Objectives and Purposes of the Convention*, CCW/GGE.1/2017/WP.2, 2017, paras. 15, 17.

47) HuffPost, Angela Kane, “Killer Robots and the Rule of Law”, <https://www.huffpost.com/entry/killer-robots-and-the-rul_b_3599657>, 검색일: 2023. 10. 11.

48) 박문언, “자율무기체계의 국제법적 허용성과 규제방안”, 101-107쪽 참조.

49) UNIDIR, *The Weaponization of Increasingly Autonomous Technologies: Considering How Meaningful Human Control Might Move the Discussion Forward*, UNIDIR, 2014, p. 3.

50) 박문언, “자율무기체계의 국제법적 허용성과 규제방안”, 102-103쪽.

51) 이를 긍정하는 견해로는 Tim McFarland, *Autonomous Weapons and Human Control*, Humanitarian Law and Policy, 2018. 한편 박문언, 위의 논문, 122쪽은 이를 부정한다.

통제’라는 용어의 사용에 부정적인 입장이며,⁵²⁾ 대신에 ‘적절한 수준의 인간 판단 (appropriate levels of human judgement)’ 개념을 채용하고 있다.⁵³⁾ 구체적인 의미는 불명확하지만, 인간의 통제를 배제하는 자율적 기능의 사용이 인간의 의도를 더욱 효과적으로 실행하고 사고를 회피할 수도 있다고 밝힌 것으로 보아⁵⁴⁾ iii)의 입장을 배척한 것으로 단정하기는 어렵다.⁵⁵⁾

자율무기체계의 통제에 있어서 ‘의미 있는 인간 통제’ 개념이 핵심적 역할을 하는 것은 맞지만, 그것이 얼마나 효과적으로 실현될 것인지는 결국 기술의 발전에 달려 있다. 특히 인공지능망을 활용하는 AI 기술의 경우 데이터를 통한 학습이 이루어지며, 대규모 언어모델의 경우 창발적 능력(emergent ability)을 보이기도 하기 때문에 기술적 통제가 쉽지 않을 수 있음을 유의해야 한다.⁵⁶⁾

III. 자율무기 개념의 구체화

1. 개념 정의의 중요성

자율무기체계에 관한 논의의 상당 부분은 그 개념에 관한 것이지만 안타깝게도 아직 국제적으로 합의된 정의는 존재하지 않는다. 이 때문에 일각에서는 개념 정의보다는 의미 있는 인간통제에 관한 논의에 집중하자고 주장하기도 한다.⁵⁷⁾ 하지만 자율

52) United States, *Human-Machine Interaction in the Development, Deployment and Use of Emerging Technologies in the Area of Lethal Autonomous Weapons Systems*, CCW/GGE. 2/2018/WP. 4, 2018, paras. 43, 46, 49.

53) Department of Defense, *Directive 3000.09: Autonomy in Weapon Systems*, 2023, §4(1).

54) United States, *Human-Machine Interaction in the Development, Deployment and Use of Emerging Technologies in the Area of Lethal Autonomous Weapons Systems*, CCW/GGE. 2/2018/WP. 4, 2018, paras. 11.

55) 이에 관하여는 미국 내에서도 혼선이 있는 것으로 보인다. 2021년 4월 미 육군 미래사령부 초대 사령관 존 머레이(John Murray) 대장은 “현재의 국방정책 하에서는 대인 살상과 관련해서 반드시 인간이 의사결정에 포함되어야 한다”고 밝혔지만, 국제전략연구소 AI 거버넌스 프로젝트의 책임자인 그레고리 앨런(Gregory C. Allen)은 그러한 주장이 DoDD 3000.09의 내용과 일치하지 않는다고 반박한 바 있다. CSIS, Gregory C. Allen, 2022. 6. 6, “DoD Is Updating Its Decade-Old Autonomous Weapons Policy, but Confusion Remains Widespread”, <<https://www.csis.org/analysis/dod-updating-its-decade-old-autonomous-weapons-policy-confusion-remains-widespread>>, 검색일: 2023. 10. 11.

56) 강화학습 기술을 활용한 AI 자율무기체계를 상정하여 통제 전략의 실효성과 한계를 검토한 연구로는 마정목, “강화학습 Q러닝 기반의 자율무기체계 통제 전략에 관한 고찰”, 한국군사학논집 제78권 제2호, 육군사관학교 화랑대연구소, 2022 참조.

무기체계의 개념은 여전히 중요한 문제이다. 이를 어떻게 정의하는지에 따라 개발, 사용 및 통제에 관한 논의 내용이 달라지기 때문이다. 예컨대 자율무기체계의 개념을 지나치게 좁게 볼 경우 그에 해당하지 않는 무기체계의 범위가 늘어나기 때문에 효과적인 통제가 어렵게 된다.⁵⁸⁾ 안타깝게도 우리나라는 자율무기체계의 개념에 대한 기본적 입장을 제시하지 못하고 있는 상황이다. 아래에서는 각국의 개념 정의를 살펴보고 개념 요소로서 고려되어야 할 사항을 검토해보기로 한다.

2. 각국의 개념 정의

2.1 미국

미국은 일찍이 2012년 국방부 훈령 3000.09를 통해 자율무기체계의 개념을 정립하였다. 이에 따르면 자율무기체계란 ‘일단 작동되면 운용자에 의한 추가적인 개입 없이도 표적을 선택하고 공격할 수 있는 무기체계’로 정의된다.⁵⁹⁾ 위 개념은 실제 자율무기체계의 개발 및 운용 과정에서 유용성을 검증받아 2023년 개정된 훈령에서도 그대로 유지되었다.

미 국방부 훈령 3000.09는 반자율무기체계와 운용자 감독 자율무기체계⁶⁰⁾도 정의하고 있다. 반자율무기체계란 ‘일단 작동되면 운용자에 의해 선택된 표적이나 표적군⁶¹⁾만을 교전하는 무기체계’를 뜻하고, 운용자 감독 자율무기체계란 ‘오작동 상황과 같이 감수할 수 없는 피해가 일어나기 전에 운용자가 개입하여 교전을 중지시킬 수 있도록 설계한 자율무기체계’를 뜻한다. 미 국방부 훈령은 자율무기체계와 반자율무기체계를 적용대상으로 하는데, 운용자 감독 자율무기체계는 자율무기체계의 범주

57) Netherlands, *Statement by the Netherlands 2018 Group of Governmental Experts on LAWS*, Characterization, 2018, p. 1.

58) ICRC, *Statement of the International Committee of the Red Cross, Convention on Certain Conventional Weapons Group of Governmental Experts on Lethal Autonomous Weapons Systems*, 2017.

59) Department of Defense, *Directive 3000.09: Autonomy in Weapon Systems*, 2012.

60) 2012년의 훈령에서는 ‘인간 감독’이라고 표현되어 있었다.

61) 표적군(target group)은 전투기 모델명과 같은 일반 분류를 포함하지 않는다고 함으로써 광범위한 표적의 선택은 의미 있는 통제로 볼 수 없음을 밝히고 있다(마정목, “미 국방부 무기체계 자율성 훈령 개정에 따른 자율무기체계 정책 분석과 이해”, 국방연구 제66권 제2호, 국방대학교 국가안전보장문제연구소, 2023, 116쪽).

아래에 속한 것으로 보고 있다. 참고로 자동화된 표적 선정 능력이 있는 유도 폭탄도 적용 대상이 되는 반면, 사이버 능력이나 지뢰 또는 무기체계가 아닌 자율/반자율 시스템 등은 적용대상에서 제외된다.

2.2 그 밖의 국가

중국은 인간에 대해 살상성을 가지고 임무수행의 전 과정에 걸쳐 자율성이 있으며, 작동된 이상 인간에 의한 종료가 불가능하고, 자율적으로 학습함으로써 발전성을 가지는 무기체계만을 자율살상무기체계로 본다.⁶²⁾ 이른바 ‘완전자율살상무기체계’만을 자율무기체계로 보아 규율하려는 입장이다.

러시아는 인간 운영자의 개입 없이 전투나 지원 임무를 수행하기 위한 완전 자율의 무인 기술수단으로서 포탄(ordnance) 이외의 것을 자율무기체계로 본다.⁶³⁾ 특히 무인 항공기(UAV)를 자율무기체계와 관련지어 논의하는 것에 부정적이다.

영국은 단순히 사전에 설정된 논리적 규칙을 따르는 자동무기와 달리 높은 수준의 의도와 목표를 이해할 수 있는 시스템으로서 인간의 감독이나 통제 없이 환경을 이해하거나 인지하여 여러 대안들 중에서 행동해야 할 방향을 결정할 수 있는 것을 자율무기로 보고 있다.⁶⁴⁾

프랑스는 지휘부의 지휘통제나 통신이 완전히 차단되어 인간 감독이 부재하는 무기체계를 자율살상무기체계라고 본다.⁶⁵⁾ 특히 자기학습능력을 중시하는데, 사전에 입력된 프로그램에 따라 인간을 살상하는 무기체계는 불확실성의 요소가 없어서 자율살상무기체계가 아니라고 본다.

한편 네덜란드는 발사된 이후에 사전에 설정된 표적의 범주와 일치하는 목표를 인간의 개입 없이 선택하고 교전하는 무기체계를 자율무기체계라고 본다.⁶⁶⁾ 특히 네덜란드는 자율무기체계의 요소로 의미 있는 인간 통제를 명시적으로 언급하면서,

62) China, *Position Paper*, CCW/GGE. 1/2018/WP. 7, para. 3.

63) Russian Federation, *Russia's Approaches to the Elaboration of a Working Definition and Basic Functions of Lethal Autonomous Weapons Systems in the Context of the Purposes and Objectives of the Convention*, CCW/GGE. 1/2018/WP. 6, paras. 2, 10.

64) UK Ministry of Defence, *Joint Doctrine Publication 0-01. 1: UK Supplement to the NATO Terminology Database*, 2018, A-2.

65) France, *Characterization of a LAWS, Meeting of Experts on Lethal Autonomous Weapons Systems*, 2016.

66) Netherlands, *op. cit.*, para. 5.

이것이 부재하는 완전자율무기체계의 개발이나 배치를 거부한다는 입장을 밝히고 있다.⁶⁷⁾

3. 자율무기체계의 개념 요소

3.1 자율성

3.1.1 전체적 접근법과 기능적 접근법

자율무기체계의 핵심적 요소는 자율성(autonomy)이다.⁶⁸⁾ 그러나 자율성의 구체적 의미는 접근방법에 따라 달라진다. 가장 큰 차이를 가져오는 것은 자율성을 무기체계 자체의 자율성으로 이해할 것인지 아니면 시스템에 있어서 특정 기능의 자율성으로 이해할 것인지이다.

초창기의 논의는 무기체계 자체의 자율성에 초점이 맞춰져 있었다. 이러한 관점은 무기체계를 수단이나 대상이 아닌 주체(agent)로 바라보고, 나아가 무기체계에 모종의 법적 지위를 결부시키려는 아이디어로 이어지기 쉽다. 자율무기체계를 활용할 경우 책임의 흠결이 발생한다는 주장이나, 지휘관 책임과 관련하여 자율무기체계를 ‘하급자’로 볼 수 있는지를 검토하려는 견해⁶⁹⁾는 이러한 입장을 전제로 한다. 이는 인공 에이전트에 법인격을 인정할 것인지와 관련된 논의를 떠오르게 한다.⁷⁰⁾

그러나 무기체계 자체의 자율성에만 주목할 경우 자율무기체계가 초래하는 구체적 위험을 식별하고 이를 통제해야 한다는 현실적 요청에 대응하기 어렵고 불필요한 오해만 불러일으킬 수 있다. 자율무기체계에서 문제되는 자율성이란 인간의 고유한 가치에 기초하여 인간 행위의 자율성을 검토하는 법철학에서의 논의와는 구별되어야 한다.⁷¹⁾ 헤인즈 역시 자율무기체계에서의 ‘자율’이라는 용어가 인간의 ‘자유 의지’와

67) Netherlands, *Ibid.*, para. 8-11.

68) 박문언, “자율무기체계의 발전과 우리 군의 현실에 대한 고찰”, *Law & Technology* 제12권 제6호, 서울대학교 기술과법센터, 2016, 52-54쪽.

69) Ronald Craig Arkin/Patrick Ulam/Alan R. Wagner, “Moral Decision-Making in Autonomous Systems: Enforcement, Moral Emotions, Dignity, Trust and Deception”, *IEEE Vol. 100 Iss. 3, Proceedings of the IEEE*, 2011, pp. 571-589.

70) 이상용, “인공지능과 법인격”, *민사법학* 제89호, 한국민사법학회, 2019 참조.

