



Center for Future Warfare Studies,
Institute of International Studies at Seoul National University |
국제문제연구소 미래전 연구센터 워킹페이퍼 No.52 (발간일: 2020.7.6.)

첨단 방위산업의 국제규범

조동준 서울대학교 정치외교학부 교수

〈차 례〉

- I. 들어가며
- II. AI 기반 자율무기의 다면성
 1. AI 기반 자율무기의 분류
 2. AI 기반 자율무기와 로봇의 연결
- III. 새 술 vs. 헌 부대
 1. 전쟁법의 세 원칙과 AI 기반 자율무기 간 불일치
 2. 인도법의 세 원칙과 AI 기반 자율무기 간 불일치
- IV. AI 무기화 반대운동의 성장
 1. 분절된 두 집단의 만남
 2. 의제화
- V. 나아가며

I. 들어가며

2018년 4월 인공지능(Artificial Intelligence, 이하 AI) 전문가 집단과 한국과학기술원(이하 KAIST) 간 상호작용은 첨단 방위산업에서 규범의 잠재적 중요성을 보여준다. AI의 군사적 이용을 우려한 전문가 집단은 2018년 2월 KAIST가 한화시스템과 함께 국방시융합연구센터를 설립하여 “군사 무기에 적용될 수 있는 AI 기술을 개발하여 세계자율무기 개발 경쟁에 참여한다”는



보도를 지적하면서(Jun 2018), KAIST 신성철 총장이 “사람의 통제가 중요하지 않은 (lacking meaningful human control) 자율무기”를 개발하지 않겠다는 약속을 표명하기 전까지 연구 교류를 중단하겠다고 선언했다 (Walsh et al. 2018).¹⁾ KAIST 신성철 총장은 KAIST가 인권과 윤리 기준을 중시하며 “인간의 존엄성에 반하는 연구 활동을 하지 않겠다”는 원론적 입장을 밝힌 후, “치명적 자율무기체계와 살인 로봇”(lethal autonomous weapons systems and killer robots)을 개발하지 않겠다고 약속하였다 (Sonh 2018). 이 사건을 통하여 AI 관련 연구활동이 자율무기의 개발 가능성 우려로 인하여 제약을 받을 수 있다는 점이 드러났다.

사람이 통제하지 못하는 자율무기를 왜 금지하려고 할까? 자율무기가 전장의 풍경을 획기적으로 바꿀 위험이 크다는 현실적 이유도 있지만 (Singer 2009, 203), 신성철 총장의 성명서에 등장하는 “인간 존엄”이 대안적 답을 제공한다. 자율무기에 의한 살상이 현재 인류가 가진 규범과 충돌하기 때문이다. 즉, 존엄하지 않는 기계가 존엄한 사람의 생명을 취할 수 없다는 믿음이다. 원거리 타격이 가능한 무기가 이미 인류에게 익숙해졌기에 무기를 조작하는 사람이 직접 살상에 관여해야 한다는 규범이 사라졌지만, 전투원이 살상 여부를 결정하는 주체이어야 한다는 규범이 존재한다. 전투 중이라도 존엄한 사람만이 전투원의 생명을 취할 수 있다는 믿음이 강고하다.

이 글은 사람의 통제 밖에 있는 자율무기를 금지하는 규범 창발을 분석한다. 이 글은 세 부분으로 구성된다. 먼저 무기에 대한 사람의 통제 정도를 이해하기 위하여 자율무기와 관련된 용어와 실제 도입된 자율무기를 검토한다. 또한, AI 기반 자율무기가 현존하지 않는 상태에서 자율무기와 관련된 다양한 용어가 경쟁하는 양상을 검토한다. 둘째, 사람의 통제를 벗어난 자율무기의 금지를 둘러싼 논쟁을 정리한다. 기존 무기를 금지 또는 규제하는 국제법의 근간이 비례성의 원칙과 차별성의 원칙으로 요약될 수 있는데, 자율무기의 속성으로 인하여 기존 국제법 원칙을 적용하기 어렵다. 셋째, AI 기반 자율무기를 금지하려는 규범의 창발 과정을 검토한다. AI 기반 자율무기가 아직 등장하지 않은 상황에서 AI 기반 자율무기를 금지하려는 운동은 이례적이다.²⁾ 이

1) 이 사건은 Korea Times의 부정확한 보도 때문에 일어났다. 상기 센터의 개소식 관련 보도 자료에 따르면, 상기 센터는 “AI를 기반으로 하는 지능형 항공기 훈련시스템을 비롯해 지능형 물체추적·인식기술, 지휘결심지원체계, 대형급 무인잠수정 복합항법 알고리즘 개발 등을 우선 과제로 선정해 산학 협동 연구개발(R&D) 방식으로 공동연구를 진행”할 계획이었다. 상기 센터의 개소식에서 한화시스템 장시권 대표이사가 “KAIST와 긴밀히 협력해 기존 무기체계에 AI 기술을 융합한 혁신적인 솔루션을 제공하는 것은 물론 글로벌 기술 경쟁력을 갖추도록 할 것”이라고 말했다(연합뉴스 2018.2.20). 상기 보도 자료와 장시권 대표이사의 발언에 ‘자율무기’가 없었지만, Korea Times의 보도에는 ‘기존 무기체계’ 대신에 ‘자율무기’가 들어갔다. Korea Times의 부정확한 보도가 AI의 무기화를 우려하는 전문가의 관심을 불러 일으켰다.

2) 미국은 2004년 사실상 완전자율무기에 해당되는 무기체계(special weapons observation remote reconnaissance direct action system)를 개발하여 M249 자동소총 또는 M240B 기관총을 장착까지 하였으나 실전배치를 하지 않았다(Jones 2007). 이 사례가 보여주듯이, AI 기반 완전 자율무기는 이미 기술적으로 개발 가능상태에 있다. 현재 국제사회는 여섯 가지 무기(대인지뢰, 독가스, 집속탄, 납탄, 실명용 레이저 무기, 팽창 탄환)을 금지하고 있다. 대인지뢰, 독가스, 집속탄, 납탄은 전장에서 이미 사용된 후 금지된 반면 실명용 레이저 무기와 팽창 탄환은 전장에서 사용되기 전 금지되었다.



현상은 객관적 위협을 줄이기 위한 기존 군축운동과 큰 차이를 보인다.

II. AI 기반 자율무기의 다면성

AI 기반 자율무기를 둘러싼 정명(正名) 논쟁은 AI 기반 자율무기의 미래를 결정할 수 있는 중요한 함의를 가진다. 반대 진영에서는 이를 ‘살인 로봇’ 또는 ‘살해봇’(slaughterbot)으로 부르면서 부정적 의미를 투영하려 한다. 반면, AI 기반 자율무기의 도입에 찬성하는 쪽에서는 전투원의 부담과 보급의 어려움을 해결하는데 도움을 주는 ‘지원 로봇’(support robot)으로 표현하여 긍정적 의미를 투영하려 한다 (Ackerman 2014). 동일한 현상을 둘러싼 정명 논쟁 가운데 AI 기반 자율무기가 놓여 있다.³⁾ 이 절은 AI 자율무기의 유형과 발전 상황을 정리한다.

