

자율무기체계와 인공지능 기반 의사결정지원체계의 군사 표적 선정

비교 분석 및 정책 권고

알렉산더 블랜차드, 로라 브룬

스톡홀름 국제평화연구소(SIPRI)

SIPRI는 분쟁, 군비, 군비통제 및 군축 연구를 전문으로 하는 독립 국제기구입니다.

1966년에 설립된 SIPRI는 정책 입안자, 연구자, 언론 및 일반 대중에게

공개 출처 기반의 데이터, 분석 및 권고를 제공합니다.

운영위원회는 연구소 발행물에 표현된 견해에 대해 책임을 지지 않습니다.

이사회

- 스테판 뢰벤, 의장(스웨덴)
- 모하메드 이븐 참바스 박사(가나)
- 찬 행 치 대사(싱가포르)
- 노하 엘-미카워 박사(이집트)
- 장마리 게헨노(프랑스)
- 라다 쿠마르 박사(인도)
- 패트리샤 루이스 박사(아일랜드/영국)
- 제시카 투크만 매튜스 박사(미국)

이사

댄 스미스(영국)

피스모모(PEACEMOMO)

피스모모는 올해로 13년째 평화와 배움, 평화와 일상을 연결함으로써

평화를 모두의 것으로 만들어가고자 활동하고 있는 비영리 사단법인입니다.

2017년부터 SIPRI 연감 요약본을 한국어로 번역하고 있습니다.

*이 번역본에 포함된 정보의 정확성과 관련하여 궁금한 점이 있는 경우,

공식 영문 보고서를 참조하시기 바랍니다.



STOCKHOLM INTERNATIONAL PEACE RESEARCH INSTITUTE

Signalistgatan 9

SE-169 70 Solna, Sweden

Telephone: +46 8 655 97 00

Email: sipri@sipri.org

Internet: www.sipri.org



PEACEMOMO

피스모모

서울특별시 서대문구 연희로 41다길 48-12, 1층

전화: 02-6351-0904

이메일: office@peacemomo.org

홈페이지: www.peacemomo.org

자율무기체계와 인공지능 기반 의사결정지원체계의 군사 표적 선정

비교 분석 및 정책 권고

알렉산더 블랜차드, 로라 브룬

발행: 2025년 6월

번역: 2025년 11월



STOCKHOLM INTERNATIONAL
PEACE RESEARCH INSTITUTE

© SIPRI 2025

<https://doi.org/10.55163/YQBY3151>

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted, in any form or by any means, without the prior permission in writing of SIPRI or as expressly permitted by law.

목차

감사의 글	1
요약	2
1장. 서론	4
본 보고서의 범위	5
개요	6
2장. 특성분석	7
자율무기체계와 인공지능 기반 의사결정지원체계의 기능	7
무력 사용 결정에 미치는 영향	9
표적 선정 주기의 범위	9
표적 결정 실행에서의 역할	10
정책적 함의	11
그림 2.1. 표적 선정 주기의 각 단계에서 자율무기체계와 인공지능 기반 의사결정지원체계의 범위	8
3장. 의도하지 않은 피해 위험	13
신뢰성 문제	13
인간-기계 상호작용의 차이	14
맥락적 변수	16
정책적 함의	17
4장. 법적 측면	19
국제인도법이 적대행위 수행에 있어 인간에게 요구하는 것과 기계에게 허용하는 것	19
법적 책임과 책무성	20
법적 평가에서의 역할	21
제36조 검토에 대한 함의	22
정책적 함의	22
5장. 정책 대응	24
접근 방식 1: 자율무기체계 관련 기존 다자간 논의에 인공지능 기반 의사결정지원체계를 구체적으로 포함시키기	24
접근 방식 2: 인공지능 기반 의사결정지원체계 전담 프로세스 구축	27
접근 방식 3: 인공지능 기반 의사결정지원체계에 대한 별도 접근 방식 취하지 않기	28
6장. 주요 결과 및 권고사항	29
주요 결과	29
권고사항	31
표 6.1. 4가지 주요 영역에서 비교된 자율무기체계(AWS)와 인공지능 기반 의사결정지원체계(AI-DSS)	32
그림 6.1. 인간-기계 상호작용, 위험 경로 및 무력 사용에서 인공지능 기반 의사결정지원체계와 자율무기체계 간의 차이점	30
저자소개	34