71) 마정목, “통제 가능한 자율무기체계의 개념과 설계에 관한 연구”, *국방정책연구* 제36권 제2호, 한국국방연구원, 2020, 90-92쪽; 김보연, “인공지능을 통한 전쟁수행은 정당인가-자율무기체계를 통한 국제인도

동일시될 수 없다고 보았다.⁷²⁾

자율성은 무기체계 자체의 자율성이 아니라 시스템에 있어서의 ‘자율적 기능’으로 이해되어야 한다.⁷³⁾ 국제적십자위원회는 이러한 입장에서 자율무기체계를 ‘핵심적 기능에서 자율성을 갖는 무기’라고 정의하였다.⁷⁴⁾ 미 국방부 훈령 3000.09 역시 ‘무기 체계의 자율/반자율 기능’을 개발하고 사용하기 위한 책임 할당과 정책을 수립하는 것을 그 목적으로 한다고 함으로써 이러한 입장을 명시적으로 채택하였다. 문제는 어떠한 기능에 자율성이 인정될 경우에 자율무기체제로 분류될 수 있는지이다. 이에 관하여는 국제적 합의가 필요하겠지만, 대체로 표적의 선택 및 교전이 이에 해당할 것이다.⁷⁵⁾

3.1.2 속성 기반 접근법과 관계 기반 접근법

자율성의 의미는 이를 무기체계 내지 그 기능이 지닌 속성에 기반하여 이해할 것인지 아니면 인간 운용자와의 관계에 기반하여 이해할 것인지에 따라 달라지기도 한다.⁷⁶⁾ 중국이나 프랑스와 같이 자율무기체계의 속성으로 자기학습능력을 요구하거나 영국과 같이 높은 수준의 인지능력을 요구하는 입장이 전자에 해당한다.⁷⁷⁾ 그러나 자율성은 시스템 자체의 성질이라기보다는 다른 특징과 관련하여 규정될 수 있는 관계적 특성이라고 보아야 한다.⁷⁸⁾ 다시 말해 인간과의 상호협력 관계를 중심으로 기능적 자율성이 어느 정도인지 검토하는 것이 좀더 타당한 접근방법이다.⁷⁹⁾

법 준수가능성을 중심으로-”, 고려법학 제107호, 고려대학교 법학연구원, 2022, 419쪽.

72) Heyns, *op. cit.*, para 43, pp. 7-8.

73) Multinational Capability Development Campaign, *Role of Autonomous Systems in Gaining Operational Access*, MCDC, 2014, p. 16.

74) International Committee of the Red Cross, *Statement at the Convention on Certain Conventional Weapons (CCW), Meeting of Experts on Lethal Autonomous Weapons Systems(LAWS)*, 2015.

75) Heyns, *op. cit.*, pp. 7-8; 박문언, “자율무기체계의 국제법적 허용성과 규제방안”, 43-46쪽은 교전에 있어서의 자율성을 자율무기체제로 인정되기 위한 최소한의 요소로 본다.

76) 속성 기반 접근법과 관계 기반 접근법의 구분은 인공 에이전트의 법인격 인정 여부에 관한 논의에서 차용한 것이다. 이상용, 앞의 논문, 7쪽 참조.

77) NATO의 조인자 그룹은 원격조정, 자동화, 학습능력이 없는 자율화, 학습능력이 있는 자율화 단계로 구분하기도 했다. Andrew P. William, et al., *Autonomous Systems: Issues for Defence Policymakers*, NATO Headquarters Supreme Allied Commander Transformation, 2015, pp. 40-52.

78) Multinational Capability Development Campaign, *op. cit.*, p. 16.

79) Defense Science Board, *Task Force Report: The Role of Autonomy in DoD Systems*, Defense Science Board, 2012, p. 4.

이러한 접근방법을 취한 대표적인 것이 바로 ‘OODA Loop’ 개념이다. 인간의 의사결정 과정은 관찰(observe), 판단(orient), 결정(decide), 행동(act)의 4단계로 나누어볼 수 있다.⁸⁰⁾ 이를 이용하면 무기체계의 자율성은 위 과정에서 인간 운용자와의 상호작용 유형에 따라 단계를 나누어볼 수 있다. 즉, 무기체계가 인간의 승인에 따라 작동하는 경우는 ‘human in the loop’, 무기체계가 독자적으로 임무를 수행하지만 그 과정에서 인간의 감독을 받는 경우는 ‘human on the loop’, 무기체계가 인간의 통제를 전혀 받지 않고 자체적으로 의사결정을 하는 경우는 ‘human out of loop’라고 할 수 있다.⁸¹⁾ 앞서 본 것처럼 이는 자율무기에 대한 통제의 관점에서 재해석될 수 있는데, 이에 따르면 각각 인간 통제, 인간 감독, 그리고 완전 자율에 해당한다고 볼 수도 있다.⁸²⁾

앞서 보았던 미 국방부 훈령 3000.09의 자율무기 개념 역시 인간과의 상호관계에 중점을 둔 것으로 이해된다. 특히 표적의 선택에 운용자가 관여하는 반자율무기는 ‘human in the loop’, 운용자에 의한 교전 중지가 가능한 운용자 감독 자율무기는 ‘human on the loop’에 해당한다고 볼 수 있다.

3.2 그 밖의 요소들

자율무기체계의 개념에 관한 각국의 입장을 살펴보면, 자율성 외에도 살상성이나 자기학습능력 등이 언급되고 있음을 알 수 있다. 그러나 이들을 자율무기체계의 개념 요소로 볼 필요는 없다. 인명 살상 없이 물자(material) 표적만을 대상으로 하는 자율적 무기체계나 자기학습능력 없는 자율적 무기체계도 규제 대상으로 삼을 필요가 있기 때문이다. 최근 들어 자율무기체계가 급속도로 발전하는데 기여하고 있는 AI 기술 역시 자율무기체계의 개념요소로 볼 것은 아니다. 기술 자체에 기반을 둔 개념은 그와 연동하여 변화할 수밖에 없다는 점에서 더욱 그러하다.⁸³⁾

4. 소결론

자율무기체계에 관한 논의는 그 활용 과정에서 생길 수 있는 위험을 효과적으로

80) Robert Coram, *Boyd: The Fighter Pilot Who Changed the Art of War*, Back bay Books, 2002, pp. 334-335, 344.

81) Singer, *op. cit.*, p. 74.

82) 박문언, “자율무기체계의 국제법적 허용성과 규제방안”, 89쪽.

83) 마정목, “미 국방부 무기체계 자율성 훈령 개정에 따른 자율무기체계 정책 분석과 이해”, 117쪽.

통제하기 위한 것이다. 이러한 목적을 위해서는 기능적, 관계적 관점에 따라 개념을 정의할 필요가 있다. 미 국방성이 채택한 ‘일단 작동되면 운용자에 의한 추가적인 개입 없이도 표적을 선택하고 공격할 수 있는 무기체계’라는 정의는 이러한 목적에 부합하는 것으로 여겨진다. 반자율무기체계나 운용자 감독 자율무기체계의 개념도 유용하게 쓰일 수 있다. 다만 미국의 MQ-1 Predator 또는 MQ-9 Reaper와 같이 원격으로 조종되는 무인항공기(UAV)는 표적 선정이나 교전이 인간에 의하여 이루어지기 때문에 자율무기체계에 해당하지 않는다. 한편, 자율무기체계 가운데 인명 살상을 목적으로 하는 것은 자율살상무기체계(Lethal Autonomous Weapon System: LAWS)⁸⁴⁾라고 부를 수 있을 것이다.

자율무기체계의 개념과 관련하여 가장 논쟁적인 주제는 완전자율무기체계(Fully Autonomous Weapon System: FAWS)를 어떻게 이해하고 다룰 것인가이다. 완전자율무기체계의 개념도 명확한 것은 아니지만 ‘의미 있는 인간 통제’가 이루어지지 않는 경우를 가리키는 것으로 이해하는 것이 문제의 식별 및 해결에 가장 유용하다.⁸⁵⁾ 완전자율무기체계의 취급에 관한 입장은 판이하다. 한편에서는 완전자율무기체계만을 자율무기체계에 해당되는 것으로 보아 다른 것들을 규율 대상에서 제외하려고 하는 반면(중국의 입장), 다른 편에서는 완전자율무기체계를 자율무기체계의 범주 밖에 있는 것으로 보고 그 밖의 것들을 중심으로 규율하려 한다(네덜란드의 입장). 다만 두 입장 모두 완전자율무기체계는 현존하지도 않고 허용되어서도 안 된다고 주장한다.⁸⁶⁾⁸⁷⁾ 생각건대 완전자율무기체계의 개념은 그다지 유용하지 않다고 여겨진다. 자율무기체계에 대한 통제를 회피하기 위하여 악용될 수 있고, 그 금지는 의미 있는 인간 통제를 관철함으로써 이를 수 있기 때문이다.

84) 치명적 자율무기라고도 번역되기도 한다. 한희원, 앞의 논문, 326쪽.

85) Netherlands, *op. cit.*, para. 10.

86) 간혹 기술의 성취에만 매몰되어 완전자율성을 추구하려는 견해가 보이기도 하는데(김종열, “자율 무기 체계의 자율성에 대한 연구”, 융합보안 논문지 제18권 제2호, 한국융합보안학회, 2018, 110쪽), 이는 부적절하다고 할 것이다.

87) 참고로 이스라엘의 IAI사는 자율적으로 적의 레이더를 식별하고 공격하는 하피(Harpy)라는 배회무기(loitering munition)를 완전 자율(fully autonomous)이라고 소개하고 있다.

IV. 윤리적 통제 방안

1. 윤리적 통제의 의의

1.1 윤리적 통제의 중요성

자율무기체계에 대한 구속력 있는 법적 통제, 특히 국제법적 통제는 국가 간의 이견으로 인해 이른 시일 안에 이루어지기 어려운 상황이다. 따라서 현재로서는 구속력 없는 유연한 방식에 의한 통제가 오히려 실효적일 수 있다. 국제사회에서 윤리원칙의 채택이나 정치적 선언 및 모범관행의 수립에 대한 목소리가 높아지는 것은 이 때문이다.

1.2 윤리이론과 자율무기체계

윤리이론은 크게 의무론적 윤리(deontological ethics)와 결과론적 윤리(consequential ethics), 그리고 덕의 윤리(value ethics)로 나누어볼 수 있다.⁸⁸⁾ 인간을 수단이 아닌 목적으로 대하라는 칸트(I. Kant)의 정언명령은 의무론적 윤리를 대변한다. 그에 비해 벤담(J. Bentham)이 주장한 ‘최대 다수의 최대 행복’은 결과론적 윤리의 핵심을 보여준다. 이는 인간을 이익형량의 대상으로 삼는 것을 허용함으로써 의무론적 윤리와 긴장 관계에 있게 된다. 한편 덕의 윤리는 의무론적 윤리와 결과론적 윤리가 근대 자유주의적 배경 아래에서 개별적 ‘행위’에 주목한 것과 달리 ‘행위자’의 품성에 주목한다. 이는 공동체 구성원의 미덕이나 인륜을 중시하는 입장으로도 연결될 수 있다.

이러한 윤리이론은 자율무기체계가 평가를 위한 논리를 제공하기도 한다.⁸⁹⁾ 의무론적 윤리는 인간이 아닌 기계가 사람을 살상하는 것은 인간의 존엄성을 부정하는 것이라는 논리로 이어질 수 있다. 반면에 결과론적 윤리는 자율무기체계를 이용하여

88) 이상용, “알고리즘 규제를 위한 지도(地圖)-원리, 구조, 내용-”, 경제규제와 법 제13권 제2호, 서울대학교 공익산업법센터, 2020, 131-132쪽 참조.

89) UNIDIR, *The Weaponization of Increasingly Autonomous Technologies: Considering Ethics and Social Values*, UNIDIR, 2014, pp. 2-5; 윤리이론은 AI 기술을 이용한 자율무기체계에 기계윤리를 구현하기 위한 방법론과도 관련이 있다. 하향식 학습방법을 이용하는 경우 행위윤리의 관점에서 도덕 규칙을 학습시키려고 하는 반면, 상향식 학습방법을 이용하는 경우 행위자의 윤리에 초점을 맞추게 된다[한국인공지능법학회, 인공지능과 법, 박영사, 2019, 71-81쪽(김효은 집필부분)].

전투원과 민간인을 보다 정확하게 구별하고 인명 살상과 재산 피해를 최소화하는 것은 바람직한 것이라는 논리로 이어질 수 있다. 그러나 윤리이론만으로 자율무기체계에 대한 구체적인 통제방안을 도출하기에는 한계가 있다. 자율무기체계의 개발 및 활용 과정에서 지켜져야 할 중요한 윤리적 원칙들을 식별하려는 실용윤리적 접근이 필요한 까닭은 이 때문이다.

2. 윤리원칙

2.1 일반적 AI 윤리원칙

2017년의 ‘아실로마 원칙’⁹⁰⁾ 이래 수많은 AI 윤리원칙이 발표되었다. 대표적으로 2019년 EU 집행위원회의 고위전문가 자문단이 발표한 ‘신뢰할 수 있는 인공지능을 위한 윤리지침’은 인간의 자율성 존중, 해악의 방지, 공정성, 설명가능성 등 4대 원칙을 밝히면서, 그 실현을 위해 안전성, 프라이버시 및 데이터 거버넌스, 비차별성, 다양성, 복지 및 환경, 투명성, 책임성 등을 강조하였다.⁹¹⁾ 같은 해 OECD 위원회는 인간 중심적 가치와 공정성, 포용적 성장과 지속 가능한 개발 및 복지, 견고성과 보안 및 안전, 투명성과 설명가능성, 책임성 등 5대 원칙을 제시하였다.⁹²⁾ 우리 정부 역시 2017년 ‘지능정보사회 윤리 가이드라인’을 통해 책무성, 통제성, 투명성, 공공성이라는 4대 원칙을 제시한 바 있다.⁹³⁾

윤리원칙들의 내용이 점차 수렴해가고 있다는 점은 주목할 필요가 있다.⁹⁴⁾ 이는 AI 기술이 지닌 특성과 활용 맥락에서 비롯된 것으로 볼 수 있다. 오늘날 AI 기술은 인공지능영역을 이용한 데이터 학습을 통해 자율적으로 작동하면서 다양한 영역에서

90) Future of Life Institute, “Asilomar AI Principles”, <<https://futureoflife.org/person/asilomar-ai-principles/>>, 검색일: 2023. 10. 11.