1. AI 기반 자율무기의 분류

AI 기반 자율무기에 관한 정명 논쟁이 일어나는 원인은 이 무기체계가 다양한 측면을 가지고 있기 때문이다. 자율성이 탐지 과정, 정보 처리과정, 타격 수단 설정, 타격 결정 과정, 실행 과정에 적용될 수 있기에 AI의 자율성을 어느 과정에 어느 정도 구현하는가에 따라 AI 무기체계는 달라질 수 있다. 즉, AI 기반 자율무기는 사람이 탐지부터 타격까지 과정을 전혀 통제할 수 없는 완전 자율무기부터 기체가 최종 실행만 담당하는 제한적 자율무기까지 포함한다. AI 기반 무기체계를 반대하는 쪽은 탐지에서 실행까지 과정에 AI의 자율성을 과도하게 강조하는 반면, 찬성하는 쪽은 사람의 통제를 과도하게 드러내는 경향을 보인다.

AI 기반 자율무기를 이해하기 위해서는 자율(autonomous) 개념에 세 가지 상호 연결된 하위 개념이 내포되어 있다는 점을 주목할 필요가 있다. 첫째, 자동(automation) 개념으로 사전에 정한 목표가 사전에 특정 상황에서 만들어졌으며 예측가능한 방식으로 실행됨을 의미한다. 자동이 예측 가능한 상황에서 작동해야 사전에 계획한 결과로 안정적으로 이어지기 때문에, 자동 개념은 단순계에서 유효하다. 둘째, 자율(autonomy)은 주어진 목표를 자급·자율(self-sufficient and self-governing) 방식으로 수행하는 능력이다. 목표가 주어진 조건 아래서 목표 실행에 필요한 수단과 방법을 스스로 선택할 수 있어야 한다. 셋째, 독립(independence)은 목표 설정에서 실행까지 모든 과정에서 스스로 규율하고 결정함을 의미한다. 독립성을 가진 개체는 사람과 동일하게 복잡계에서 활동한다 (김민혁·김재오 2020, 177-181; 김종열 2018, 102; Group of Governmental Experts of the High Contracting Parties to CCW 2018).

³⁾ AI 자율무기의 군사적 장점으로 전투원 부담 경감, 예산 절감, 인간 전투원이 감당하지 못하는 전투 역량, 인간의 관여로 인한 오류 감소, 윤리적 장점으로 전투 현장에서 인간 감정의 부작용 통제, 인간 인지능력의 한계 극복, 전투원의 생존 본능에 따른 오발 사고 방지 등이 제기된다. 반면, 윤리적 단점으로 비인간에 의한 살상 대상으로 인간 탐지, 비인간에 의한 인간 살상 결정, 전투시 잘못에 대한 책임 소재의 불분명 등이 제기된다 (Etzioni and Etziono 2017, 72-76; Sharkey 2019, 76-78).



AI 기반 자율무기는 자율성의 정도와 인간 통제의 정도에 따라 분류된다. 첫째, 인간 통제 하 자율무기(human on the loop)는 사람이 탐지에서 실행까지 전체 과정에서 AI의 활동을 감시하고 중지시킬 수 있는 무기체계를 의미한다. 즉, AI의 자율성이 활용되지만, 무기체계를 작동하는 사람이 중요한 결정을 내린다. 사람이 탐지할 지역과 대상을 먼저 설정하면 AI의 자율성이 뒤따라 나오고, AI의 탐지 활동으로 얻어진 자료를 사람이 평가하고 타격 대상을 설정하면 타격 결심을 내리면, AI의 자율성에 따라 실행된다. 사람이 최종 실행과정을 검토하면서, 사전 입력된 실행 계획과 실제 활동의 차이 여부에 따라 실행을 취소할 수 있게 된다.

반(半)자율무기(human in the loop)는 인간이 AI의 활동을 특정 단계에서 통제하지만, 다른 단계에서는 AI의 자율성이 인간의 통제 밖에서 구현되는 무기체계를 의미한다. 사람이 여러 정찰 자료를 종합하여 타격 목표를 사전에 설정하고 타격을 결심하면, 사람이 정한 조건에 부합하는 타격 목표를 AI 기반 자율무기가 타격한다. 이 무기체계가 실행된 이후에 사람은 이를 취소시킬 수 없다. 사전 입력된 타격 목표를 자동적으로 찾아 타격하는 유도무기체계가 이 범주에 든다. 실행 후 취소할 수 없기에 ‘발사 후 망각’(fire and forgot)이라고 불린다.

치명적 자율무기(human out of the loop, fully autonomous)는 일단 만들어지면 AI가 사람의 관여 없이 탐지, 타격 대상 설정, 타격 결정, 실행까지 감당하는 무기다. 완전 자율무기로 불리기도 하는데, 사람이 이 무기체계를 만드는 시점까지 통제한다. 현재까지는 구상 단계이지만, 이 무기가 일단 실행되면, 사전 입력된 프로그램에 따라 작동될 뿐이다. 엄밀한 기준에서 ‘살인 로봇,’ ‘살해봇’이 이 범주에 들어간다.

2. AI 기반 자율무기와 로봇의 연결

AI 기반 자율무기가 부정적으로 ‘살인 로봇,’ 긍정적으로 ‘지원 로봇’으로 표현되는 이유는 크게 두 가지로 나눌 수 있다. 첫째, 로봇이 가진 기술적 특성이다. 로봇은 프로그램에 따라 움직이는 자기통제적 기계로 단순계에서 임무를 자동적으로 수행한다. 로봇이 복잡계에서 자기규율적 방식으로 목표를 수행할 수 있는 가능성을 가진 기계로 묘사되지만, 로봇은 실상 주어진 상황에서 주어진 임무를 자동적으로 실행한다. 로봇이 가진 자동 속성이 AI 기반 자율무기와 일부 부합한다.

둘째, AI가 대중에게 익숙한 로봇에 부합한다. 영화 트랜스포머, 터미네이터, 스타트랙 등에서 좋은 역할을 수행하는 로봇과 악역을 담당하는 로봇이 등장한다. 영화에서 전투를 수행하는 로봇이 사람보다는 기계에 가까운 모습을 하거나(e.g., ‘트랜스포머’에 등장하는 autobots과 decepticons), 사람에 가까운 모습을 하거나(e.g., ‘I·robot’에 등장하는 NS-5 Sonny), 사람과 똑같은 모습을 하고 사람의 행동을 따라할 수도 있지만(e.g., ‘스타트랙’에 등장하는 Lt. Data), 공통적으로 복잡계 상황에서 자기규율적으로 움직이면서 목표를 수행하며, 심지어 목표까지 설정한다. 대중에게 각인된 로봇이 AI의 속성과 일부 부합한다.



AI 기반 자율무기와 로봇의 속성이 일부 부합하지만, 양자의 연결이 자연스럽게 않다. 먼저 현재 언급되는 AI 기반 자율무기와 로봇이 큰 차이를 보인다. AI 기반 자율무기가 이론상 목표까지 자율적으로 설정할 수 있어 복잡계에서도 작동 가능한 무기인 반면, 로봇은 단순계에서 자동으로 반복 임무를 수행한다. AI의 무기화에 반대하는 입장에서는 로봇과 자율무기를 연결시키지만, 실상 로봇과 AI는 상이한 개념에 기반한다. 둘째, 미래과학 관련 영화와 소설에서 등장하는 로봇이 긍정적인 모습과 부정적 모습을 동시에 갖추고 있기 때문이다. 예를 들어, 심지어 영화 터미네이터에서 동일한 외양을 가진 T-800이 1편에서는 부정적 모습으로, 2편에는 긍정적 모습으로 나타난다. 로봇 앞에서 “살인”(killer), “살행”(slaughter), “지원”(support)과 같은 수식어가 붙을 때, 특정한 의미를 가진다.