감사의 글

SIPRI와 저자들은 이 간행물에 대한 재정 지원을 아끼지 않은 네덜란드, 스웨덴, 스위스 외교부에 진심으로 감사드립니다. 또한 이 보고서 집필 과정에서 귀중한 통찰력과 지원을 제공해 준 SIPRI 동료인 마르타 보, 뱅상 블라냉, 네타 구삭, 더스틴 루이스, 쥘 팔레이어에게 감사드립니다. 아울러 안나 앤더슨, 루퍼트 바렛 테일러, 네할 부타, 잉빌드 보데, 제시카 도르시, 마틴 하그스트뤼프, 샬리 롱워스, 아서 홀랜드 미셸, 안나 나디바이제가 프로젝트 전반에 걸쳐 제공한 의견, 토론, 피드백에도 감사드립니다. 마지막으로 SIPRI 편집부의 면밀한 편집 작업에도 감사드립니다.

요약

자율무기체계(AWS)가 제기하는 인도주의적 우려 및 법적 우려는 오랫동안 국제 정책 논의의 주제였으며, 최근에는 군사 분야 인공지능(AI) 도입 논의에서도 다뤄지고 있습니다. 인공지능 기반 의사결정지원체계(AI-DSS)의 군사적 활용에 대한 관심이 높아지면서, 이러한 체계를 국제 정책 논의에서 어떻게 다루어져야 하는지 고려할 필요성이 대두되고 있습니다. 본 보고서는 군사 표적 선정에 사용되는 자율무기체계와 인공지능 기반 의사결정지원체계를 각각의 특성과 의도하지 않은 피해 위험, 법적 측면 및 정책 대응 관점에서 비교합니다. 또한 이 문제에 대한 정책 입안자들의 판단을 돕기 위해 여러 주요 연구 결과를 제시하고 있습니다.

군사 표적 선정에 사용되는 자율무기체계와 인공지능 기반 의사결정지원체계는 모두 표적 선정 결정에서 인간의 역할에 영향을 미칩니다. 그러나 주요 차이점은 표적 선정 주기에서의 활용 범위입니다. 자율무기체계는 임무 실행 단계에 국한되는 반면, 인공지능 기반 의사결정지원체계는 여러 단계에 걸쳐 더 광범위하게 사용됩니다. 다만 두 체계의 실제 배치 방식에 따라 구분이 모호해질 수 있습니다.

자율무기체계와 인공지능 기반 의사결정지원체계 모두 의도하지 않은 피해 위험을 수반하지만, 이러한 위험은 서로 다른 방식으로 발현됩니다. 두 체계 모두 자율성과 인공지능의 알려진 기술적 한계 때문에 발생하는 신뢰성 문제를 공유하지만, 인간 기계 상호작용 방식이 다르기 때문에 서로 다른 결과를 낳을 수 있습니다. 표적 식별과 교전 사이에 직접적인 경로가 있는 자율무기체계의 경우 위험이 직접적입니다. 예를 들어 잘못된 표적 식별은 인간의 개입 없이 즉각적인 치명적 행동으로 이어질 수 있습니다. 인간에게 결과를 제공하는 인공지능 기반 의사결정지원체계의 경우 위험이 간접적입니다. 즉, 인간이 잘못된 표적 식별에 따라 행동할 경우 피해가 발생합니다.

자율무기체계와 인공지능 기반 의사결정지원체계는 모두 사용자가 국제인도법(IHL) 의무 이행을 위해 이러한 체계에 얼마나 의존할 수 있는지, 그리고 책임과 책무성을 어떻게 보장할 것인지에 대한 의문을 제기합니다. 그러나 두 체계는 별개의 법적 과제를 야기합니다. 자율무기체계의 경우 자율적인 무력 행사에 대한 우려와 사용자가 체계의 영향을 합리적으로 예측하고 통제할 수 있는지에 대한 우려가 제기됩니다. 인공지능 기반 의사결정지원체계의 경우 자동화 편향(Automation bias) 등으로 인해 인간이 법적 판단을 내릴 때 인공지능에 과도하게 의존할 수 있다는 우려가 제기되며, 이는 잠재적으로 인간이 체계의 권고사항에 수동적으로 동의하는 상황으로 이어질 수 있습니다. 인공지능 기반 의사결정지원체계의 활용은 국제인도법을 넘어 국제인권법과 같은 법적 틀에도 시사하는 바가 있습니다.