91) AI HLEG, *Ethics Guidelines for Trustworthy AI* High-Level Expert Group on AI, European Commission, 2019. 신뢰할 수 있는 인공지능을 합법적이고(lawful), 윤리적이며(ethical), 강건한(robust) 인공지능으로 정의하였다.

92) OECD Council, *Recommendation of the Council on Artificial Intelligence*, OECD, 2019.

93) 정보문화포럼/한국정보화진흥원, 지능정보사회 윤리 가이드라인 및 지능정보사회 윤리현장, 한국정보화진흥원, 2017.

94) 윤혜선, “인공지능 규제 정책에 관한 연구: 주요국의 규제 정책 사례를 중심으로”, 정보통신정책연구 제 26권 제4호, 정보통신정책학회, 2019 참조.

인간의 판단과 의사결정을 돕고 있다. 자율적으로 작동하기 때문에 인간 중심적 가치나 책임성이 요구되고, 논리적 알고리즘보다는 인공지능망을 통해 작동하기 때문에 투명성이나 설명가능성이 강조되며, 데이터 학습에 의존하기 때문에 프라이버시 및 데이터 거버넌스가 주목받을 수밖에 없다. 또한 판단과 의사결정에 활용되기 때문에 공정성, 다양성 등이 강조되며, 범용성을 지닌 신기술인 만큼 그 성능이나 위험성이 충분히 파악되지 않은 탓에 강건성, 안전성, 보안 등이 요구될 수밖에 없다.

자율무기체계와 같은 군사 영역에서도 AI 기술을 이용하는 경우가 많으므로 위와 같은 일반적 윤리원칙은 대체로 군사 분야, 나아가 군사 분야를 간접적으로 지원하는 국방분야에서 활용되는 AI⁹⁵⁾의 경우에도 유효하다고 할 수 있다. 그러나 군사 AI의 경우 영역의 특수성으로 인해 몇 가지 특별한 고려가 필요할 수 있다. 우선 군사 분야는 필연적으로 인명 살상이나 재산 피해가 뒤따를 수밖에 없고, 정당한 무력행사인 경우 그것이 적법하게 평가될 수도 있다. 따라서 위협의 예방이나 완화를 위한 특별한 주의가 필요하고, 생명·신체·재산상의 안전을 넘어 인간의 존엄성과 같은 보다 높은 차원의 윤리적 원칙이나 국제인도법의 준수가 강조될 필요가 있다. 그리고 적과의 대결이라는 군사 분야의 특성상 성능이나 보안 또는 강건성이 강조될 수밖에 없어 공정성, 다양성, 프라이버시 등의 요구는 다소 물려설 여지가 있다. 다만 군사 분야를 간접적으로 지원하는 국방 분야에서는 일반적 윤리원칙이 보다 유효하게 적용될 수 있을 것이다. 아래에서는 군사 분야의 특수성이 반영된 대표적인 윤리원칙들을 살펴보기로 한다.

2.2 CCW 자율무기체계 지도원칙

CCW는 수 년간의 논의 끝에 2019년 자율무기체계의 11가지 지도원칙을 발표하였다. 이는 CCW 정부전문가그룹의 가장 큰 성과로서 이후 자율무기체계와 관련된 윤리적 원칙의 발전에 크게 공헌하였다. 그 내용의 요지는 다음과 같다.⁹⁶⁾

95) 국방 AI는 주로 타겟 식별, 자동무기 발사체계, 시뮬레이션, 예측(및 추천), 사이버보안, 보급-의무 등의 영역에서 개발이 이루어지고 있다. 특히 군사 분야의 경우 인식, 판별, 결심 단계에서 AI 기술이 활용되고 있다(YuLong Zhang, et al., *Application of artificial intelligence in military: From projects view*, 2020 6th International Conference on Big Data and Information Analytics, IEEE, 2020); 그 밖에 이종관/한창희, 앞의 논문, 785-787쪽도 참조.

96) United Nations, *Meeting of the High Contracting Parties to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have*

① 자율살상무기체계를 포함한 모든 무기체계에 국제인도법이 적용될 것(국제인도법 적용), ② 책임은 기계에게 전가될 수 없으며 인간에게 책임이 유지될 것(인간 책임), ③ 인간과 기계의 상호작용은 전체 수명주기 동안 보장되어야 하며, 작전적 맥락과 무기체계의 특성 및 성능이 고려될 것(인간기계 상호작용), ④ 신무기의 개발·배치는 책임 있는 지휘계통을 통해 국제법에 의해 보장될 것(국제법 적용), ⑤ 신무기 또는 전쟁수단의 연구·개발·획득에 있어 국제법적 금지 여부를 검토할 것(무기의 적법 검토), ⑥ 물리적 방호, 사이버 안전조치, 테러집단에 의한 오용의 위험을 고려할 것(다양한 위험의 고려), ⑦ 자율살상무기체계의 설계·개발·평가·배치시 위험평가 및 완화조치를 시행할 것(시험평가), ⑧ 자율살상무기체계 관련 신기술 사용시 국제인도법과 다른 국제법적 의무를 준수할 것(국제법적 의무 준수), ⑨ 자율살상무기체계 관련 신기술을 인간과 같이 취급하지 않을 것(의인화 금지), ⑩ 지능형 자율기술의 평화적 사용이나 발전을 저해하지 않을 것(신기술의 평화적 발전), ⑪ CCW가 자율살상무기체계 문제의 논의를 위한 가장 적절한 국제기구임을 선언함(CCW의 역할).

2.3 미 국방부 AI 윤리원칙

미국은 2018년 국방 AI 전략을 발표하면서 전략적 접근방법의 하나로 군사 AI 윤리를 선도할 것을 제시하였다.⁹⁷⁾ 이에 따라 학계, 산업계, 시민단체 등 각계의 의견을 취합하는 신중한 절차를 거쳐 2020년에 전투원과 비전투원 모두에게 적용될 수 있는 군사 AI 윤리의 5가지 원칙이 마련되었다.⁹⁸⁾ 그 내용은 다음과 같다.

① 국방부 구성원은 적절한 수준의 주의와 판단력을 행사하는 한편 AI 능력의 개발, 배치, 사용에 대한 책임을 질 것(reliable: 책임성), ② AI 능력에 있어 의도하지 않은 편향을 최소화하기 위한 조치를 취할 것(equitable: 형평성), ③ 관계자들이 투명하고 감사 가능한 방법론, 데이터 출처, 설계 절차 및 문서화 등을 포함하여, AI 능력에 적용될 수 있는 기술, 개발 절차, 운용 방법을 적절하게 이해할 수 있도록 개발할

Indiscriminate Effect, CCW/MSP/2019/9, 2019, p. 10. Annex III. Guiding Principles affirmed by the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons System 참조.

97) Department of Defense, *Summary of the 2018 Department of Defense Artificial Intelligence Strategy*, Department of Defense, 2018.

98) Department of Defense, "Ethical Principles for Artificial Intelligence", <https://www.ai.mil/docs/Ethical_Principles_for_Artificial_Intelligence.pdf>, 검색일: 2023. 10. 11.

것(traceable: 추적가능성), ④ AI 능력의 용도는 명시적으로 잘 정의되어야 하며, AI 능력의 안전성, 보안성, 효과성은 정의된 용도 범위에서 전 생애주기에 걸쳐 시험되고 보장될 것(reliable: 신뢰성), ⑤ 의도된 기능을 다하도록 설계하되, 의도하지 않은 결과를 탐지하여 회피할 수 있고, 의도하지 않은 작동을 할 경우 중단 또는 정지할 수 있을 것(governable: 통제가능성).

3. 모범 관행의 수립과 정치적 선언

자율무기체계를 통제하기 위하여 윤리원칙을 활용하려는 것은 법적 구속력이 없어 동의를 이끌어내기 쉽고 유연한 규율이 가능하기 때문이다. 비록 엄밀한 의미에서는 윤리원칙과 같은 규범적 통제의 범주에 속하지는 않지만 모범 관행을 수립하는 것도 유연한 규율이라는 맥락에서 유사한 효과가 있을 수 있다. 영국 정부는 사이버 분야의 탈린 매뉴얼과 같은 매뉴얼을 마련하자고 제안하기도 했다.⁹⁹⁾

정치적 선언을 하는 방안도 현실적인 대안이 될 수 있다. 이와 관련하여 2023년 네덜란드와 한국이 공동으로 개최한 ‘인공지능의 책임 있는 군사적 이용에 관한 고위급회의(REAIM)’가 주목받고 있다. 미국과 중국 등 60여 국의 정부 관계자와 국제기구 종사자를 비롯하여 2000여 명이 참가한 위 회의에서 이스라엘을 제외한 참가국들은 ‘공동행동 촉구서’에 서명함으로써 국제법에 따른 의무와 국제 안보 및 안정, 책임성을 훼손하지 않는 방향으로 군사 AI를 개발하고 사용할 것을 선언하였다.¹⁰⁰⁾ 위 공동행동 촉구서에는 인간 책임, 위험의 평가 및 경감, 데이터 거버넌스, 안전조치 및 인간의 감독, 운전자 교육, 모범적 관행의 공유, 교류와 협력 등을 포함한 25개의 조항이 담겨 있다.¹⁰¹⁾ 특히 미 국방부는 위 공동행동 촉구서와는 별도로 ‘AI 및 자율성의 책임 있는 군사적 이용에 관한 정치적 선언문’을 마련하였다.¹⁰²⁾ 위 선언문은 여러 해에 걸쳐 다듬어진 미 국방부의

99) GGE, *Report of the 2022 session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems*, 2022.

100) 경향신문, “국제법 의무·책임성 훼손 없는 군용 AI 개발을”, 2023. 2. 17, <<https://www.khan.co.kr/world/world-general/article/202302172118025>>, 검색일: 2023. 10. 11.

101) Government of the Netherland, “REAIM 2023 Call to Action”, <<https://www.government.nl/documents/publications/2023/02/16/ream-2023-call-to-action>>, 검색일: 2023. 10. 11.

102) U.S. Department of State, “Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy”, <<https://www.state.gov/political-declaration-on-responsible-military-use-of-artificial-intelligence-and-autonomy/>>, 검색일: 2023. 10. 11.

AI 전략, AI 윤리원칙, 이행지침 등에 담긴 핵심적이고 실천적인 원칙들을 담고 있다. 다만 미국의 정치적 선언은 국제적 논의 없이 일방적으로 이루어졌다는 점에서 한계가 있다.¹⁰³⁾ 아마도 서울에서 개최하기로 예정된 2024년 REAIM에서 위 선언문에 바탕을 둔 합의문의 채택이 시도되지 않을까 예상해본다.

V. 법적 통제 방안

1. 법적 통제의 의의

1.1 법적 통제의 중요성

현행 국제법이나 우리나라의 국내법에는 자율무기체계를 직접적인 규율 대상으로 하는 규정이 존재하지 않는다. 현실적으로 윤리원칙이나 정치적 선언 등을 통한 규율이 자율무기체계의 통제에 어느 정도 역할을 하는 것은 사실이지만, 구속력 있는 법적 규율이 가지는 예측 가능성이나 강제력에 미칠 바가 아님은 명백하다. 본 장에서는 국제법과 국내법으로 나누어 자율무기체계에 대한 법적 통제 가능성을 검토하고자 한다. 다만 그 전에 법이론과 자율무기체계의 관계 및 자율무기체계의 법적 지위에 관한 논의를 간략히 살펴본다.

1.2 법이론과 자율무기체계

학자에 따라서는 법이론과 자율무기체계의 허용성 사이에 일정한 관련이 있다고 보기도 한다. 즉, “자율무기체계의 절대적 금지를 주장하는 견해는 법이 도덕적이어야 한다고 보는 자연법론에 가까운 반면, 자율무기체계의 허용가능성을 고려하는 견해는 법실증주의에 근접해 있다”는 것이다.¹⁰⁴⁾ 이는 자율무기체계의 법 준수능력을 놓고 벌어지는 찬반론과 궤를 같이 한다고 할 수 있다. 다만 이러한 법철학적 논쟁은 구체적인 대안을 제시해주지는 못한다는 한계가 있다.¹⁰⁵⁾

103) LawFare, Michael Depp, “Making Unilateral Norms for Military AI Multilateral”, 2023. 4. 6, <<https://www.lawfaremedia.org/article/making-unilateral-norms-military-ai-multilateral>>, 검색일: 2023. 10. 11.

104) 김보연, 앞의 논문, 417쪽.

1.3 자율무기체계의 법적 지위

자율무기체계 자체에 형사책임을 인정할 수 있을까. 이는 이론적으로 자율무기체계의 작동으로 인한 법익 침해나 그 위험의 발생을 형법상 ‘행위’로 볼 수 있는지, 그리고 자율무기체계가 형사책임의 ‘주체’가 될 수 있는지에 달려 있다.