AI 자율무기와 로봇을 연결시키는 담론은 장점과 단점을 동시에 가지고 있었다 (Carpenter 2016, 58-64). 먼저 단점으로 시간차이다. AI 기반 자율무기가 로봇처럼 작동하는 현상은 미래과학 작품에서만 등장한다. 완전 자율적인 AI 기반 자율무기가 현존하지 않다. 이런 상황에서 앞으로 등장할 AI 기반 자율무기에 로봇을 덧붙이면 위험이 먼 미래에 구현되는 것으로 보인다. 위험이 가까이 있다는 신호가 전달되어야 사회운동이 성공할 수 있는데, ‘로봇’이란 단어가 사회운동의 시급성을 약화시키는 효과를 가진다.

반면, 로봇이 대중에게 익숙한 개념이기 때문에 특정 사회운동이 제기하려는 위험을 상대적으로 쉽게 대중에게 전달할 수 있다. 완전히 새로운 현상을 대중에게 전달하기 보다는 대중이 이미 익숙한 현상에 사회운동이 전달하려는 의미를 덧붙임이 유리할 수 있다. ‘로봇’이란 단어가 AI의 무기화와 반대하는 사회운동을 대중에게 쉽게 다가갈 수 있도록 돕는다. 또한, ‘로봇’이 익숙한 개념이기 때문에 AI의 무기화에 반대하는 여러 분파의 사회운동간 차이점과 이질성을 줄이는 효과를 가진다. 기존 사회운동세력, 학계, 종교계 등 다양한 배경을 가진 분파가 공통 언어를 공유하게 함으로써 분파 간 갈등을 줄였다.

로봇과 AI 자율무기를 연결시키는 시도가 AI 기반 자율무기에 대한 선호에 영향을 미칠까? Young and Carpenter(2018)는 설문조사기법을 동원하여 미래과학 관련 작품에 노출된 정도가 높을수록, 교육 수준이 높을수록 AI 기반 자율무기에 반대하는 경향을 확인하였다. 또한, 응답자가 접한 로봇이 주는 공포감이 높을수록 AI 기반 자율무기에 반대하는 경향을 확인하였다. 반면, 동일한 무기체계를 ‘살인 로봇’ 또는 ‘완전 자율무기’(fully autonomous weapons)로 부르는 프레임효과는 유의미한 결과로 이어지지 않았다. 미래과학 관련 영화와 소설이 AI 기반 자율무기에 대하여 이미 대중에게 부정적 상을 전파한 상황에서, 용어 선택이 결정적 차이를 만들어내지는 못한다는 함의를 가진다.

III. 새 술 vs. 현 부대



AI 기반 자율무기의 위험은 아직 완전하게 현실화되지 않았다. AI가 인류에게 새로운 현상이기 때문에, 이를 금지할 근거가 아직 정리되지 않았다. 과거 무기를 금지하기 위하여 인류가 합의한 원칙이 AI 기반 자율무기에 적용되기 쉽지 않은 상황이다. 최근 성장한 인도법도 AI 기반 자율무기를 금지하는데 한계를 보인다. 이 절은 AI의 무기화를 둘러싼 법적 논쟁을 정리한다.

1. 전쟁법의 세 원칙과 AI 기반 자율무기 간 불일치

사회운동세력이 AI의 잠재적 무기화에 대한 반대를 주도하는 이유는 기존 국제법 체계 안에서 이를 다루기 어렵기 때문이다. 인류는 전투 중 동원되는 무기, 전투 방식을 규율하는데 크게 세 원칙을 발전시켰다.⁴⁾ 첫째, 구별 원칙이다. 전투는 전투원 사이에서만 일어나야 하기 때문에 전투 중 민간인과 전투원이 구별되어야 하고, 전투원 중에서도 전투 의지를 가진 전투원과 투항 의지를 가진 전투원이 구별되어야 한다는 당위에 인류가 합의하였다. 구별 원칙이 무기에 적용되면, 전투원과 민간인을 구별하지 않는 무기의 금지로 이어진다. 구체적으로 핵무기 금지, 집속탄 금지 등으로 나타난다. 또한, 구별 원칙에 부합할 수 있는 무기체계는 구별 원칙에 부합하게 사용되어야 한다. 예를 들어 군사시설을 파괴하기 위하여 소이탄을 사용할 수 있지만, 민간 시설에 대하여 소이탄을 사용해서는 안 된다. 구체적으로 구별 원칙이 무기 사용과 연결되면, 문화재 폭격 금지, 민간인 대상 폭격 금지 등으로 나타난다.

둘째, 비례 원칙이다. 전쟁이 정치적 목적을 달성하기 위한 불가피한 과정으로 수용하더라도 전쟁의 목표를 달성하기 위한 수단이 전쟁의 목표에 어울려야 한다는 당위다. 무기에 비례성의 원칙이 적용되면, 독가스 금지, 납탄 금지, 대인지뢰 금지, 레이저 무기 금지 등으로 나타난다. 또한, 비례 원칙에 부합할 수 있는 무기체계는 전투에 사용될 때도 비례 원칙이 적용되어야 한다. 구체적으로 상대방의 폭력에 대한 비례적 대응, 과도하거나 불필요한 고통을 주는 행위의 금지, 전투 목표와 수단 간 비례성 등으로 이어진다.

셋째, 책임 원칙이다. 전투 중 특정한 행동에 대하여 책임이 명확해야 한다는 당위인데, 이는 불법적 전투 행위의 책임 소재를 추적하여 처벌할 수 있도록 함으로써 불법적 전투 행위를 막으려 한다. 책임은 불법적 무기의 제조, 도입, 사용에 대한 책임과 합법적 무기의 불법적 사용에 대한 책임으로 나뉜다.

기존 전쟁의 세 원칙은 AI 기반 자율무기를 규제하는데 있어 약점을 보인다. 첫째, 책임 원칙을 충족할 수 없다. AI 기반 자율무기는 인간의 통제가 제한되기 때문에, 사고와 불법행위에 대한 책임 소재를 파악하기 어렵다. 예를 들어, 현재 사용되고 있는 반자율무기를 사용하는 과정에

4) 전쟁법의 주요 원칙으로 군사적 필요가 추가될 수 있다. 제네바협약 제1추가의정서(1977년) 제52조는 군사목표를 “그 성질, 위치, 사용 또는 장래 활용 가능성을 통해 상대방에게 군사적으로 효과적인 기여를 하고 당시 상황에 비춰 그것의 전부 또는 일부를 파괴하거나 포획하는 것이 명백한 군사적 이익을 제공하는 물건”으로 정의함으로써, 군사적 필요에 의해서만 전투행위가 진행되어야 한다는 점을 명확히 한다.



서 오발사고가 발생할 경우, 탐지부터 실행까지 과정에서 사람과 시가 함께 관여하기 때문에 오류의 소재를 파악하기 어렵다. 만약 시의 잘못으로 오발사고가 났다면, 프로그램을 만들기로 결정한 사람, 프로그램을 만든 사람, 잘못된 프로그램을 실행한 사람 등에게 책임 소재가 분산된다.