두 체계의 비교를 통해 군사 분야 인공지능과 관련된 다자간 논의를 진행하고 있는 정책 입안자들에게 세 가지 접근 방식을 제안할 수 있습니다. (a) 자율무기체계 관련 다자간 노력에 인공지능 기반 의사결정지원체계를 구체적으로 포함시키기, (b) 새로운 인

3 자율무기체계와 인공지능 기반 의사결정지원체계의 군사 표적 선정

공지능 기반 의사결정지원체계 전담 프로세스를 구축하기, (c) 인공지능 기반 의사결정지원체계에 대한 별도 접근 방식을 취하지 않기. 각 방안에는 정책 입안자들이 고려해야 할 특정 함의와 득실이 있습니다. 자율무기체계와 인공지능 기반 의사결정지원체계 간의 실질적인 유사점과 차이점을 볼 때 이러한 접근 방식 중 하나가 적합함을 알 수 있습니다. 그러나 궁극적으로 채택되는 접근 방식은 정치적, 제도적 필요를 반영해야 합니다.

이러한 결과를 바탕으로 본 보고서는 세 가지 권고사항을 담았습니다. 첫째, 각국은 다양한 접근 방식 간의 득실을 인지하고, 인공지능 기반 의사결정지원체계를 위한 전담 다자간 프로세스의 필요성을 고려해야 합니다. 둘째, 인공지능 기반 의사결정지원체계에 대한 향후 정책 노력은 자율무기체계 거버넌스, 특히 인간-기계 상호작용 및 감독, 법률 준수 및 책무성 문제와 관련하여 얻은 통찰을 바탕으로 해야 합니다. 셋째, 인공지능 기반 의사결정지원체계는 자율무기체계 논의에서 누락되었거나 충분히 탐구되지 않은 별도의 문제를 제기하므로, 정책 입안자는 주요 지식 격차를 파악하고 군사적 의사결정에 인공지능을 책임감 있게 통합하며, 무력 사용 결정에서 인간의 주체성을 지원하기 위한 연구를 추진해야 합니다.

1장. 서론

인간의 개입 없이 표적을 선택하고 무력을 행사할 수 있는 자율무기체계(AWS)는 무력 사용에서 인간의 역할을 바꾸고 다양한 인도주의적, 법적, 윤리적 문제를 제기합니다.¹ 이에 따라 자율무기체계는 국가적, 지역적, 국제적 차원에서 지속적인 정책 논쟁의 주제가 되었습니다.² 이 논쟁의 가장 중요한 국제 포럼은 유엔으로, 제네바에서 개최되는 특정재래식무기금지협약(CCW) 당사국회의 주관 하에 각국은 자율무기체계에 대한 적절한 거버넌스 대응 방안을 논의하고 있습니다.³ 자율무기체계는 유엔 총회 제1위원회, 인공지능의 책임 있는 군사적 이용에 관한 고위급회의(REAIM), 그리고 미국의 인공지능 및 자율성의 책임 있는 군사적 이용에 관한 정치선언 등 인공지능(AI)의 군사적 도입에 관한 국제 정책 논의가 급증하는 가운데 두드러지게 등장하고 있습니다.⁴ 인공지능은 자율무기체계를 가능하게 하는 여러 기술 중 하나이지만, 자율무기체계는 오랫동안 군사 분야에서 인공지능을 활용하는 명확하고 두드러지며 극적인 사례였습니다.