일부 학자들은 AI의 작동이 형법상 행위로 평가될 수 있는지 여부는 행위론에 따라 달라진다고 본다. 즉, 인과적 행위론이나 목적적 행위론에 의할 경우 학습 및 판단능력을 갖춘 AI의 작동은 형법상의 행위로 인정될 수 있지만,¹⁰⁶⁾ 사회적 행위론이나 인격적 행위론에 의할 경우 AI가 사회적 의미를 안다거나 그 작동을 인격의 표출로 볼 수 없어 행위성을 인정하기 어렵다는 것이다.¹⁰⁷⁾

그러나 자율무기체계에 형사책임을 인정할 수 있는지의 문제는 행위성 여부에 앞서 형사책임 주체성, 구체적으로는 범죄능력과 수형능력의 인정 여부에 의하여 판단되어야 할 것이다. 특히 형사법 영역에서 자율무기체계의 책임 주체성을 판단함에 있어서는 민사법상 법인격 유무 판단의 경우와는 달리 ‘책임주의’가 관철될 필요가 있다.¹⁰⁸⁾ 즉, 자의식과 자기성찰능력이 없어 인간성이 결여된 자율무기체계는 인간의 존엄성 원리로부터 도출되는 비난가능성이 없고,¹⁰⁹⁾ 도덕적 자기결정 및 위법행위의 회피능력이 없으므로¹¹⁰⁾ 범죄능력을 인정할 수 없다. 또한 형벌로 인한 교육이나 예방효과를 기대하기 어려우므로 수형능력 역시 인정되지 않는다.¹¹¹⁾ 인간과 구별할 수 없는 수준에 이른 이른바 강인공지능의 경우 달리

105) 김보연, 위의 논문, 418쪽.

106) 안성조, “인공지능 로봇의 형사책임: 논의방향의 설정에 관한 몇 가지 발전적 제언”, 법철학연구 제20권 제2호, 한국법철학회, 2017, 84쪽.

107) Sabine Gless/Emily Silverman/Thomas Weigend, “If Robots Cause Harm, Who is to Blame? Self-driving Cars and Criminal Liability”, *New Criminal Law Review Vol. 19 No. 3.*, University of California Press, 2016, p. 6; 임석순, “형법상 인공지능의 귀속책임”, 형사정책연구 제27권 제4호, 한국형사법무정책연구원, 2016, 75쪽.

108) 이상용, “인공지능과 법인격”, 8쪽; 윤지영/김한균/감동근/김성돈, 법과학을 적용한 형사사법의 선진화방안 8: 인공지능기술, 한국형사정책연구원, 2017, 234쪽 이하(김성돈 집필부분); 국제법상 인공지능에 대한 법인격 인정 여부에 관한 비판적 검토로는 Simon Chesterman, “Artificial Intelligence and the Limits of Legal Personality”, *International and Comparative Law Quarterly Vol. 69 Iss. 4*, Cambridge University Press, 2020 참조.

109) 안성조, 앞의 논문, 82쪽.

110) Asaro, “On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanization of Lethal Decision-making”, p. 693; Sharkey, *op. cit.*, p. 790; Sassòli, *op. cit.*, p. 323.

볼 수도 있겠지만,¹¹²⁾ 현재 이러한 AI는 존재하지 않을 뿐 아니라 장차 출현 가능성을 예단하기도 어렵다.

자율무기체계의 책임 주체성에 관한 논의는 국제인도법상 전투원의 지위를 인정할 것인지 여부에 관한 논의와도 관련이 있다. 국제인도법상 전투원은 교전자격을 갖추고 적대행위에 참여할 권리를 갖는 자로 이해된다.¹¹³⁾ 일부 견해는 인간 전투원을 대신하여 자율무기체계가 의사결정과 교전행위를 한다면 단순한 무기가 아니라 전투원의 속성을 갖는다고 보기도 한다.¹¹⁴⁾ 그러나 앞서 책임 주체성과 관련하여 살펴본 것처럼 현 단계의 기술 수준에서는 자율무기체계에 전투원의 지위를 인정할 수는 없다. 한편 이 문제는 뒤에서 볼 지휘관 책임의 성부와도 밀접한 관련이 있다.

2. 국제법적 통제

2.1 서설

기술과 문화의 놀라운 발전에도 불구하고 인류는 아직 전쟁이라는 폭력을 종식시키지 못했다. 그러나 그로티우스(H. Grotius)가 ‘정당한 무력사용’이라는 개념을 제시한 이래¹¹⁵⁾ 적어도 군사상 필요성과 인도적 고려라는 한계 내에서만 전쟁이 정당화된다는 관념은 널리 받아들여졌다. 이처럼 무력 분쟁의 수단을 통제하기 위한 일련의 국제법 체계를 국제인도법(International Humanitarian Law: IHL)이라고 한다.¹¹⁶⁾ 현대 국제인도법의 토대는 1899년과 1907년의 헤이그 육전규칙과 1949년에 체결된

111) Sparrow, “Killer Robots”, p. 72; Bonnie Docherty, *Mind the Gap: The Lack of Accountability for Killer Robots*, Human Rights Watch, 2015, p. 19.

112) Jack M. Beard, “Autonomous Weapons and Human Responsibility”, *Georgetown Journal of International Law* Vol. 45, Georgetown University Law Center, 2014, p. 663.

113) 안준형, “국제인도법상 불법전투원의 법적 지위”, 서울국제법연구 제28권 제1호, 서울국제법연구원, 2021, 135쪽.

114) Tim McFarland, *Autonomous Weapon Systems and the Law of Armed Conflict: Compatibility with International Humanitarian Law*, Cambridge University Press, 2020, pp. 69-70, 159.

115) Stephen C. Neff(ed.), *Hugo Grotius on the Law of War and Peace*, Cambridge University Press, 2012, pp. 326-327.

116) 국제인도법은 전쟁법(law of war) 또는 무력충돌법(laws of armed conflict)으로 불리기도 한다. 국제인도법은 무력충돌 상황에서 민간인과 비전투원 보호 문제를 규율하는 것으로 주로 무력충돌이 아닌 상황에서 인권문제를 다루는 국제인권법의 특별법이라고 할 수 있다.

4건의 제네바협약¹¹⁷⁾에 의하여 마련되었다. 이후 제네바협약은 1977년에 2건의 추가 의정서¹¹⁸⁾(이하 제1추가의정서 및 제2추가의정서라고 한다)에 의하여 보충되었다.

현대 국제인도법은 신무기나 신기술에 의해 야기된 새로운 도전에 여러 방식으로 적응하고 발전해왔다.¹¹⁹⁾ 그렇다면 자율무기체계에 대하여도 국제인도법이 적절히 대응할 수 있을까. 이 문제는 인도주의 원칙을 정한 마르텐스 조항, 구별·비례·사전 예방을 내용으로 하는 무력충돌의 3대 원칙, 그리고 국제인도법 위반으로 인한 책임 등 3가지 측면에서 검토할 수 있다.

2.2 마르텐스 조항

제1추가의정서는 “인도원칙 및 공공양심의 명령으로부터 연원하는 국제법 원칙의 보호”를 규정하고 있다[§1(1)]. 이러한 법리는 1899년 러시아 외교관 마르텐스(F. Martens)의 제안에 따라 헤이그 제2협약 전문에 처음으로 규정되었기 때문에 ‘마르텐스 조항(Martens Clause)’이라고 불리기도 한다.

인도원칙과 공공양심을 강조하는 마르텐스 조항은 국제법상의 원칙인 동시에 윤리적 원칙의 성격도 지닌다. 자율무기체계 반대론자는 자율살상무기의 사용은 인간의 존엄성을 침해하는 것으로서 공공양심에 반하여 마르텐스 조항을 위배하는 것이라고 주장한다.¹²⁰⁾ 그러나 이에 대하여는 인간이 설계한 프로그램에 따라 임무를 수행하는 이상 인간에 의한 결정이 이루어지지 않는다고 볼 수 없으므로¹²¹⁾ 장거리 미사일이나 포탄 또는 지뢰의 경우와 달리 볼 필요가 없다거나,¹²²⁾ 보다

117) 육전에 있어서의 군대의 부상자 및 병자의 상태개선에 관한 1949년 8월 12일자 제네바협약(제I협약), 해상에 있어서의 군대의 부상자, 병자 및 조난자의 상태개선에 관한 1949년 8월 12일자 제네바협약(제II협약), 포로의 대우에 관한 1949년 8월 12일자 제네바협약(제III협약), 전시에 있어서의 민간인의 보호에 관한 1949년 8월 12일자 제네바협약(제IV협약) 등이 그것이다.

118) 1949년 8월 12일자 제네바 제협약에 대한 추가 및 국제적 무력충돌의 희생자 보호에 관한 의정서(제I의정서), 1949년 8월 12일자 제네바 제협약에 대한 추가 및 비국제적 무력충돌의 희생자 보호에 관한 의정서(제II의정서).

119) 이장희, “신무기기술의 사용과 국제인도법의 적용 문제”, 인도법논총 제33호, 대한적십자사 인도법연구소, 2013, 43쪽; 드론 작전에서의 국제인도법적 쟁점에 관하여는 조홍제/강호중, “드론의 군사적 활용에 따른 국제법적 쟁점-차별의 원칙과 비례성 원칙을 중심으로”, 한국항공우주정책·법학회지 제35권 제1호, 한국항공우주정책·법학회, 2020 참조.

120) Sparrow, “Robots and Respect: Assessing the Case Against Autonomous Weapon Systems”, pp. 106-107, 110.

121) Sassòli, *op. cit.*, pp. 314-315.

122) Ryan Jenkins/Duncan Purves, “Robots and Respect: A Response to Robert Sparrow”, *Ethics & International*

정밀한 무기를 사용하여 희생을 최소화하는 것이 오히려 인도주의적이라는 설득력 있는 반론이 제기되고 있다.¹²³⁾

2.3 무력충돌의 3대 원칙

제1추가개정서는 무력충돌시 준수해야 할 기본원칙으로서 구별의 원칙, 비례의 원칙, 그리고 사전예방의 원칙을 규정하고 있다. 이는 전쟁방법과 수단을 무제한으로 선택할 수 없다는 국제인도법의 이념을 구체화한 것이다.

2.3.1 구별의 원칙

구별의 원칙(principle of distinction)은 멀게는 손자병법이나 성 아우구스티누스의 저작에서도 찾아볼 수 있는 오랜 역사를 지닌 원칙으로서, 민간인과 전투원, 민간물자와 군사목표물의 구별을 전제로 민간인에 대한 공격, 민간물자에 대한 공격, 그리고 무차별적인 공격은 금지된다는 원칙을 말한다[제1추가개정서 §48, §51(4), §52].¹²⁴⁾

구별의 원칙 적용과 관련하여 몇 가지 주의할 점이 있다. 우선 민간인이라도 적대행위에 직접 가담하는 동안에는 보호되지 않는다[§51(3)]. 이 때 ‘적대행위에 직접 가담’한다는 것의 의미가 항상 명백한 것은 아니다.¹²⁵⁾ 예컨대 적군에 대한 편의 제공이 모두 이에 해당하는 것은 아니며,¹²⁶⁾ 군사물자의 생산이나 군수지원과 같은 지원행위가 이에 포함되는지에 관하여도 견해가 대립한다.¹²⁷⁾ 또한 전투원이라 할지라도 전의를 상실한 자에 대한 공격은 금지된다[§41(1)]. ‘전의를 상실한 자’의 의미도 판단이 쉽지 않은데, 제1추가개정서는 적대 당사국의 권력 내에 있거나, 항복할 의사를 분명히 표시하였거나, 의식 상실이나 부상 등으로 자신을 방어할 수 없는 자로서 적대행위를 하지 않고 도피하려 하지 않는 경우에 이에 해당한다고 규정하고 있다[§41(2)].

Affairs Vol. 30 Iss. 3, Cambridge University Press, 2016, pp. 394-395.

123) Lark, *op. cit.*, pp. 69-70.

124) 참고로 비국제적 무력충돌에 적용되는 제2추가개정서는 민간인에 대한 공격만을 금지할 뿐 민간물자에 대한 공격은 금지하지 않고 있다(§13).

125) ICRC는 피해 유발의 목적, 행위와 결과의 관련성, 행위가 상대방의 피해를 위해 고안될 것 등의 요소를 제시한다. Nils Melzer, *Interpretive Guidance on the Notion of Direct Participation in Hostilities under International Humanitarian Law*, ICRC, 2009, p. 46.

126) Department of Defense, *Law of War Manual*, Office of General Counsel, 2016, p. 233.

127) 미국은 이런 경우에도 직접적 적대행위 가담으로 보고 있다. Department of Defense, *Ibid.*, p. 229.

그 밖에 의무요원이나 종교요원은 전투원이 아니므로 민간인과 동일하게 보호된다거나[§43(2)], 양자의 구별이 불명확한 경우 민간인과 민간시설로 간주해야 한다는[§50(1), §52(3)] 등의 보충적 조항들이 존재한다.

인간의 개입 없이 표적을 선정할 수 있는 자율무기체계를 사용할 경우 구별의 원칙이 준수될 수 있을까. 자율무기체계는 인간보다 정보처리 및 식별능력이 우월하기 때문에 일반적으로는 민간인과 전투원 그리고 민간물자와 군사목표물을 인간보다 정확하게 구별할 수 있다. 그러나 실제 전장에서는 상식이나 인간의 의도에 대한 이해를 필요로 하는 불명확한 상황이 많이 발생하는데, 자율무기체계는 아직 이러한 능력을 갖고 있지 않다.¹²⁸⁾ 예를 들어 전의를 상실한 적 또는 반대로 그런 것처럼 위장하고 있는 적[§37(1)],¹²⁹⁾ 적대행위에 가담하고 있는 민간인, 단속적으로 적대행위에 가담하거나,¹³⁰⁾ 적의 시설보호를 위해 자발적으로 인간방패로 참여하거나,¹³¹⁾ 정당방위를 행사하는 민간인, 목적이 불분명하거나 민간인과 군에 의하여 모두 사용되는 이중용도 물건 등의 경우에는 인간의 도움 없이 자율무기체계가 공격의 허용 여부를 판단하기 어렵다.