둘째, 시 기반 자율무기가 구별 원칙을 어느 정도 수용할 수 있는지 명확하지 않다. 탐지 단계에서 타격 대상을 사람보다 더 정확하게 구별할 수 있는지 여부가 기술 발달에 달려있기 때문이다. 시의 무기화에 반대하는 쪽은 시가 탐지 단계에서 타격 대상을 구별할 때 일으킬 수 있는 오류에 초점을 맞춘다. 반면, 시의 무기화에 찬성하는 쪽은 시의 구별 능력이 사람의 인지 능력보다 더 뛰어나다는 점에 초점을 맞춘다. 기술이 더 발달하면서, 구별성 쟁점에서는 시의 무기화를 반대하는 쪽이 수세에 몰리는 경향이 보인다.

셋째, 시 기반 자율무기가 비례 원칙을 어느 정도 수용할 수 있는지 명확하지 않다. 시의 무기화를 반대하는 쪽에서 시의 무기화가 과도한 전투 행위로 이어질 위험성을 경고한다. 시의 무기화를 이룬 전투 세력이 자신의 피해를 상대적으로 걱정하지 않기에 과도한 대응으로 이어질 위험을 지적한다. 반면, 시를 무기체계에 도입하자는 쪽은 시 기반 자율무기가 전투 중 사람의 보복 심리, 생존 본능 등을 막을 수 있다고 한다. 사람의 약점이 전투 현장에서 발현되지 않도록 함으로써 전장에서 오히려 인도적 원칙이 준수되게 할 수 있다고 한다. 비례성 원칙에서는 양측의 입장이 팽팽히 맞선다.

정리하면, 기존 전쟁법은 시 기반 자율무기의 도입에 대하여 명확한 답을 제공하지 못한다. 시 기반 자율무기가 다양한 측면을 가지고 있고, 오히려 기존 전쟁법의 원칙에 더 부합하는 영역마저 존재한다.

2. 인도법의 세 원칙과 시 기반 자율무기 간 불일치

전투 행위에 인도적 고려가 필요하다는 입장이 전쟁 과정 중 인도법의 발전으로 이어지고 있다. 전쟁을 규율하는 국제인도법(International Humanitarian Law)은 1899년 '육전의 법 및 관습에 관한 협약(헤이그 제2협약)'[Convention with Respect to the Laws and Customs of War on Land (Hague II)]의 전문에 있는 마르텐스 조항(Martens Clause)으로 거슬러 올라간다. 이 조항은 아래와 같다.

보다 완비된 전쟁법에 관한 법규가 제정되기까지는 당사국이 채택한 규정에 포함되지 않은 사례에서는 주민과 전투원이 인도의 법칙 및 공공양심의 요구로부터 유래하는 국제법 원칙의 보호 및 지배하에 있음을 선언하는 것이 타당하다고 생각하며 ...

상기 문안은 현존 전쟁법이 규율하지 못하는 전투 행위 또는 무기가 등장할 경우 관례, 인도의 법칙, 공공양심이 대안적 법리로 적용될 수 있음을 의미한다. 이 협약 등장 이후 전쟁 자체의



불법화, 대인지뢰금지, 소년병 금지 등 국제인도법이 상기 문안에 기반하여 등장하였다. 즉, 상기 문안은 기존 엄격한 전쟁법 체계에 새로운 규범이 결합될 수 있는 기회의 창이었다 (Mero 2017; Ticehurt 1997).

자율무기에 반대하는 쪽은 마르텐스 조항을 활용하여 세 가지 쟁점을 국제인도법에 투영하려 한다. 첫째, 비인간 주체가 전투원의 목숨을 끊을 수 있는지에 관한 윤리성이다. AI 기반 자율무기의 윤리성 기계가 과연 사람의 목숨을 끊는 행위가 정당할 수 있는가? AI의 무기화를 반대하는 쪽에서는 기계에 의한 살인이 정당하지 않다고 주장한다 (International Committee of the Red Cross 2018, 11-12; Sharkey 2010).

둘째, 인과관계를 예상할 수 있는 예측가능성(predictability)이다. 인류의 양심과 인도적 원칙은 무기 사용이 예측 가능한 범위 안에서 이루어져야 한다는 당위를 포함한다. 과연 AI 기반 자율무기가 초래할 수 있는 결과는 예측 가능한가? AI의 무기화에 반대하는 쪽은 AI를 만드는 인간의 능력을 의심스럽게 바라본다. 사람이 AI 기반 자율무기를 만들 때는 사람이 예측가능한 범위 안에서 프로그램을 만들고 감독할 수 있지만, 실제 AI가 활동하는 환경이 인간의 예측 범위보다 더 복잡적일 수 있다. AI가 인간이 사전에 설계한 방향으로 작동하려 해도 복잡계에서는 제대로 작동할 수 없을 위험이 있다.

셋째, 사용하는 수단의 효과가 믿을 수 있고 일관되게 작동하는 신뢰성(reliability)이다. AI의 무기화에 반대하는 쪽은 AI의 학습능력을 의심스럽게 바라본다. AI가 인간의 감독 아래 학습을 거쳐 실제 작동하기 시작하면, “스스로 학습”할 수 있다. AI 기반 자율무기가 복잡계에서 스스로 제작자의 의도와 다르게 학습하고 부정적 결과를 초래할 수 있다. 2004년 개봉된 i-ROBOT에서 NS-5 Sonny가 제작사의 의도와 다르게 움직이듯이, 미래 AI 기반 자율무기가 인간의 통제를 벗어날 위험성을 제기한다.

반면, AI의 무기화에 긍정적인 입장은 모든 열병기가 일정 정도 사람의 통제 밖에 있음을 강조하며 인간과 무기 간 상호작용을 평가하는 기준이 연속적이라는 현실을 언급한다. 완전 자율무기가 측면에서 문제가 될 수 있겠지만, 현존 AI 기반 자율무기가 사람과 무기 간 상호작용을 전제하기 때문에 이 쟁점이 크게 부각될 이유가 없다는 입장이다. 또한, 신뢰성과 예측가능성에서 인간의 한계는 모든 영역에서 일어나기 때문에, AI에만 특별히 적용될 필요가 없다는 입장이다.

정리하면, 국제인도법이 AI 기반 자율무기의 도입에 대하여 부정적이지만, 국제인도법이 기존 전쟁법보다 구속성이 약하다. 따라서 국제인도법도 AI 기반 자율무기의 도입을 막을 만큼 강한 구속성을 가지지 못한다.

IV. AI 무기화 반대운동의 성장

2013년 Human Rights Watch 연결망 안에서 AI 무기화를 반대하는 입장을 가진 일군의 사



회운동가들은 지뢰금지, 인권 옹호 등 영역에서 오랫동안 공동보조를 취했던 우호 비정부단체를 모아 ‘킬러로봇 중단 캠페인’(International Campaign to Stop Killer Robots)이라는 초연결망을 만들었다. 이 연결망은 매우 짧은 시간 간에 AI의 무기화를 국제사회의 의제로 만들었고, 우호세력을 확보하였다. 이 절은 AI 무기화 반대운동의 성장을 검토한다.