정책 입안자들은 군사 분야 인공지능 도입이 자율무기체계 관련 문제보다 훨씬 더 광범위하고 다면적인 사안이라는 점을 점차 인식하고 있습니다. 이는 이스라엘의 가자지구 점령과 러시아-우크라이나 전쟁 참전국을 포함한 현대 무력 분쟁에서 인공지능 기반 의사결정지원체계(AI-DSS)의 활용 사례가 보고되면서, 촉발된 관심의 결과이기도 합니다.⁵ 인공지능 기반 의사결정지원체계는 인공지능을 통합한 전산화 도구로, 표적 결정을 포함한 군사적 의사결정을 돕는 정보를 제공합니다.⁶ 이러한 체계는 상황 인식을 강화하는 등 군사 표적 선정을 개선하기 위한 의도였지만, 인공지능의 알려진 한계와 결

1 Boulanin, V. 및 Verbruggen, M., 무기 시스템의 자율성 개발 매핑(SIPRI: 스톡홀름, 2017년 11월), 24-27 쪽; 국제적십자위원회(ICRC), '자율무기체계에 대한 ICRC 입장', 2021년 5월 12일.

2 Blanchard, A. 등, '자율무기체계에 대한 정책 논쟁의 딜레마', SIPRI 논평, 2025년 2월 6일.

3 과도하게 해롭거나 무차별적인 효과를 갖는 것으로 간주될 수 있는 특정 재래식 무기의 사용에 대한 금지 또는 제한에 관한 협약(CCW 협약), 1981년 4월 10일 서명 개방, 1983년 12월 2일 발효.

4 각각 유엔 총회 결의안 79/239, 2024년 12월 24일 참조; REAIM, '행동을 위한 청사진', 2024년 9월 10일; 미국 국무부 군비통제·억제·안정국, '인공지능과 자율성의 책임 있는 군사적 사용에 대한 정치선언', 2023년 11월 9일.

5 AI Now Institute, '안전과 전쟁: 군사 AI 시스템의 안전 및 보안 보장', 2024년 6월 25일; Davies, H. 및 Abraham, Y., '공개: 이스라엘군이 방대한 팔레스타인 감시 데이터 모음을 사용하여 ChatGPT 유사 도구 개발', The Guardian, 2025년 3월 6일; Abraham, Y., "'라벤더': 가자에서 이스라엘의 폭격을 지휘하는 AI 기계', +972 Magazine, 2024년 4월 3일; 이스라엘방위군(IDF), 'IDF의 정보 처리에서의 데이터 기술 사용', 보도자료, 2024년 6월 18일; Bendett, S., '러시아-우크라이나 분쟁에서 AI의 역할과 함의', Russia Matters, 2023년 7월 20일; Farnell, R. 및 Coffey, K., '전쟁 계획에서 AI의 새로운 지평: AI 에이전트가 군사적 의사결정에 혁명을 일으킬 수 있는 방법', Belfer Center for Science and International Affairs, 2024년 10월 11일; Hunder, M., '우크라이나, AI 모델 훈련을 위해 방대한 전쟁 데이터 수집', Reuters, 2025년 12월 20일; Bondar, K., '우크라이나의 군사 AI 생태계 이해', Center for Strategic and International Studies(CSIS), 2024년; 및 Bergengruen, V., '기술 거물들이 우크라이나를 AI 전쟁 연구소로 만든 방법', Time Magazine, 2024년 2월 8일.

6 Boulanin, V., 'AI 기반 군사적 의사결정의 위험과 이점', 편집 R. Geiß 및 H. Lahmann, Research Handbook on Warfare and Artificial Intelligence(Edward Elgar: Cheltenham, 2024); Baxter, C. A., '인공지능을 이용한 전술적 수준 표적 선정 향상', Field Artillery, vol. 1(2024); Nadibaidze, A., Bode, I. 및 Zhang, Q., 군사적 의사결정지원체계의 AI: 개발 및 논쟁 검토(Center for War Studies: Odense, 2024년 11월); Holland Michel, A., Decisions, Decisions, Decisions, Decisions: Computation and Artificial Intelligence in Military Decision-making(ICRC: Geneva, 2024); Probasco, E. 외, '군사적 의사결정을 위한 AI: 장점 활용 및 위험 회피', 보안 및 신기술센터(CSET) 이슈 브리핑, 2025년 4월.