이처럼 자율무기체계가 구별의 원칙을 준수할 수 있는지 여부가 상황에 따라 달라진다면 절충적 방안을 마련해야 할 것이다. 즉, 명확히 구별이 가능한 경우와 그렇지 않은 경우로 범주를 나누어 전자의 경우에만 자율무기체계를 사용하도록 하는 방안이나,¹³²⁾ 후자의 경우에는 운전자 감독 자율무기체계만 사용하도록 하는 방안이 고려될 수 있을 것이다.

2.3.2 비례의 원칙

비례의 원칙(principle of proportionality)이란 직접적인 군사적 이익에 비하여 민간

128) Sharkey, *op. cit.*, p. 789; Heyns, *op. cit.*, paras. 67-68; Bonnie Docherty/Erik Neunswander/M. Karir/Kate Flinner, *Losing Humanity, The Case against Killer Robots*, Human Rights Watch, 2012, pp. 30-31.

129) Robert Sparrow, "Twenty Seconds to Comply: Autonomous Weapon Systems and the Recognition of Surrender", *International Law Studies Vol. 91*, US Naval War College Press, 2015, p. 709.

130) 이러한 경우에도 보호 대상이 되는지 여부에 관하여 이른바 '회전문 이론(revolving door)'이라는 논쟁이 있어 왔다. 이에 관한 설명은 박문언, "자율무기체계의 국제법적 허용성과 규제방안", 149-151쪽 참조.

131) 합동참모본부는 자발적 인간방패의 경우 민간인으로서의 보호를 그대로 인정하고 있다(합동참모본부, 작전법 실무서, 2016, 52쪽).

132) Schmit/Thurnher, *op. cit.*, p. 246.

인의 인명피해나 민간물자에 대한 피해가 과도한 공격은 금지된다는 원칙이다[제1추가의정서 §51(5)(b), §57(2)(a)(iii), (b)].¹³³⁾ 비례의 원칙은 합리적이고 설득력 있는 것이지만 실제 적용에 있어서 여러 한계점이 지적되어 오기도 했다. 우선 비례의 원칙을 인명피해에 적용하는 것에 대하여는 인간의 생명에 대한 형량이 불가능하다는 비판이 꾸준히 제기되어 왔다.¹³⁴⁾ 전체적 맥락과 다양한 변수를 고려한 고도의 전략적 평가가 요구되기 때문에 그 준수 여부를 정량화하기 어렵다는 한계도 지적된다.¹³⁵⁾ 공격 당시 지휘관이나 전투원의 판단을 기준으로 판단할 것인지 아니면 합리적인 전투원과 지휘관을 기준으로 판단할 것인지도 문제가 된다. 유고슬로비아 국제형사재판소(ICTY)는 후자의 입장에서 동일한 환경에서 충분한 정보를 보유한 사람이 이용 가능한 정보를 합리적으로 활용하였을 경우를 기준으로 판단한 바 있다.¹³⁶⁾

자율무기체계를 사용할 경우 비례의 원칙이 준수될 수 있을까. 인간의 판단 역시 완벽하지 않다면서 이를 긍정하는 견해도 있지만,¹³⁷⁾ 위에서 본 것처럼 비례의 원칙은 어려운 가치 판단과 전체적 맥락을 고려한 질적 판단을 요구하기 때문에 대체로 이를 부정하는 견해가 많다.¹³⁸⁾ 그러나 구별의 원칙에서와 마찬가지로, 비례의 원칙 판단이 가능한 경우와 그렇지 않은 경우를 구분하여 전자의 경우에만 자율무기체계를 사용하는 절충적 접근이 현실적이라고 생각된다.¹³⁹⁾ 예컨대, 군사적 이익과 부수적 피해 사이의 비교형량시 양자가 모두 시설 내지 물건인 경우에는 자율무기체계에 공격 여부의 판단을 맡길 수도 있겠지만, 적군과 민간인의 인명피해는 비교형량이 불합리하므로 인간의 통제가 필요하고, 적 시설물 파괴로 민간인의 인명피해가 예상

133) 비례의 원칙은 적 전투원과 군사목표물의 피해에 대하여는 적용되지 않는다(Stuart Casey-Maslen/Steven Haines, *Hague law interpreted: the conduct of hostilities under the law of armed conflict*, Hart Publishing, 2018, pp. 15-16).

134) United Nations, *Final Report to the Prosecutor by the Committee Established to Review the NATO Bombing Campaign against the Federal Republic of Yugoslavia*, para. 48; Supreme Court of Israel, *Public Committee against Torture in Israel v. Government of Israel*, Case No. HCJ 769/02, 13 Dec. 2006, para. 46.

135) Casey-Maslen, *op. cit.*, p. 16; Sassòli, *op. cit.*, p. 332.

136) United Nations, *Prosecutor v. Galic*, IT-98-29-T, Trial Chamber I, 2003, para. 58; 미국 전쟁법 매뉴얼은 공격 당시 지휘관이나 전투원이 취득할 수 있었던 정보를 기준으로 판단하도록 하고 있다[Department of Defense, *Law of War Manual*, pp. 58, 266].

137) Schmitt, *op. cit.*, p. 26.

138) Heyns, *op. cit.*, paras. 70-74; Docherty, *Losing Humanity: The Case against Killer Robots*, p. 33; Bill Boothby, "How Far Will the Law Allow Unmanned Targeting to Go?", in D. Saxon (ed.), *International Humanitarian Law and the Changing Technology of War*, Martinus Nijhoff, 2013, p. 469.

139) Sassòli, *op. cit.*, p. 338.

되는 경우에는 전적으로 인간이 결정해야 한다는 식의 접근이 가능할 것이다.¹⁴⁰⁾ 참고로 미 국방성이 2023년 발표한 ‘AI 및 자율성의 책임 있는 군사적 이용에 관한 정치적 선언문’은 “국가는 핵무기 사용과 관련된 주권적 결정을 알리고 실행함에 있어 중요한 모든 행위에 인간의 통제와 관여를 유지해야 한다”는 내용을 포함하고 있다.¹⁴¹⁾

한편, 자율무기체계가 인간의 개입 없이 스스로 표적을 선택하고 공격하는 경우에는 통상적인 국제인도법의 경우보다 기준을 엄격하게 설정하여야 할 것이다. 즉, 이 경우 군사적 이익은 전략적 이익이 아닌 구체적 공격상황에서의 작전적 이익으로 제한되어야 하고, 이익의 범위가 불명확한 아군의 보호는 군사적 이익에서 제외되는 것이 적절하다.¹⁴²⁾

2.3.3 사전예방의 원칙

사전예방의 원칙(principle of precautions)은 공격의 계획이나 결정 또는 군사작전의 수행에 있어 민간인 및 민간물자의 피해를 방지하기 위해 실행 가능한 모든 예방조치를 해야 한다는 원칙을 말한다(제1추가 의정서 §57).¹⁴³⁾ 위 원칙은 군사작전 단계에서는 물론 무기의 설계·배치, 심지어 사용 후 폐기에 이르기까지 모든 단계에서 적용된다.¹⁴⁴⁾ 다만 예방조치가 대체로 적절하게 이루어졌다면 일부 미흡한 부분이 있더라도 원칙이 지켜진 것으로 평가된다.¹⁴⁵⁾

140) 박문언, “자율무기체계의 국제법적 허용성과 규제방안”, 181-187쪽.

141) DoD Bureau of Arms Control, Verification and Compliance, “Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy”, 2023. 11. 1, <<https://www.state.gov/political-declaration-on-responsible-military-use-of-artificial-intelligence-and-autonomy/>>, 검색일: 2023. 10. 11; 국제사법재판소는 국가의 사활이 걸린 중대한 위협 아래에서는 핵무기의 사용도 정당화될 수 있다고 한 바 있다 (INTERNATIONAL COURT OF JUSTICE, *Legality of the Threat or Use of Nuclear Weapons*, Advisory Opinion of 8 July 1996, ICJ Reports, p. 226, para 96-97). 그러나 핵무기를 인간의 통제와 관여가 배제된 자율무기체계에 탑재하는 것은 철저히 금지되어야 할 것이다(류병운, 앞의 논문, 75쪽).

142) 사실 국제인도법상 직접적인 군사적 이익에 전략적 이익이나 아군의 보호가 포함될 수 있는지 자체에 관하여도 이견이 없는 것은 아니다. 이들 쟁점에 관한 견해의 대립은 박문언, “자율무기체계의 국제법적 허용성과 규제방안”, 162-169쪽 참조.

143) 1996년 특정재래식무기사용금지조약(CCW) 제2의정서도 ‘실행가능한 예방조치’를 규정하고 있다 [§3(10)].

144) Alexandra Boivin, “The Legal Regime Applicable to Targeting Military Objectives in the Context of Contemporary Warfare”, *Research Paper Series No. 2*, Criminal Justice Commission, 2006, p. 35.

145) United Nations, *Final Report to the Prosecutor by the Committee Established to Review the NATO Bombing*

사전예방의 원칙은 자율무기체계의 개발, 검증·확인(verification and validation: V&V), 시험·평가(test and evaluation: T&E), 전술·기술·절차의 확립 및 사용자 교육, 교신 두절 시 또는 문제 발생시의 조치를 포함한 통제방안 확보 등 모든 과정에 적용되어야 한다. 예컨대 교신 두절 시 자율무기체계에 허용되는 작동 범위는 미승인표적/승인표적, 공격/방어, 살상/비살상 등의 범주로 나누어 미리 설계될 수 있을 것이다.¹⁴⁶⁾ 그리고 해킹이나 사이버공격 등의 문제 발생 시 운용자에게 통보하고, 필요한 경우 공격을 중단하거나 시스템이 종료되도록 하며,¹⁴⁷⁾ 관련 데이터를 기록하여 조사·검증할 수 있도록 하는 방안도 검토될 수 있다. ‘의미 있는 인간 통제’는 이 모든 과정에서 나침반 역할을 할 수 있을 것이다.¹⁴⁸⁾ 미 국방부 훈령 3000.09는 사전예방의 원칙을 구체화하기 위한 방안을 마련하고 있는데, 이에 관하여는 뒤에서 좀 더 자세히 살펴볼 것이다.

사전예방의 원칙은 인도주의와 같은 윤리적 원칙과도 관련이 있다. 앞서 윤리원칙에서 살펴본 투명성, 설명가능성, 추적가능성은 문제점을 파악하고 해결방안을 마련함으로써 사고 예방에 기여한다는 측면에서 사전예방의 원칙의 이행방안이 될 수 있다. 한편 신무기의 법적 검토의무(제1추가개정서 §36) 역시 사전예방의 원칙과 관련이 있는데, 이에 관하여는 아래에서 항을 달리하여 살펴보기로 한다.

2.4 신무기의 법적 검토와 특정 무기의 금지·제한

제1추가개정서는 체약 당사국에게 신무기, 전투수단 또는 방법의 연구·개발·획득 및 채택에 있어서 그 사용이 국제법에 의하여 금지되는지 여부를 검토하여 결정하도록 할 의무를 부과하고 있다(§36). 적법성 검토는 무기체계의 전체 생명주기 동안

Campaign against the Federal Republic of Yugoslavia, para. 29.

146) Paul Scharre, *Autonomous Weapons and Operational Risk: Ethical Autonomous Project*, Center for a New American Security, 2016, p. 45 참조.

147) 만약 자율무기체계의 문제점을 확인했다라도 인간이 자율무기체계를 중단하거나 종료하지 못한다면 실질적으로 사전예방의 원칙이 준수되지 않고 있는 것과 같다(Peter Margulies, “Making Autonomous Weapons Accountable: Command Responsibility for Computer-Guided Lethal Force in Armed Conflicts”, *Roger Williams University Legal Studies Paper No. 166*, Roger Williams University School of Law, 2016, p. 25).

148) Noel Sharkey, “Towards a Principle for the Human Supervisory Control of Robot Weapons”, *Politica & Società* 2/2014, Società Editrice Il Mulino, 2014, p. 305.

요구되며, 성능에 중대한 변화가 있는 경우에도 요구된다. 자율무기체계 역시 다른 신무기와 마찬가지로 이러한 법적 검토의무의 대상이 된다.¹⁴⁹⁾

한편, 어떠한 무기들은 국제법상 사용이 금지되기도 한다. 이들은 i) 그 자체로 금지되는 무기와 ii) 사용되는 방법에 의하여 금지되는 무기로 나누어볼 수 있다. 성질상 과도한 손해나 불필요한 고통을 초래하는 무기의 사용은 그 자체로 금지된다 [제1추가 의정서 §35(2)]. 성질상 무차별적인 살상을 야기하는 무기의 사용 역시 금지된다[§51(4)]. 이러한 국제인도법의 원칙을 구체화하기 위하여 마련된 특정재래식무기금지협약은 개별의정서에서 구체적으로 사용이 금지되는 무기를 정하고 있다.¹⁵⁰⁾ 한편, 성질상 금지되는 무기가 아니더라도 사용 방법에 따라 구별·비례·사전예방 등의 원칙이 준수될 수 없는 무기 역시 사용이 제한될 수 있다.¹⁵¹⁾ 다만 이에 대하여는 적법한 무기가 위법한 방법으로 사용될 수 있더라도 그것만으로 사용이 금지되지는 않는다는 반론이 제기되기도 한다.¹⁵²⁾

자율무기체계는 금지되는 무기일까. 자율무기체계는 속성상 반드시 과도한 손해나 불필요한 고통을 초래하는 것은 아니기 때문에 그 자체로 금지된다고는 볼 수 없다.¹⁵³⁾ 다만 앞서 보았던 것처럼 구별의 원칙이나 비례의 원칙을 지키기 어려운 상황이 있을 수 있으므로 그러한 경우에는 자율무기체계를 사용하지 않거나 운용자의 감독을 받도록 하는 방식으로 사용 방법을 제한할 필요가 있다.