1. 분절된 두 집단의 만남

AI 무기화 반대운동은 2009년 이전 연결되지 않았던 두 이질적 집단의 만남으로 시작되었다. 첫째, 인도적 군축운동을 전개하던 사회운동세력이다. 1990년대 대인지뢰금지를 제기했던 국제 대인지뢰금지운동(International Campaign to Ban Landmines, ICBL)⁵⁾이 예상하지 못할 정도로 빠르게 대인지뢰와 집속탄을 금지하는 국제협약을 이끌어내는 과정에서, 인도적 군축운동이 성장했다. 동시에 ICBL의 성공에 자극을 받아 다른 영역에서 군축 또는 무기규제에 관심을 가진 단체가 등장했다. 2007년 당시 군축운동가들은 운동 역량을 집중할 새로운 대상을 찾고 있었고,⁶⁾ 다수 우호세력이 주변에 존재했다.

AI 무기화 반대운동은 ‘국제인권감시기구’와 연계된 군축전문가로부터 시작하였다.⁷⁾ 2007년부터 2010년 사이 이들은 AI 기반 자율무기의 등장 조짐을 바라보면서 AI 기반 자율무기를 사회운동의 의제로 삼을지 여부를 타진했다. 의제화는 외부 환경의 변화와 규범창발자의 의도적 행위 간 상호작용의 결과인데, 이들은 1990년대 대인지뢰금지 규범의 발전과 성문화 과정, 2000년대 집속탄금지 규범의 발전과 성문화 과정에 참여하였다. 집속탄금지운동이 성공적으로 마쳐질 즈

- 5) 1991년까지 대인지뢰와 관련하여 개별적으로 활동하던 6개 비정부기구가 대인지뢰금지를 목적으로 결성한 비정부기구 연합체이다. 프랑스에 기반을 둔 ‘국제장애위원회’(Handicap International, 1982년 창립), 영국에 기반을 둔 ‘지뢰조언집단’(Mines Advisory Group, 1989년 창립), 독일에 기반을 둔 ‘국제의료진’(Medico International, 1968년 창립), 미국에 기반을 둔 ‘인권을 위한 의사협회’(Physicians for Human Rights, 1986년 창립), 국제사회에서 인권감시에 관여하는 ‘국제인권감시기구’(Human Rights Watch, 1988년 창립), 베트남전쟁 참전 군인으로 구성된 ‘미국 베트남 전쟁 퇴역 군인재단’(Vietnam Veterans of America Foundation, 1980년 창립)이 비정부기구연합체에 참여했다. ICBL은 1997년 ‘오타와협약’(Convention on the Prohibition of the Use, Stockpiling, Production and Transfer of Anti-Personnel Mines and on their Destruction, 1999년 발효)에서 중요한 역할을 담당했다(조동준 2011, 29-34). ICBL은 집속탄금지운동도 전개하여 2008년 ‘더블린협약’(Convention on Cluster Munitions, 2010년 발효)으로 이어졌다.
- 6) 인도적 군축운동은 ‘핵무기폐기국제운동’(International Campaign to Abolish Nuclear Weapons, 2007년 창립), ‘폭발무기금지 국제연결망’(International Network on Explosive Weapons, 2011년 창립), ‘킬러로봇 중단 캠페인’(2013년 창립)으로 분화되었다(Goose and Wareham 2016, 28-29). ICBL 창립에 관여했거나 ICBL과 함께 활동했던 비정부기구가 ‘폭발무기금지 국제연결망’과 ‘킬러로봇 중단 캠페인’의 창립에서 핵심 역할을 담당했다. 반핵운동에 관여하던 비정부기구가 ‘핵무기폐기국제운동’으로 결집하였다.
- 7) ‘국제인권감시기구’는 AI의 무기화 반대 운동을 처음 시작했는데, 이 단체는 소련과 동유럽에서 투옥된 인사를 위한 인권감시와 옹호 활동을 하던 Helsinki Watch(1978 창립)을 모체로 Americas Watch(1981년 창립), Asia Watch(1985년 창립), Africa Watch(1988 창립), Middle East Watch(1989년 창립)를 아우르는 연결망이다. 이 단체는 과거 국가와 무장단체의 인권 침해를 감시하는 활동에 집중했지만, 최근 대인지뢰금지, 소년병 등 전쟁과 인권 간 연계 쟁점을 다룬다. ‘국제인권감시기구’는 ‘폭발무기금지 국제연결망’과 ‘킬러로봇 중단 캠페인’의 창립 회원이다.



음, 군축전문가들은 새로운 무기를 금지하기 위하여 의제를 탐색했다.

둘째, 로봇과 AI의 윤리적 측면에 관심을 가졌던 연구진이다. AI와 로봇과 관련된 연구를 진행하던 전문가는 두 현상과 관련된 윤리적 문제를 대처하기 위한 움직임을 조직화했다. 영국에 기반을 둔 ‘인공지능과 모의행동 학회’(Society for the Study of Artificial Intelligence and Simulation of Behaviour, 1964년 창립) 안에는 오래 전부터 인공지능의 윤리적 문제를 탐색하는 연구진이 있었는데, 2005-2006년 사이 전문 연구모임이 결성되기 시작하였다.⁸⁾ AI에 기반한 로봇이 구현될 수 있는 개연성이 높아짐에 따라, 해당 분야에서 전문성을 가진 연구자와 해당 분야의 윤리적 측면에 관심을 가진 과학철학자가 모임으로 결집하기 시작하였다.

2009년 설립된 ‘로봇무기통제 국제위원회’(International Committee for Robot Arms Control, 이하 ICRC)가 AI와 로봇 관련 연구집단과 사회운동세력을 연결했다. ICRC는 로봇 무기의 등장을 막기 위한 목적으로 결성된 연구자 집단이다. 이 모임은 AI 전문가인 노엘 샤키(Noel Sharkey, 영국 셰필드대학교 명예교수), 물리학자 유르겐 알트만(Jürgen Altmann, 독일 도르트문트대학교 교수), 과학철학자 피터 아사로(Peter M. Asaro, 미국 뉴스쿨 교수), 과학철학자 로버트 스페로우(Robert Sparrow, 호주 모나쉬대학교 교수)로 최초 구성되었는데, AI와 로봇의 무기화를 공통적으로 우려했다. 이들은 과학철학 관련 모임을 통하여 모임을 형성했고, AI 기반 로봇의 무기화에 대한 반대 의견을 사회운동을 발전시킬 의향을 가졌다 (Bolton and Mitchell 2020, 29-40).

ICRC 창립 회원은 AI 기반 로봇의 무기화에 대한 우려를 확산시킬 목적으로 회원 확대에 나섰다. 알트만 교수의 제안에 따라 German Foundation for Peace Research와 영국에 기반을 둔 Joseph Rowntree Charitable Trust로부터 재정 지원을 받아, 독일에서 학회를 기획했다. 이 학회에 AI와 로봇 관련 연구진은 물론 독일 외무성의 군축담당자를 초청하여, 회원 확보는 물론 독일 정부의 지원을 탐색하려 했다. 2010년 학회를 거치면서 회원이 14명이 증가해, ICRC가 연구자 집단으로서 모습을 갖추었다. 이 과정에서 AI와 로봇 관련 전문가보다는 과학철학자가 더 많이 참가하는 경향을 보였다 (Carpenter 2014, 94-95).