합되어 표적 결정을 형성하는 능력이 인도주의적, 법적, 윤리적 문제를 야기할 수 있음을 의미하기도 합니다.⁷

이는 다음과 같은 주요 정책 질문들을 제기합니다. 현재 자율무기체계에 집중된 군사 분야 인공지능 도입 및 자율성에 대한 글로벌 정책 논의에서 인공지능 기반 의사결정지원체계를 어떻게 다룰 수 있을까요? 즉, 인공지능 기반 의사결정지원체계와 자율무기체계를 같은 포럼과 절차에서 함께 다루야 할까요, 아니면 별도의 포럼이나 절차에서 따로 논의해야 할까요? 정책 입안자들이 이 질문을 검토할 수 있도록 본 보고서는 자율무기체계와 인공지능 기반 의사결정지원체계를 그 특성, 의도하지 않은 피해 위험, 그리고 법적 측면에서 비교합니다. 또한 인공지능 기반 의사결정지원체계와 자율무기체계가 유사하거나 상이한 기술적, 법적, 정책적 과제를 제기하는지, 그리고 제기한다면 어느 정도까지 그러한지를 검토합니다. 본 보고서는 두 가지 측면에 기여하고자 합니다. 첫째, 자율무기체계 관련 다자간 프로세스에 참여하는 정책 입안자들에게, 해당 논의에 인공지능 기반 의사결정지원체계를 포함하는 것이 적절한지에 대한 근거를 제공합니다. 둘째, 군사 분야 인공지능 거버넌스에 대한 국제 정책 이니셔티브에 참여하는 정책 입안자들에게, 인공지능 기반 의사결정지원체계를 다루기 위한 유용한 접근 방식을 제시합니다. 여기에는 자율무기체계 관련 글로벌 정책 논의에서 얻을 수 있는 교훈도 포함됩니다.

본 보고서의 범위

인공지능 기반 의사결정지원체계는 공급망 물류 및 유지보수, 분쟁 예측, 위게임(전쟁 모의훈련) 등 다양한 군사적 목적으로 활용될 수 있지만, 표적 선정을 위한 사용은 인공지능의 고위험 군사적 적용으로 간주됩니다.⁸ 따라서 본 보고서는 ‘군사 표적 선정’에 사용되는 인공지능 기반 의사결정지원체계에 초점을 맞춥니다. ‘군사 표적 선정(military targeting)’이란 ‘전쟁 당사자(국가군대 또는 조직된 무장단체)가 통제할 수 없는 개인이나 대상에 대해 무력을 사용하는 것’으로, 공격 또는 방어 시 의도적이고 동적인 무력 행사를 포함하는 개념으로 널리 이해됩니다.⁹

본 보고서는 자율무기체계 및 인공지능 기반 의사결정지원체계와 관련된 모든 문제를 포괄적으로 다루지는 않습니다. 오히려 자율무기체계에 대해 국제 정책 논의에서 큰 관심을 받고, 기존 정책 전문성이 축적된 분야와 쟁점에 초점을 맞춰, 두 기술 간의 보다 실질적인 비교를 촉진하는 것을 목표로 합니다. 인도주의적 위험, 국제인도법(IHL), 인간-기계 상호작용 등의 영역이 이에 해당합니다. 본 보고서에서 다루지 않았지만, 자

7 ICRC, ‘군사 분야에서의 인공지능에 관한 유엔 사무총장 제출서’, 2025년 4월 17일.

8 Boulanin, ‘AI 기반 군사적 의사결정의 위험과 이점’(각주 6)

9 ICRC, ‘국제인도법에 따른 표적 선정’, 법은 전쟁에서 어떻게 보호하는가?, ICRC 온라인 사례집, [n.d.]; Duchaine, P. A. L., Schmitt, M. N. 및 Osinga, F. P. B., ‘서론’, Duchaine, Schmitt 및 Osinga 편집, Targeting: The Challenges of Modern Warfare(T. M. C. Asser Press: 헤이그, 2016).

율무기체계와 인공지능 기반 의사결정지원체계 모두와 관련해 더 많은 주의를 기울여야 할 문제들도 있습니다(예: 국제인권법, 안보, 윤리 등).

본 보고서는 기존 SIPRI 연구, 학술 및 정책 문헌 검토, 그리고 엄선된 전문가들과의 협의를 바탕으로 작성되었습니다.