2.5 국제법 위반에 따른 책임

2.5.1 국제형사책임

지금까지 국제인도법의 여러 원칙들을 살펴보았다. 그러나 아무리 훌륭한 원칙이라도 그 위반에 대한 책임을 묻지 못한다면 유명무실하게 될 수밖에 없다. 2002년발효

149) 김자희, “치사성 자율무기시스템에 대한 적법성 검토”, 인도법논총 제38호, 대한적십자사 인도법연구소, 2018, 26-27쪽.

150) 제1의정서는 탐지 불능의 파편 발생 무기, 제2의정서는 지뢰, 부비트랩 및 기타 장치, 제3의정서는 소이성 무기, 제4의정서는 실명 무기, 제5의정서는 전쟁 잔류 폭발물에 대하여 규정하고 있다. 대인지뢰와 집속탄에 관하여는 별도의 조약이 존재한다.

151) ICRC, *A guide to the Legal Review of New Weapons, Means and Methods of Warfare: Measure to Implement Article 36 of Additional Protocol I of 1977*, Cambridge University Press, 2006.

152) Dunlap, *op. cit.*, p. 66.

153) Schmitt/Thurnher, *op. cit.*, pp. 243-265; Boothby, *op. cit.*, pp.57-62; Sassóli, *op. cit.*, pp. 308-310.

된 「국제형사재판소에 관한 로마규정」(이하 ‘로마규정’이라고 하며, 본 항의 조문은 특별한 언급이 없는 한 로마규정의 조문을 뜻한다)¹⁵⁴⁾은 심각한 국제인도법 위반 행위들을 전쟁범죄로 규정하여 행위자를 처벌하도록 하고 있다(§8). 미국이나 중국을 비롯한 강대국들이 비준하지 않고 있다는 한계에도 불구하고 로마규정은 국제인도법의 실효성을 한 단계 높인 것으로서 의의가 크다.

① 개인의 형사책임

전쟁범죄를 저지른 자연인은 개인으로서 형사책임을 부담한다[§25(1)(2)]. 단독정범(간접정범 포함)은 물론 공동정범이나 교사범 또는 방조범으로서 처벌받거나 미수범으로 처벌받을 수도 있다[§25(3)].

다만 주관적 구성요건과 관련하여 다소 문제점이 있다. 일부 전쟁범죄는 과실범의 형태로 되어 있기도 하지만,¹⁵⁵⁾ 로마규정은 원칙적으로 ‘의도와 인식을 지니고(with intent and knowledge)’ 범죄를 저지르는 경우에만 처벌한다(§30). 이는 우리 형법상 주관적 구성요건요소로서의 고의 개념과는 다소 차이가 있다.¹⁵⁶⁾ 즉, 우리 형법의 경우 ‘범죄사실이 발생할 위험을 용인하는 내심의 의사’가 있는 경우에도 미필적 고의(dolus eventualis)를 인정하지만,¹⁵⁷⁾ 로마규정은 이러한 경우는 고의범으로 처벌하지 않는 것이다.¹⁵⁸⁾

자율무기체계를 배치·사용한 결과 민간인에 대한 공격 등이 발생한 경우 현실적으로 그러한 결과에 대한 의도나 인식이 있었다고 인정되는 경우는 드물 것이다. 그러나 그러한 결과 발생에 대한 위험을 용인했다고 볼 수 있는 경우는 충분히 예상할 수 있다. 이 때문에 일부 견해는 자율무기체계를 비롯한 자동화된 무기의 위험성을 통제하기 위해서는 로마조약의 개정이나 판례¹⁵⁹⁾ 발전을 통한 국제관습법의 성립을

154) Rome Statute of the International Criminal Court. 1998. 7. 17. 채택되어 2002. 7. 1.일 발효되었다. 한국은 2000. 3. 8. 서명하고 2002. 11. 8. 국회에서 비준하였다.

155) “군사적 필요에 의하여 정당화되지 아니하며 불법적이고 무분별하게 수행된 재산의 광범위한 파괴 또는 징수”가 이에 해당한다[§8(2)(a)(iv)].

156) 김기준, 국제형사법, 박영사, 2017, 199쪽.

157) 우리 형법상 미필적 고의는 범죄사실의 발생 가능성에 대한 인식이 있고 나아가 범죄사실이 발생할 위험을 용인하는 내심의 의사가 있는 경우에 인정된다(대법원 2017. 1. 12. 선고 2016도15470 판결).

158) Bemba 사건에서 국제형사재판소는 직접고의 내지 제1급 고의(미국 모범형법전의 ‘purposely’에 상응)와 간접고의 내지 제2급 고의(미국 모범형법전의 ‘knowingly’에 상응)만을 고의로 인정하였다(Bemba, PTC II, Decision on the Confirmation of Charges, Case No. ICC-01/05-01/08-424, 15 June 2009).

159) Lubanga, PTC I, Decision on the Confirmation of Charges, Case No. ICC-01/04-01/06, 29 January 2007, para.

통해 미필적 고의의 경우에도 형사책임을 인정할 필요가 있다고 주장하기도 한다.¹⁶⁰⁾

② 지휘관책임

군지휘관(또는 사실상 군지휘관으로서 행동하는 자)은 자신의 실효적인 지휘(또는 권위)와 통제 하에 있는 군대가 전쟁범죄를 범하고 있거나 범하려 한다는 사실을 알았거나 또는 알았어야 함에도 불구하고 이를 방지하거나 억제하기 위하여 필요하고 합리적인 조치를 취하지 않은 경우 군대를 적절하게 통제하지 못한 결과로서의 형사책임을 부담한다[§28(a)]. 로마규정은 이와 유사하게 하급자의 행위에 대한 상급자의 책임도 규정하고 있다[§28(b)]. 지휘관 책임은 군대나 하급자의 기본범죄를 전제로 하지만, 대리책임(vicarious liability)이나 엄격책임(strict liability)이 아니라 자신의 과실로 인한 책임이다.¹⁶¹⁾ 한편, 지휘관이 군대 구성원이나 하급자와 공동으로 범죄 행위를 한 경우에는 앞서 본 개인으로서의 형사책임을 부담한다.

인간이 아닌 자율무기체계와 사이에서도 지휘관계를 인정할 수 있을까. 이에 관하여는 i) 기계와 사이에서는 지휘관계가 인정될 수 없다는 견해,¹⁶²⁾ ii) 자율적 판단능력을 가지는 자율무기체계와의 관계에서는 지휘관계가 인정될 수 있다는 견해,¹⁶³⁾ iii) 지휘관 책임을 유추적용할 수 있다는 견해¹⁶⁴⁾ 등이 제시된 바 있다. 동물의 경우 규범을 익힐 수 없고 훈련된 대로만 움직인다는 점에서 도구에 가까우므로 사용자가 지휘관책임을 지지 않는다. 이에 비해 아동의 경우 비록 형사책임능력은 없지만[§26, §31], 인간으로서 자율적 판단이 가능하므로 소년병의 지휘관은 지휘관책임을 질 수 있다.¹⁶⁵⁾ 자율무기체계는 동물과 아동 가운데 어느 것과 가까울까. 앞서 본 것처럼 자율무기체계에는 형사책임 주체성이 인정될 수 없고, 자율무기체계의 작동은 형사

352. (범죄의 결과가 범죄자의 작위 또는 부작위로 인해 발생할 것이라는 위험을 알았음에도 자신이 그 위험을 받아들인 경우도 고의의 범위로 포함시켰다).

160) 박문연, “자율무기체계의 국제법적 허용성과 규제방안”, 243-245쪽.

161) UNITED NATIONS, In The Appeals Chamber, IT-96-21-A, 20 Feb. 2001, para. 313; William A. Schabas, *An Introduction to the International Criminal Court 4th ed.*, Cambridge University Press, 2011, p. 237.

162) Thompson Chengeta, “Defining the Emerging Notion of Meaningful Human Control in Weapon Systems”, *N.Y.U. J International Law and Politics Vol. 49 No. 3*, New York University Journal of International Law and Politics, 2017, p. 880.

163) Ronald Arkin, et al., *op. cit.*, pp. 13, 18.

164) Heyns, *op. cit.*, p. 15.

165) *Prosecutor v. Dyilo (Lubanga)*, ICC-01/04-01/06-2842, judgement, trial Chamber, Mar. 14. 2012. 참고로 15세 미만의 아동을 군대에 징집 또는 모병하거나 그들을 적대행위에 적극적으로 참여하도록 하는 경우 그 자체가 전쟁범죄가 된다[§8(2)(b)(xxvi)].

법상 행위로 평가될 수 없다. 또한 지휘관책임 역시 형사책임인 이상 죄형법정주의가 적용되므로(§23, §24) 유추해석은 허용되지 않는다. 따라서 인간 지휘관과 자율무기체계와의 관계에서는 지휘관계가 인정되지 않는다. 다만 자율무기체계를 사용한 부하와 그 지휘관 사이의 지휘관계가 인정될 수 있음은 물론이다.¹⁶⁶⁾

2.5.2 국제민사책임

① 개인의 책임

자율무기체계를 사용한 전쟁범죄를 처벌함으로써 국제인도법의 원칙을 관철해가는 것도 중요하지만 위법한 자율무기체계의 작동으로 인해 피해를 입은 사람들을 민사법적으로 구제해주는 것 역시 중요한 일이다. 책임을 부담할 자로는 소프트웨어 개발자, 제조자, 판매자, 사용자, 지휘관 등이 검토될 수 있다.

민사법에서는 원칙적으로 과실책임의 원칙이 지배한다. 그러나 자율무기체계의 경우 피해자 보호를 위하여 과실의 증명이 없는 경우에도 책임을 부담시켜야 한다는 견해가 있다. 영미 보통법에서는 전통적으로 극히 위험한 활동에 대하여 과실을 요하지 않는 엄격책임이 인정되어 왔는데, 공격을 위한 자율무기체계의 사용은 성질상 이러한 범주에 속하므로 무과실책임이 규정되어야 한다는 것이다.¹⁶⁷⁾ 그러나 무기체계의 사용이 위험성을 내포하는 것은 사실이지만 국제법상 그로 인한 피해에 대하여 무과실책임을 인정한 예는 찾아보기 힘들다.¹⁶⁸⁾

② 국가의 책임

현재로서는 피해자가 가해자의 과실을 증명해야 한다. 그러나 복잡하고 자율적으로 작동하는 자율무기체계의 경우 가해자의 과실을 증명하는 것은 매우 어려운 일이다.¹⁶⁹⁾ 이 때문에 자율무기체계의 개발이나 획득을 주도한 국가에게 책임을 부과하는 것이 보다 현실적이라는 주장이 있다.¹⁷⁰⁾ 이와 관련하여 2001년 국제법위원회에서

166) Advisory Council on International Affairs, *Autonomous Weapon Systems: The Need for Meaningful Human Control*, Commissie van Advies inzake Volkenrechtelijke Vraagstukken, 2015, p. 30.

167) 박문언, “자율무기체계의 국제법적 허용성과 규제방안”, 246-248쪽.

168) 우주법이나 국제환경법 등의 영역에서는 엄격책임의 법리를 정한 조약들이 존재한다.

169) Marta Bo, “Autonomous Weapons and the Responsibility Gap in light of the Mens Rea of the War Crime of Attacking Civilians in the ICC Statute”, *Journal of International Criminal Justice* Vol. 19 Iss. 2, Oxford University Press, 2021, p. 276.

170) Daniel N. Hammond, “Autonomous Weapons and the Problem of State Accountability”, *Chicago Journal of*

마련한 「국제적 불법행위에 대한 국가책임 규정」¹⁷¹⁾을 살펴볼 필요가 있다. 위 문서는 국가가 국제적 의무를 위반한 경우 국제적 불법행위 책임을 부담한다고 규정하고 있다(§2). 일부 견해는 위 문서가 의무위반 외에 고의나 과실적 요소를 언급하고 있지 않은 점에 주목하기도 한다.¹⁷²⁾ 다만 국내법원이나 국제법원이 국가의 책임을 인정하는 판결을 과연 내릴 수 있을지 의문이 드는 것도 사실이다.¹⁷³⁾

3. 국내법적 통제

3.1 한국법상 자율무기체계의 통제

헌법상 조약과 국제관습법은 국내법과 같은 효력을 가진다(제6조 제1항). 그러나 한국의 국내법 가운데에도 자율무기체계 통제를 위하여 활용될 수 있는 것들이 존재한다.

우선 형사책임의 경우 사안에 따라 형법상 과실치사상죄(제266조, 제267조, 제228조)나 군형법상 부하범죄 부진정죄(제93조) 등이 적용될 수 있음은 물론이지만, 무엇보다 로마규정을 국내법으로 제정한 「국제형사재판소 관할 범죄의 처벌 등에 관한 법률」이 중요한 역할을 한다. 다만 위 법은 로마규정과 달리 지휘관 책임을 부하의 범죄사실을 ‘알았던’ 경우에만 한정하는 한편(제5조), 과실의 경우 지휘관의 직무태만 죄를 별도로 규정하고 있다(제15조 제2항). 지휘관책임의 법적 성질에 관하여는 통제 의무 불이행을 내용으로 하는 독자적 책임이라는 독립책임설과 하급자의 범행에 상급자가 참가한 것으로 보는 참여책임설의 대립이 있다.¹⁷⁴⁾ 위 법은 독립책임설의 입장에서 규정한 것으로 보이지만¹⁷⁵⁾ 실질적으로는 로마규정과 크게 다르지 않다고 할 것이다. 다만 한국 형법은 미필적 고의의 경우 고의범으로 처벌한다는 점에서 로마규정보다는 형사책임이 인정되는 범위가 넓을 수 있다.