2011년 군축 관련 비정부기구인 ‘36조’(Article 36)의 리처드 모이어 집행이사(Managing Director Richard Moyes)가 ICRC 회원으로 가입하면서,⁹⁾ ICRC에 포함된 전문가와 사회운동 세력이 만났다. 모이어 집행이사는 ICBL의 대인지뢰금지운동과 집속탄금지운동에 깊이 관여한 인사로 인도적 군축운동세력과 연결되어 있었다. 그는 ICRC가 느슨한 연구자 모임이 아니라 정식으로 비정부기구로 전환되어 사회운동에 참여하길 권유했다. 또한, ICRC 회원을 인도적 군축운동과 연결시켰다. 2011년까지 연결되지 않았던 두 집단이 모이어 집행이사의 중개로 연결되

⁸⁾ Future of Humanity Institute(2005 창립), Society for the Study of Artificial Intelligence and Simulation of Behaviour, Bristol Robotics Lab(2006년 창립) 등이 이 시기 조직화되었다.

⁹⁾ ‘36조’(Article 36)는 제네바협약 제1 추가의정서 36조(Article 36 of the 1977 Additional Protocol I to the Geneva Conventions)의 이행을 촉구하기 위한 목적으로 만들어진 비정부기구인데, 해당 조항은 새로운 무기의 금지 여부를 당사국이 결정할 수 있다는 내용을 담는다.



었다 (Ibid., 114-117).

두 이질적 집단은 AI 기반 로봇의 무기화 반대 운동에서 역할을 분담했다. ICRC는 AI와 로봇 관련 전문 지식을 인도적 군축운동에 제공하고, 전문성에 기반한 의견을 대외적으로 제시함으로써 인도적 군축집단의 신뢰성을 높였다.¹⁰⁾ 반면, 인도적 군축운동에 참여했던 활동가 집단은 AI와 로봇의 무기화를 규정하는 담론을 만들어 국제사회의 의제로 만들고, 우호세력을 확대하는 작업을 담당했다. 1990년대 진행되었던 대인지뢰금지운동과 유사한 과정이 AI와 로봇 영역에서 다시 전개되었다.

2. 의제화

AI 기반 자율무기가 인도적 군축운동의 의제로 부상한 배경은 ‘테러와의 전쟁’에서 미군의 무인기 사용과 깊게 관련된다. 이라크와 아프가니스탄에서 전투원이 직접 관여하는 작전이 제약받자 미군은 무인기를 활용하기 시작하였다.¹¹⁾ 무인기가 ‘테러와의 전쟁’에 도움을 준다는 언론의 보도가 대중에게 우호적으로 수용되는 현상이 나타나면서, 인도적 군축운동은 AI의 무기화를 우려하기 시작하였다. 초기에는 정찰 수단으로 사용되던 무인기의 활용 영역이 넓어지는 동시에 표적의 탐색에 AI 기법이 활용되면서, AI 기반 자율무기의 위험이 점차 현실화되기 시작하였다.

인도적 군축운동은 AI의 무기화를 부정적으로 표현함으로써, 무인기 사용에 대한 언론과 대중의 허용 태도를 바꾸려하였다. 2010년부터 AI의 무기화에 관한 연구를 시작했는데, 이들의 내부 작업은 2012년 *Losing Humanity*로 집약되었다 (Human Rights Watch 2012). 인도적 군축운동은 “사람의 개입 없이 타격 목표를 선택하고 공격할 수도 있는 완전 자율무기(fully autonomous weapons that could select and engage targets without human intervention)”를 ‘살인로봇’(killer robot)으로 표현하였다 (Ibid., 1). AI 기반 자율무기의 부정적 측면을 드러내는 ‘의제화’(framing) 단계가 시작되었다.¹²⁾ 이후 인도적 군축운동 안에서 ‘킬러 로봇’이라는 용어를 일관되게 사용하였다.

2013년 인도적 군축운동은 ‘킬러로봇 중단 캠페인’으로 결집했다. 창립 회원단체로 ‘퍼그워시 회의’(Pugwash Conferences on Science and World Affairs, 1957년 창립), ‘국제사면위원

10) ICRC가 사회운동에 참여하면서 역설적으로 내부 동력을 상실했다. ICRC 회원 중 AI와 로봇 관련 전문가는 전문성에 기반한 의견 교환을 중시하고 완전 자율무기의 무기화에만 반대 입장을 밝히는 반면, 과학철학자와 사회운동가는 인도적 군축운동의 대의를 중시하고 AI와 로봇의 일반적 윤리에 관심을 가졌다. ICRC 안에서 두 세부 집단 간 차이가 봉합되지 않자, 2019년 이후 ICRC의 활동이 거의 정체 상태로 접어들었다 (Bolton Mitchell 2020, 42-49).

11) 2007년 11월 7일 미군은 최초로 MQ-9A Reaper 무인기에서 유도폭탄을 발사하여 아프가니스탄 반군을 공격했다. 아프가니스탄에서 활동하던 미군이 반군의 거점 좌표를 무인기 운용 부대에 전했고 무인기를 작동하던 요원이 아프가니스탄 상공에 있던 무인기로부터 유도폭탄을 발사했다(Tiernan 2007). MQ-9A Reaper가 AI 기반 자동항법장치에 기반하여 작동되기 때문에, 이 사건이 AI가 실제 전장에서 살상에 활용된 첫 사례로 기록된다.

12) AI 무기화에 반대하는 사회운동가 사이에서 최소한 2011년까지는 AI 기반 자율무기에 로봇 단어를 결부하는 수사를 고려하지 않았다 (Carpenter 2016, 57).



회'(Amnesty International, 1961년 창립), '국제장애위원회'(Handicap International, 1982년 창립), '국제인권감시기구'(Human Rights Watch, 1988년 창립), '여성 노벨상 수상자 계획'(Nobel Women's Initiative, 2006 창립), '그리스도의 평화 국제 위원회'(Pax Christi International, 1945년 창립), '평화와 자유를 위한 여성 국제연맹'(Women's International League for Peace and Freedom, 1915년 창립)가 참여했다. 이 가운데 '국제장애위원회'와 '국제인권감시기구'는 ICBL의 창립회원이고, '퍼그워시회의'를 제외한 다른 단체가 모두 ICBL의 회원단체이다. 핵군축에 초점을 맞추던 '퍼그워시회의'를 제외하면, 사실상 ICBL의 연결망이 AI와 로봇 영역으로 옮겨온 셈이다.

'킬러로봇 중단 캠페인'은 AI의 무기화를 국제사회의 의제로 투영하는데 성공했다. 남아프리카공화국에서 인권운동에 관여하던 크리스토프 헤인즈 교수(Christof Heyns, 남아공 프리토리아 대학교 교수)가 2010년 '초법적, 약식 또는 자의적 처형에 관한 특별보고관'(Special Rapporteur on extrajudicial, summary or arbitrary executions)으로 임명되었는데, 인권 연결망을 통하여 AI의 무기화에 관심을 가지게 되었다. 2013년 그는 AI의 무기화에 관한 보고서를 국제연합 인권이사회에 제출했는데, AI에 관한 인도적 군축운동의 용어를 그대로 수용하지 않았지만 (Heyes 2013),¹³⁾ AI의 무기화 가능성을 제기했고 AI의 무기화를 중단할 것을 촉구했다. AI의 무기화가 국제연합 인권이사회에서 화두가 되었다.