International Law Vol. 15 Iss. 2, University of Chicago Law School, 2015.

171) Articles on the Responsibility of States for Internationally Wrongful Acts.

172) 박문연, “자율무기체계의 국제법적 허용성과 규제방안”, 210쪽.

173) 한희원, 앞의 논문, 356쪽.

174) ICTY, *Prosecutor v. Delalić et al.*, No. IT-96-21-T, para. 1285; *The Prosecutor v. Jean-Pierre Bemba Gombo*, ICC-01/05-01/08, para. 752.

175) 이윤제, “국제범죄에 대한 지휘관책임의 연구”, 서울대학교대학원 박사학위논문, 2016, 150쪽.

민사책임의 경우 사안에 따라 민법상 일반불법행위나 제조물책임법상의 책임이 적용될 수 있을 것이다. 다만 자율무기체계를 사용한 국가를 상대로 책임을 물을 경우에는 상당한 법적 장애가 있을 수 있다. 우선 국가가 손해를 유발한 군의 행위에 대해 국가행위로서 주권면제를 주장하는 경우 피해자에 대한 보상은 사실상 어려워진다. 그리고 군대 파견국과 접수국 사이의 주둔군지위협정(Status of Forces Agreements)에서 면책에 관한 규정을 둔 경우에도 배상청구가 어려워질 수 있다. 다만 이러한 면책을 자율무기체계에 대한 반대의 근거로 삼는 것¹⁷⁶⁾은 부적절하다. 주권면제나 면책조항으로 인한 문제는 자율무기체계에 국한된 것이 아니기 때문이다.¹⁷⁷⁾

3.2 미 국방부 훈령

자율무기체계의 위법한 사용으로 인한 민형사상 책임을 묻는 것은 어디까지나 문제가 발생한 뒤에 이루어지는 사후적 통제이다. 보다 바람직한 것은 문제가 발생하지 않도록 예방하는 것이다. 그런 점에서 자율무기체계의 운용과 통제에 관한 개념과 전략·전술 및 절차를 확립하여 행정법규나 규칙 또는 지침의 형태로 담아내는 것이 아마도 실효적인 국내법적 통제 방안이라고 할 수 있다. 우리나라는 아직 이러한 방안을 마련하지 못하고 있는데,¹⁷⁸⁾ 미국의 사례가 좋은 참고가 될 수 있다.¹⁷⁹⁾

미 국방부는 2012년 국방부 훈령 3000.09를 통해 무기체계의 자율성에 대한 정책을 세계 최초로 공개한 이래, 2018년 국방 AI 전략, 2020년 AI 윤리원칙, 2022년 책임 있는 AI 전략 및 실행안을 차례로 발표하였으며,¹⁸⁰⁾ 2023년에는 그 동안의 성과들을 담아 위 훈령을 개정하였다. 위 훈령은 i) 무기체계의 자율/반자율 기능을 개발하

176) Docherty, *op. cit.*, pp. 27-28.

177) 박문연, “자율무기체계의 국제법적 허용성과 규제방안”, 207쪽.

178) 자율무기체계의 도입에 따른 전략, 작전, 전술의 변화를 고찰한 것으로는 조현석, “인공지능, 자율무기체계와 미래 전쟁의 변환”, 21세기정치학회보 제28권 제1호, 21세기정치학회, 2018, 128-134쪽; 자율무기체계 도입을 가정하여 전략적 차원과 전투수행 기능별로 운용방안을 검토한 연구로는 장용, “미래 무인자율무기체계의 군사적 운용에 관한 연구”, 한국군사 제11권, 한국군사문제연구원, 2022.

179) 한편 호주의 자율무기체계에 대한 국내법적 통제 내용에 관하여는 김민혁/김재오, “자율살상무기체계에 대한 국제적 쟁점과 선제적 대응방향”, 국방연구 제63권 제1호, 국방대학교 국가안전보장문제연구소, 2020, 193-194쪽.

180) 마정목, “미 국방부 무기체계 자율성 훈령 개정에 따른 자율무기체계 정책 분석과 이해”, 124쪽.

고 사용하기 위한 책임 할당과 정책 수립, ii) 자율/반자율 무기체계의 실패 영향과 확률을 최소화하기 위한 가이드라인 수립 등을 목표로 한다.

미 국방부 훈령 3000.09의 주요 내용은 다음과 같다.¹⁸¹⁾ ① 자율무기체계의 개념 정립(§G.2), ② ‘적절한 수준의 인간 판단(appropriate levels of human judgement)’의 행사라는 기본원칙과 이를 위한 강건성, 안전 및 보안, 인간-기계 인터페이스, 투명성 등의 강조[§1.2(a)], ③ 국제인도법 등 법규의 준수[§1.2(b)], ④ 최고위급 정책결정자의 승인 사항 및 예외[§1.2(c)(d)], ⑤ 기존 절차에 의한 군사판매/이양[§1.2(e)], ⑥ 미 국방부 AI 윤리원칙의 적용[§1.2(f)], ⑦ 부서별 책임의 할당[§2], ⑧ 검증확인 및 시험평가 절차의 수립[§3] 등이 그것이다. 특히 ④와 관련하여, 자율/반자율 운용 모드를 가진 무기체계를 포함한 자율무기체계의 경우 원칙적으로 최고위급 정책결정자의 승인이 요구되지만,¹⁸²⁾ 물자(material) 표적에 대한 방어 차원에서의 자율무기체계 활용은 예외적으로 승인 절차가 필요 없다.¹⁸³⁾ 바꾸어 말하면 승인절차를 거치기만 하면 운용자의 감독 없이 표적을 선정하고 교전하는 자율살상무기도 활용할 수 있다는 뜻이 된다. 다만 그 전제는 적절한 수준의 인간 판단의 행사가 확보되어야 한다는 것이다. ‘의미 있는 인간 통제’의 개념을 정교하게 발전시켜야 하는 이유이다.

VI. 결론

AI를 비롯한 기술의 발전에 따라 세계 각국이 자율무기체계 개발에 나서고 있다. 자율무기체계는 그 정확성으로 인해 인명 손실과 재산 피해를 줄일 수도 있지만, 전쟁의 문턱을 낮추고, 인간의 존엄성을 위협하며, 책임의 공백을 가져올 것이라는 우려 때문에 반대하는 견해도 많다. 그러나 단순한 찬반론을 넘어 ‘의미 있는 인간 통제’를 확보하는 것이 중요하다. 그 통제 방안으로는 윤리적 통제와 법적 통제가

181) 자세한 내용은 마정목, 위의 논문 참조.

182) 재래식 무기를 개조하여 자율무기체계로 만들거나 이전에 승인된 자율무기체계가 다시 개조된 경우에도 승인을 받도록 하고 있다. 이 정도 난이도의 승인 절차는 전례 없는 것이라고 한다(Allen, *op. cit.*).

183) 최고위급 정책결정자의 승인 없이 통상의 획득절차를 거치면 되는 경우는, i) 자율 모드가 없는 반자율 무기체계, ii) 적의 포화공격(saturation attack)이나 급박한 공격(time-critical attack)을 국지적으로 방어하기 위해 물자 표적을 선택하고 교전하는데 사용되는 운용자 감독 자율무기체계, iii) 원격조종 또는 자율 기체(vehicles or vessels)를 방어하기 위해 물자 표적을 선택하고 교전하는데 사용되는 운용자 감독 자율무기체계, iv) 물자표적에 대해 비살상, 비물리적인 무력을 행사하는 자율무기체계 등이다.

있을 수 있다.

그러나 논의의 전제로서 우선 자율무기체계의 개념을 정립할 필요가 있다. 이에 관하여 많은 논쟁이 있어 왔지만, 미 국방부의 입장처럼 기능적·관계적 관점에 따라 ‘일단 작동되면 운용자에 의한 추가적인 개입 없이도 표적을 선택하고 공격할 수 있는 무기체계’ 정도로 정의하는 것이 적절할 것이다. 한편 완전자율무기체계 개념을 둘러싼 소모적인 논란에 휘말리기보다는 의미 있는 인간 통제를 실현하는 방안에 주목해야 할 것이다.

윤리적 통제를 위해서는 윤리이론을 바탕으로 실천적 윤리원칙을 마련하는 방안을 검토할 필요가 있다. 그동안 제안되어 온 AI 윤리원칙들을 바탕으로 군사 분야의 특수성을 고려하여 윤리원칙을 마련할 수 있다. CCW의 자율무기체계 지도원칙이나 미 국방부의 AI 윤리원칙 등이 그 사례이다. 모범 관행의 수립이나 정치적 선언 역시 윤리원칙과 마찬가지로 구속력 없는 연성 규범으로서 어느 정도 효과를 발휘할 수 있다.

법적 통제는 국제법과 국내법으로 나누어 검토할 수 있다. 다만 그 전에 자율무기체계의 법적 지위가 문제될 수 있는데, 행위성과 책임주체성이 인정되기는 어려울 것이다. 국제법적 측면에서는 국제인도법의 원칙들과 그 위반에 대한 민형사상의 책임이 문제된다. 자율무기체계는 국제인도법상 그 자체로 금지되는 무기라고는 할 수 없지만, 고차원적인 판단이 요구되는 상황에서는 인간의 도움 없이 구별의 원칙과 비례의 원칙을 준수하기는 어려우므로 일정한 경우 사용방법 상의 제한이 이루어질 필요가 있다. 자율무기체계를 사용한 전쟁범죄에 대하여는 로마규정에 따른 형사책임이 인정되는데, 다만 자율무기체계를 부하로 취급하여 지휘관책임을 물을 수는 없다. 자율무기체계의 위법한 사용으로 인한 피해자는 관련자들을 상대로 민사책임을 물을 수도 있다.

국내법적 통제로서는 사후적으로 민형사상의 책임을 묻는 것 외에 자율무기체계의 운용과 통제에 관한 사전적 규율을 마련하는 방안이 검토되어야 한다. 미국은 국방부 훈령을 통해 ‘적절한 수준의 인간 판단의 행사’를 기본원칙으로 하여 자율무기체계를 개발·운용하고 실패의 영향과 확률을 최소화하기 위한 규율을 마련하였다. 우리도 이를 참고하여 자율무기체계에 관한 개념, 전략 및 전술, 절차를 확립하여 법령이나 지침으로 문서화할 필요가 있다.

참고문헌

1. 단행본

김기준, 국제형사법, 박영사, 2017.

한국인공지능법학회, 인공지능과 법, 박영사, 2019.

합동참모본부, 작전법 실무서, 2016.

Bill Boothby, “How Far Will the Law Allow Unmanned Targeting to Go?”, in D. Saxon (ed.), *International Humanitarian Law and the Changing Technology of War*, Martinus Nijhoff, 2013, pp. 43-63.

Dan Saxon, *Fighting Machines: Autonomous Weapons and Human Dignity*, University of Pennsylvania Press, 2022.

Department of Defense, *Law of War Manual*, Office of General Counsel, 2016.

ICRC, *A guide to the Legal Review of New Weapons, Means and Methods of Warfare: Measure to Implement Article 36 of Additional Protocol I of 1977*, Cambridge University Press, 2006.

Nils Melzer, *Interpretive Guidance on the Notion of Direct Participation in Hostilities under International Humanitarian Law*, ICRC, 2009.

Paul Scharre, *Army of None: Autonomous Weapons and the Future of War*, W. W. Norton & Company, 2018.

Peter Warren Singer, *Wired for War: The Robotics Revolution and Conflict in the 21st century*, Penguin Books, 2009.

Robert Coram, *Boyd: The Fighter Pilot Who Changed the Art of War*, Back bay Books, 2002.

Stephen C. Neff(ed.), *Hugo Grotius on the Law of War and Peace*, Cambridge University Press, 2012.

Stuart Casey-Maslen /Steven Haines, *Hague law interpreted: the conduct of hostilities under the law of armed conflict*, Hart Publishing, 2018.

Tim McFarland, *Autonomous Weapon Systems and the Law of Armed Conflict: Compatibility with International Humanitarian Law*, Cambridge University Press, 2020.

William A. Schabas, *An Introduction to the International Criminal Court 4th ed.*, Cambridge University Press, 2011.

Yuna Huh Wong, et al., *Deterrence in the Age of Thinking Machines*, Rand Corporation, 2020.

2. 학술지

김민혁, “북한 인공지능 기술의 군사화와 우리 군의 대응 무기체계 발전방향 연구”, 한국IT서비스학회지 제20권 제1호, 한국IT서비스학회, 2021, 29-40쪽.

김민혁/김재오, “자율살상무기체계에 대한 국제적 쟁점과 선제적 대응방향”, 국방연구 제63권 제1호, 국방대학교 국가안전보장문제연구소, 2020, 171-204쪽.

김보연, “인공지능을 통한 전쟁수행은 정당인가 -자율무기체계를 통한 국제인도법 준수가능성을 중심으로-”, 고려법학 제107호, 고려대학교 법학연구원, 2022, 413-448쪽.

김자희, “치사성 자율무기시스템에 대한 적법성 검토”, 인도법논총 제38호, 대한적십자사 인도법연구소, 2018, 9-41쪽.

류병운, “드론과 로봇 등 자율무기의 국제법적 적법성”, 홍익법학 제17권 제2호, 홍익대학교 법학연구소, 2016, 61-80쪽.

마정목, “통제가능한 자율무기체계의 개념과 설계에 관한 연구”, 국방정책연구 제36권 제2호, 한국국방연구원, 2020, 81-102쪽.