2013년 11월 AI의 무기화가 국제사회의 군축의제가 되었다. '특정 재래식무기 금지협약(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 협약)'의 당사국 회의에서 AI의 무기화가 논의되었다.¹⁴⁾ ICRC가 해당 회의에 초청되어 의견 발표를 했고, 당사국 대표와 질의응답시간을 가졌다. CCW 협약 당사국 회의는 2014년 비공식 전문가 회의를 열기로 했다. 이후 매년 비공식 전문가 회의가 열리며, CCW 협약 당사국 회의에서 AI의 무기화가 논의된다.

인도적 군축운동은 AI의 무기화를 국제사회의 의제로 투정한 이후, 우호단체를 모으고 유명인, 정치인을 대상으로 지지 세력을 넓혔다. 2018년 말까지 3000명의 유명인과 학자, 200개 과학기술 관련 기업의 지지를 확보했다. 동시에 AI의 무기화에 반대하는 입장을 가진 국가의 공개적 성명을 이끌었다.¹⁵⁾

13) 이 보고서는 '살인 로봇'이 '감정을 자극하는' 용어로 비난받을 수 있다고 하며 이를 직접 사용하지 않았지만, 치명적 자율무기를 배치하는 행위가 사람을 해충으로 대우하는 행위로 묘사될 수 있다고 하였다 (Heyes 2013, para.95). 상기 표현은 "치명적 자율무기의 배치가 사람에게 귀뿔을 놓는 것과 같으며, 그런 행위는 적을 해충으로 대우하는 것"이라는 문구에서 나왔다(Sparrow 2011, 125).
14) CCW 협약은 전쟁 목적에 비하여 과도한 상해를 초래하거나 민간인에게도 피해를 끼치는 무기를 금지하거나 규제하기 위한 국가 합의로 1983년 발효되었다. 현재 국제사회에서 CCW 당사국 회의는 가장 큰 군축회의이다.
15) 현재까지 총 28개국(파키스탄, 에콰도르, 이집트, 로마 교황청, 쿠바, 가나, 볼리비아, 팔레스타인, 짐바브웨, 알제리, 코스타리카, 멕시코, 칠레, 나카라과아, 파나마, 페루, 아르헨티나, 베네수엘라, 과테말



2018년 과학기술계 지도자의 모임인 Web Summit에 참석한 구테라스 UN 사무총장 (Secretary-General António Guterres)은 아래와 같이 AI 무기화 운동에 힘을 보탤다.

인공지능의 무기화는 중대한 위협이다. 자율적으로 타격 대상을 선정하고 타격할 수 있는 능력을 갖춘 기계로 인하여 확산 방지, 전장에서 국제인도법과 인권법의 준수가 어려워졌거나 미래에 어려워질 것이다. (이 쟁점이) 나에게서는 매우 명확하다. 사람의 목숨을 거둘 수 있는 능력과 재량을 가진 무기는 정치적 수용 불가능하며, 도덕적으로 역겹다. 이런 무기는 국제법을 통하여 금지되어야 한다(Guterres 2018).

AI 무기화에 반대하는 담론은 현재 어디에 있는가? 국제규범의 생애주기 개념에 비추어 보면,¹⁶⁾ 증폭 단계의 초입에 있다. AI의 무기화를 실제 구현하는 기술적 능력을 갖추지 못하며 강대국과 동맹관계를 맺지 않는 중소국이 이미 AI의 무기화에 반대하는 입장을 밝혔다. 즉, AI 무기화 반대 담론이 국제사회의 일부에서 수용된 셈이다. 이 담론이 더 많은 국가의 지지를 얻어 확산될지 또는 무산될지 갈림길에 있다.

V. 나아가며

AI 기반 자율무기를 금지하려는 사회운동은 1990년대 대인지뢰금지운동과 2000년대 집속탄 금지운동의 경험을 AI 영역에 투영하고 있다. 두 무기체계의 금지를 이끌어낸 인도적 군축운동은 2007-2009년 새로운 군축운동의 대상을 탐색하는 과정에서 핵무기, 인구밀집지역을 공격하는 무기, AI 기반 자율무기를 후보로 선정하면서 분화되었다. 앞서 두 무기체계의 금지운동이 성공을 거두는 사이 고양된 에너지가 세 분야로 분화된 것이다. 비정부기구 '36조'가 세 영역에 모두 관여하는 현상에서 보여지듯이 인도적 군축운동 사이에 연결점이 있지만, 세 영역 간 분화도 기정사실이 되었다.

AI 무기화에 반대하는 군축운동세력은 비정부기구 '36조'의 중개로 AI와 로봇 전문가 집단과 만나게 되었다. 군축운동세력은 AI와 로봇 전문가 집단이 비정부기구로 변모하는데 도움을 주고, ICRC의 도움을 얻어 군축운동의 신빙성을 높이는데 성공했다. 반면, ICRC가 느슨한 전문가 모임에서 비정부기구로 변모하는 과정에서 내적 통일성이 높아졌지만, 연구에 초점을 맞추는 집단과 사회운동에 초점을 맞추는 집단 사이의 차이로 인하여 무력하게 되었다. 전문가 집단이 사

라, 브라질, 이라크, 우간다, 오스트리아, 중국, 지부티, 콜롬비아, 엘살바도르, 모로코)이 지지 의사를 밝혔다.

16) 국제규범은 출현(emergence), 증폭(cascade), 그리고 내화(internationalization) 등 세 단계의 생애 주기를 갖는다. 출현 단계는 특정 담론이 사회적 동의를 얻어 규범으로 바뀌는 과정, 증폭 단계는 특정 담론이 규범으로 인정을 받아 급속하게 확산되는 과정, 내화 단계는 특정 담론이 당연한 행동 기준으로 인정을 받는 단계이다(Finnmore and Sikkink 1998, 894-905).



회운동세력과 만나면서 전문가 집단의 장점이 사라졌다.

인도적 군축운동세력은 시의 무기화와 로봇을 엮어 ‘살인 로봇’이라는 용어를 고안함으로써 대중의 관심을 얻는데 성공했다. 아직 구현되지 않는 ‘살인 로봇’이라는 용어를 사용함으로써 긴박성을 다소 떨어뜨리는 효과를 피할 수 없었지만, 대중에게 익숙하지 않는 AI 기반 자율무기를 대중에게 익숙한 로봇으로 소개함으로써 대중에게 다가갈 수 있었다. 시의 무기화를 ‘살인 로봇’으로 의제화하는데 성공했다.

더 나아가 인도적 군축운동세력은 시의 무기화 반대 운동에 국가를 끌어들이고 있다. 기술적 능력의 한계로 인하여 시의 무기화를 할 수 없으며 강대국으로부터 거리를 두는 중소국이 인도적 군축운동세력에 동조한다. 이념적 동조라기보다는 강대국과의 군사적 격차를 더 이상 확대하지 않으려는 현실적 고려가 더 강하게 보인다. 반면, 기술적 능력을 가지거나 강대국과 밀접한 관계를 유지하는 국가는 조심스러운 행보를 보인다. 자국이 직접 시의 무기화를 구현할 수 있는 가능성, AI 무기화 금지가 과도한 규제로 이어질 수 있는 가능성, 강대국의 반발을 초래할 수 있는 가능성 등을 동시에 고려하는 듯 보인다.