_____, “강화학습 Q러닝 기반의 자율무기체계 통제 전략에 관한 고찰”, 한국군사학논집 제78권 제2호, 육군사관학교 화랑대연구소, 2022, 425-446쪽.

_____, “미 국방부 무기체계 자율성 훈령 개정에 따른 자율무기체계 정책 분석과 이해”, 국방연구 제66권 제2호, 국방대학교 국가안전보장문제연구소, 2023, 111-130쪽.

박문언, “자율무기체계의 발전과 우리 군의 현실에 대한 고찰”, Law & Technology 제12권 제6호, 서울대학교 기술과법센터, 2016, 51-67쪽.

_____, “자율무기체계에 대한 국제적 논의와 우리 군의 대비방향”, 국방논단 제1920호, 한국국방연구원, 2022, 1-12쪽.

안성조, “인공지능 로봇의 형사책임: 논의방향의 설정에 관한 몇 가지 발전적 제언”, 법철학연구 제20권 제2호, 한국법철학회, 2017, 77-122쪽.

- 안준형, “국제인도법상 불법전투원의 법적 지위”, 서울국제법연구 제28권 제1호, 서울국제법연구원, 2021, 131-165쪽.
- 유준구, “자율살상무기체계의 논의 동향과 쟁점”, 정책연구시리즈 2019-18, 국립외교원 외교안보연구소, 2019, 1-32쪽.
- 윤혜선, “인공지능 규제 정책에 관한 연구: 주요국의 규제 정책 사례를 중심으로”, 정보통신정책연구 제26권 제4호, 정보통신정책학회, 2019, 135-176쪽.
- 이상용, “인공지능과 법인격”, 민사법학 제89호, 한국민사법학회, 2019, 3-49쪽.
- _____, “알고리즘 규제를 위한 지도-원리, 구조, 내용-”, 경제규제와 법 제13권 제2호, 서울대학교 공익산업법센터, 2020, 129-159쪽.
- 이수훈/유영철, “러시아 안보·국방 분야의 인공지능 기술 도입과 개발 현황: 분석과 한계”, 중소기업 제44권 제3호, 한양대학교 아태지역연구센터, 2020, 87-116쪽.
- 이장희, “신무기기술의 사용과 국제인도법의 적용 문제”, 인도법논총 제33호, 대한적십자사 인도법연구소, 2013, 39-57쪽.
- 이종관/한창희, “미래전과 국방 인공지능 체계”, 한국통신학회논문지 제44권 제4호, 한국통신학회, 2019, 782-790쪽.
- 임예준, “인공지능 시대의 전쟁자동화와 인권에 관한 소고 -국제법상 자율살상무기의 규제를 중심으로-”, 고려법학 제92호, 고려대학교 법학연구원, 2019, 265-302쪽.
- 장 용, “미래 무인자율무기체계의 군사적 운용에 관한 연구”, 한국군사 제11권, 한국군사문제연구원, 2022, 135-155쪽.
- 조현석, “인공지능, 자율무기체계와 미래 전쟁의 변환”, 21세기정치학회보 제28권 제1호, 21세기정치학회, 2018, 115-140쪽.
- 조흥제/강호중, “드론의 군사적 활용에 따른 국제법적 쟁점-차별의 원칙과 비례성 원칙을 중심으로-”, 한국항공우주정책·법학회지 제35권 제1호, 한국항공우주정책·법학회, 2020, 127-152쪽.
- 한희원, “인공지능(AI) 치명적 자율무기(LAWs)의 법적·윤리적 쟁점에 대한 기초연구”, 중앙법학 제20집 제1호, 중앙법학회, 2018, 325-365쪽.
- Alexandra Boivin, “The Legal Regime Applicable to Targeting Military Objectives in the Context of Contemporary Warfare”, *Research Paper Series No. 2, Criminal Justice*

- Commission, 2006, p. 56.
- Charles J. Dunlap Jr., “Accountability and Autonomous Weapons: Much Ado about Nothing?”, *Temple International & Comparative Law Journal No. 30*, Temple University School of Law, 2016, pp. 63-76.
- Christopher P. Toscano, “Friend of Humans: An Argument for Developing Autonomous Weapon Systems”, *Journal of National Security Law & Policy Vol. 8 No. 1*, Georgetown University Law Center, 2015, pp. 1-82.
- Daniel N. Hammond, “Autonomous Weapons and the Problem of State Accountability”, *Chicago Journal of International Law Vol. 15 Iss. 2*, University of Chicago Law School, 2015, pp. 652-687.
- Filippo Santoni de Sio/Jeroen van den Hoven, “Meaningful Human Control over Autonomous Systems: A Philosophical Account”, *Frontiers in Robotics and AI Vol. 5 Article 15*, Frontiers Media S.A., 2018, pp. 1-14.
- Jack M. Beard, “Autonomous Weapons and Human Responsibility”, *Georgetown Journal of International Law Vol. 45*, Georgetown University Law Center, 2014, pp. 617-681.
- Kenneth Anderson/Matthew Waxman, “Law and Ethics for Robot Soldiers”, *Policy Review Dec. 2012 & Jan. 2013 Iss. 176*, Hoover Institution Press, 2013, pp. 35-49.
- Marco Sassóli, “Autonomous Weapons and International Humanitarian Law: Advantages, Open Technical Questions and Legal Issues to Be Clarified”, *International Law Studies Vol. 90*, US Naval War College Press, 2014, pp. 308-340.
- Marta Bo, “Autonomous Weapons and the Responsibility Gap in light of the Mens Rea of the War Crime of Attacking Civilians in the ICC Statute”, *Journal of International Criminal Justice Vol. 19 Iss. 2*, Oxford University Press, 2021, pp. 275-299.
- Martin Lark, “The future of Killing: Ethical and Legal Implications of Fully Autonomous Weapon Systems”, *Salus Journal Vol. 5 No. 1*, Charles Sturt University, 2017, pp. 62-73.
- Michael Schmitt/Jeffrey Thurnher, “Out of the Loop: Autonomous Weapon Systems and the Law of Armed Conflict”, *Harvard National Security Journal Vol. 4*, Harvard

- Law School Journals, 2013, pp. 231-281.
- Noel E. Sharkey, "The Evitability of Autonomous Robot Warfare", *International Review of the Red Cross* Vol. 94 No. 886, Cambridge University Press, 2012, pp. 787-799.
- _____, "Towards a Principle for the Human Supervisory Control of Robot Weapons", *Politica & Società* 2/2014, Società Editrice Il Mulino, 2014, pp. 305-324.
- Paul W. Kahn, "The Paradox of Riskless Warfare", *Philosophy & Public Policy Quarterly* Vol. 22 No. 3, George Mason University, 2002, pp. 2-8.
- Peter Asaro, "How Just Could a Robot War Be?", *Current Issues In Computing and Philosophy* Vol. 175, IOS Press, 2008, pp. 50-64.
- _____, "On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanization of Lethal Decision-making", *International Review of the Red Cross* Vol. 94 Iss. 886, Cambridge University Press, 2012, pp. 687-709.
- Peter Margulies, "Making Autonomous Weapons Accountable: Command Responsibility for Computer-Guided Lethal Force in Armed Conflicts", *Roger Williams University Legal Studies Paper No. 166*, Roger Williams University School of Law, 2016, pp. 1-26.
- Robert Sparrow, "Killer Robots", *Journal of Applied Philosophy* Vol. 24 No. 1, Wiley, 2007, pp. 62-77.
- _____, "Twenty Seconds to Comply: Autonomous Weapon Systems and the Recognition of Surrender", *International Law Studies* Vol. 91, US Naval War College Press, 2015, pp. 699-728.
- _____, "Robots and Respect: Assessing the Case Against Autonomous Weapon Systems", *Ethics & International Affairs* Vol. 30 No. 1, Cambridge University Press, 2016, pp. 93-116.
- Ronald Craig Arkin/Patrick Ulam/Alan R. Wagner, "Moral Decision-Making in Autonomous Systems: Enforcement, Moral Emotions, Dignity, Trust and Deception", *IEEE* Vol. 100 Iss. 3, Proceedings of the IEEE, 2011, pp. 571-589.
- Ryan Jenkins/Duncan Purves, "Robots and Respect: A Response to Robert Sparrow", *Ethics & International Affairs* Vol. 30 Iss. 3, Cambridge University Press, 2016, pp.

391-400.

Sabine Gless/Emily Silverman/Thomas Weigend, “If Robots Cause Harm, Who is to Blame? Self-driving Cars and Criminal Liability”, *New Criminal Law Review Vol. 19 No. 3.*, University of California Press, 2016, pp. 412-436.

Simon Chesterman, “Artificial Intelligence and the Limits of Legal Personality”, *International and Comparative Law Quarterly Vol. 69 Iss. 4*, Cambridge University Press, 2020, pp. 819-844.

Thomas W. Simpson/Vincent C. Müller, “Just war and robots’ killings”, *The Philosophical Quarterly Vol. 66. No. 263*, Oxford University Press, 2016, pp. 302-322.

3. 학위논문

박문언, “자율무기체계의 국제법적 허용성과 규제방안”, 서울대학교대학원 박사학위 논문, 2019.

이윤제, “국제범죄에 대한 지휘관책임의 연구”, 서울대학교대학원 박사학위논문, 2016.

4. 웹자료

CSIS, Gregory C. Allen, 2022. 6. 6, “DoD Is Updating Its Decade-Old Autonomous Weapons Policy, but Confusion Remains Widespread”, <<https://www.csis.org/analysis/dod-updating-its-decade-old-autonomous-weapons-policy-confusion-remains-widespread>>, 검색일: 2023. 10. 11.

5. 기타자료

윤지영/김한균/감동근/김성돈, 법과학을 적용한 형사사법의 선진화방안 8: 인공지능 기술, 한국형사정책연구원, 2017.

Andrew P. William, et al., *Autonomous Systems: Issues for Defence Policymakers*, NATO Headquarters Supreme Allied Commander Transformation, 2015.

Bonnie Docherty, *Mind the Gap: The Lack of Accountability for Killer Robots*, Human Rights Watch, 2015.

Bonnie Docherty/Erik Neunschwander/M. Karir/Kate Flinner, *Losing Humanity, The Case*

- against Killer Robots*, Human Rights Watch, 2012.
- Christof Heyns, *Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions*, UN Doc. A/HRC/23/47, 2013.
- Defense Science Board, *Task Force Report: The Role of Autonomy in DoD Systems*, Defense Science Board, 2012.
- Department of Defense, *Summary of the 2018 Department of Defense Artificial Intelligence Strategy*, Department of Defense, 2018.
- Elsa B. Kania, *Chinese Military Innovation in Artificial Intelligence*, Center for a New American Security, 2019.
- European Parliament, *European Parliament Resolution of 12 September 2018 on Autonomous Weapon Systems(2018/2752(RSP))*, European Parliament, 2018.
- GGE, *Report of the 2022 session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems*, GGE, 2022.
- GGE, *Report of the 2023 session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems*, GGE, 2023.
- Marco Sassòli, *Potential Advantages for the Respect of International Humanitarian Law, Professionals in Humanitarian Assistance and Protection(PHAP)*, 2013.
- Multinational Capability Development Campaign, *Role of Autonomous Systems in Gaining Operational Access*, MCDC, 2014.
- Paul Scharre, *Autonomous Weapons and Operational Risk: Ethical Autonomous Project*, Center for a New American Security, 2016.
- Philip Alston, *Interim Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions*, UN Doc. A/65/321, 2010.
- Tim McFarland, *Autonomous Weapons and Human Control*, Humanitarian Law and Policy, 2018.
- UNIDIR, *The Weaponization of Increasingly Autonomous Technologies: Considering How Meaningful Human Control Might Move the Discussion Forward*, UNIDIR, 2014.

[Abstract]

Law and Ethics of Autonomous Weapon System

Sang Yong Lee*

Advancements in technology, particularly in AI, have spurred countries worldwide to develop autonomous weapon systems (AWS). AWS offer the potential to reduce human casualties and property damage due to their precision but also raise concerns about warfare thresholds, human dignity, and accountability gaps.

Addressing these concerns requires establishing meaningful human control over AWS through ethical and legal frameworks. Defining AWS as systems capable of autonomous target selection and engagement, as per the U.S. Department of Defense, is essential.

Ethical control involves developing practical ethical principles, adapted from AI ethics, as a foundation. These principles should address the unique challenges of military applications, such as the CCW's AWS guiding principles and the U.S. Department of Defense's AI ethics principles.

Legal control spans international and domestic law. AWS blur the lines of responsibility, making determining their legal status challenging. While international law addresses principles like proportionality and distinction, implementing them autonomously is difficult. War crimes committed with AWS can be prosecuted under the Rome Statute, but holding commanders accountable for AWS actions is complex. Victims can pursue civil liability against responsible parties.

Domestically, regulations must encompass post facto civil and criminal liability and proactive rules governing AWS. The U.S. has set guidelines emphasizing human judgment levels to minimize AWS-related risks. Other nations can use similar frameworks to formalize AWS concepts, strategies, tactics, procedures, and administrative regulations.

In conclusion, the global development of AWS, driven by technological progress, presents both benefits and concerns. To navigate this complex landscape, it is essential to ensure

* Professor, Law School of Konkuk University

substantial human control over AWS, defining the concept, implementing ethical principles, and establishing legal controls. Such measures are crucial to address the ethical and legal implications arising from the evolution and deployment of autonomous weapon systems.

[Key Words] AWS, LAWS, MHC, REAIM, DoDD 3000.09