AI 무기화 반대운동의 향방이 어떻게 될까? 현재 인도적 군축운동은 대인지뢰금지운동과 집속탄금지운동의 경험을 AI 무기화에 적용하려 한다. CCW 협약의 당사국 회의를 활용하고 있지만, 자율무기를 구현할 수 있는 기술적 능력을 갖춘 국가가 AI 무기화 반대운동에 쉽게 동의하지 않는다. 시에 대한 과도한 규제로 이어질 수 있는 위험이 있기 때문이다. 따라서 국가차원의 협의체에서 다수 국가의 동의를 얻을 시의 무기의 금지로 이어질 가능성이 낮다.

CCW 협약 당사국 회의가 인도적 군축세력에게 우호적인 장소가 되지 않는다면, 인도적 군축운동이 CCW 당사국 회의를 우회하고 직접 국제사회의 구성원과 함께 시의 무기화를 금지하는 협약 초안을 만들고 국가를 압박하는 현상을 보게 될 수도 있다.¹⁷⁾ 인도적 군축운동이 저개발 비동맹 중소국과 연합하여 집속탄금지규범을 성문화시키는 현상이 다시 나타날 수 있다. 일부 강대국이 시의 무기화에 반대하는 금기를 넘어 AI 기반 자율무기를 개발하고 활용하는 결과가 나올 수도 있다. 시의 무기화로 인한 윤리적 비난보다 안보적 고려를 더 중시할 수도 있기 때문이다. 이 경우 핵무기확산금지조약처럼 일부 강대국만 AI 기반 자율무기를 가진 후 진입장벽을 세울 수 있다.

인도적 군축운동에게 최악 시나리오는 시의 무기화가 허용되는 현상이다. 이렇게 되면, 공격하는 쪽에서 전쟁으로 인한 전투원의 피해를 피하면서 은밀하게 무력행위를 진행할 수 있다고 믿게 될 수 있다. 당연히 전쟁의 임계점이 낮아질 위험이 커진다. 더 나아가 모든 국가가 AI 무기로 무장하는 또 한 차례 군비경쟁이 일어날 수 있다.

17) 엑스워시 장관은 대인지뢰금지규범을 확산시키기 위한 공간을 CCW 평가회의가 아니라 오타와 프로세스로 바꾼 일을 “우리는 미뉴엣(minuet)을 추지 않기로 했습니다. 사람들을 모아 회합을 열었습니다.”라고 표현했다. CCW 평가회의를 보폭이 좁고 우아한 몸짓을 특징으로 하는 궁정 춤곡인 미뉴엣에 비유한 반면, 오타와 프로세스는 사람의 회합에 비유한 것이다(Noma 2005, 23).



참고문헌

- 김민혁·김재오. 2020. "자율살상무기체계에 대한 국제적 쟁점과 선제적 대응방향." 국방연구 63(1):171-203.
- 김종열. 2018. "자율 무기체계의 자율성에 대한 연구." 융합보안 논문지 18(2):101-111.
- 연합뉴스. "한화시스템·KAIST, AI 기술 탑재 첨단무기 공동연구." 연합뉴스(2018.2.20).
- 조동준. 2011. "정치 환경의 변화와 대인지뢰금지규범의 확산: 대인지뢰에 대한 의제화(framing)와 한국대인지뢰대책회의의 활동을 중심으로." 평화학연구 12(1):27-57.
- Ackerman, Evan. "U.S. Army Considers Replacing Thousands of Soldiers With Robots." IEEE Spectrum(2014.1.22).
- Bolton, Matthew Breay, Sarah Njeri and Taylor Benjamin-Britton, eds. 2020. Global Activism and Humanitarian Disarmament. Gewerbstrasse, Switzerland: Palgrave Macmillian.
- Bolton, Matthew Breay and Cayman C. Mitchell. 2020. "When Scientists Become Activists: The International Committee for Robot Arms Control and Politics of Killer Robots." in Matthew Breay Bolton, Sarah Njeri and Taylor Benjamin-Britton, eds. 2020. Global Activism and Humanitarian Disarmament. Gewerbstrasse, Switzerland: Palgrave Macmillian, 27-58.
- Carpenter, Charli. 2014. Lost Causes: Agenda Vetting in Global Issue Networks and the Shaping of Human Security. Ithaca, NY: Cornell University Press.
- Etzioni, Amitai and Oren Etzioni. 2017. "Pros and Cons of Autonomous Weapons Systems." Military Review (May-June 2017).
- Finnmore, Martha and Cathryn Sikkink. 1998, "International Norm Dynamics and Political Change", International Organization 52(4):887-917.
- Goose, Stephen D. and Mary Wareham. 2016. "The Growing International Movement Against Killer Robots. Harvard International Review 37(4):28-33.
- Group of Governmental Experts of the High Contracting Parties to CCW. "Categorizing lethal autonomous weapons systems – A technical and legal perspective to understanding LAWS" (CCW/GGE.2/2018/WP.2, 2018.8.24.).
- International Committee of the Red Cross. Ethics and Autonomous Weapon Systems: An Ethical Basis for Human Control? (2018.4.3).
- Jones, Kim. "Special Weapons Observation Remote recon Direct Action System (SWORDS)." North Atlantic Treaty Organization Science and Technology Organization (NATO STO) MP-AVT-146-36(2007).



- Jun, Ji-hye. "Hanwha, KAIST to Develop AI Weapons." Korea Times(2018.2.25).
- Mero, Theodor. 2017. "The Martens Clause, Principles of Humanity, and Dictates of Public Conscience." American Journal of International Law 94(1):78-89
- Noma, Emiko. ed. 2005. "Distinguished Lecture Series, The Honorable Lloyd Axworthy", Joan B. Kroc Institute for Peace & Justice,
- Sharkey, Amanda. 2019. "Autonomous Weapons Systems, Killer Robots and Human Dignity." Informational Technology 21:75-87.
- Sharkey, Noel. 2010. "Saying 'No!' to Lethal Autonomous Targeting." Journal of Military Ethics 9(4):369-383.
- Singer, Peter. 2009. Wired for War. London, UK: Penguin Group.
- Sohn, Ji-young. "AI experts call off boycott of KAIST over alleged development of 'killer robots'." Korea Herald (2018.4.9.).
- Sparrow, Robert. 2011. "Robotic Weapons and the Future of War." in Jessica Wolfendale and Paolo Tripodi, eds. New Wars and New Soldiers: Military Ethics in the Contemporary World. Burlington, VA: Ashgate, 117-133.
- Ticehurst, Rupert. 1997. "The Martens Clause and the Laws of Armed Conflict." International Review of the Red Cross 317 (1997.4.30).
- Tiernan, Trevor. "Reaper drops first precision-guided bomb, protects ground forces." U.S. Air Forces Central Command(2007.11.9).
- Walsh, Toby et al. "Open Letter to Professor Sung-Chul Shin, President of KAIST from some leading AI researchers in 30 different countries." (March 2018).
- Wolfendale, Jessica and Paolo Tripodi, eds. 2011. New Wars and New Soldiers: Military Ethics in the Contemporary World. Burlington, VA: Ashgate.
- Young, Kevin L. and Charli Carpenter. 2018. "Does Science Fiction Affect Political Fact? Yes and No: A Survey Experiment on "Killer Robots." International Studies Quarterly 62(3):562-576.