개요

인공지능 기반 의사결정지원체계와 자율무기체계가 공동 정책 접근 방식을 필요로 하는지, 아니면 별개의 정책 접근 방식을 필요로 하는지를 이해하기 위한 토대를 마련하기 위해 세 장에 걸쳐 비교 분석하였습니다. 2장에서는 자율무기체계와 인공지능 기반 의사결정지원체계를 정의하고, 각 기술의 특성이 표적 결정에서 어떻게 작동하는지 비교합니다. 3장에서는 군사 표적 선정에 자율무기체계와 인공지능 기반 의사결정지원체계를 사용할 때 의도하지 않은 피해 위험이 어떻게 나타나는지를 살펴봅니다. 4장에서는 두 체계가 국제인도법 준수와 관련하여 유사한 과제를 제기하는지, 아니면 다른 과제를 야기하는지 등 법적 측면을 비교합니다. 5장에서는 이러한 비교 분석을 토대로, 두 체계가 동일한 유형의 위험 완화 조치를 요구하는지 또는 별도의 접근이 필요한지를 분석하고, 이를 바탕으로 세 가지 정책 방안을 제시합니다. 마지막으로 6장에서는 주요 연구 결과와 권고사항을 제시합니다.

2장. 특성분석

이 장에서는 자율무기체계와 인공지능 기반 의사결정지원체계의 특징을 군사 표적 선정 주기에서의 기능 및 역할 측면에서 비교합니다. 표적 선정 주기 모델은 군사 교리에 따라 상이합니다. 본 장(그림 2.1)에서 사용된 모델은 공개적으로 이용 가능한 교리들을 요약한 것으로, 네 가지 주요 단계로 구성됩니다. (1) 표적 선정 노력의 포괄적인 목표와 세부 목적의 수립, (2) 표적 식별과 표적 목록 작성 및 가장 적합한 표적 공격 수단 결정, (3) 표적 인물 또는 대상을 식별하고 공격하는 임무 실행, (4) 공격의 효과성 평가.¹⁰ 실제로 이러한 단계는 순차적으로만 이루어지지 않고 반복적·양방향적이며, 때로는 동시에 진행됩니다.¹¹ 즉, 표적에 대한 무력 사용은 3단계인 임무 실행에서 발생하지만, 무력 사용 결정은 1, 2, 3단계에서 내려질 수 있습니다.

자율무기체계와 인공지능 기반 의사결정지원체계의 기능

자율무기체계는 일단 활성화되면 인간의 개입 없이 표적을 식별하고 선택하며 무력을 적용할 수 있는 무기입니다.¹² 자율무기체계는 사전 프로그래밍된 표적 프로필과 센서 및 소프트웨어를 통해 ‘인식’한 기술 지표에 따라 작동됩니다. 인간 운용자는 자율무기 체계를 감독하거나 재정의할 수 있는 선택권을 유지할 수 있으나, 그 작동에는 인간의 개입이 필수적이지 않습니다. 자율무기체계는 작동을 위해 반드시 인공지능으로 뒷받침될 필요는 없습니다(예: 지뢰는 때때로 자율무기체계로 분류됩니다). 실제로 자율무기체계의 국제 거버넌스를 논의하는 주요 다자간 포럼인 유엔 특정재래식무기금지협약 산하 치명적 자율무기체계(LAWS) 분야 신기술 정부전문가그룹(GGE)은 이러한 체계 논의에 기술적으로 중립적인 입장을 취하며, 인공지능을 자율무기체계의 필수 구성 요소로 보지 않았습니다.¹³ 그러나 자율무기체계 국제 정책 논의는 주로 머신러닝(ML)과 딥러닝(DL) 등 인공지능 기술의 발전과 깊이 연관되어 있습니다. 이러한 발전은 자율무기체계가 역동적이고 비정형적이며 개방적인 환경에서 작동하고, 개인을 표적으로 삼을 수 있게 만들 가능성을 높입니다.¹⁴

10 예를 들어, 미 공군, “Targeting”, 공군 교리 출판물 3-60, 2021년 11월 12일, 7쪽; 북대서양조약기구(NATO) 표준화 사무국, ‘공동 표적 선정을 위한 연합군 공동 교리’, NATO 표준 AJP-3.9 B판 1, 2021년 11월, 1.5절; 호주 방위군 전쟁 센터, “Targeting”, 작전 시리즈 ADDP 3.14, 2판, 2009년 2월 2일, 1.9절 및 4.8-4.9절을 참조하십시오.

11 Ekelhof, M. A., ‘Lifting the fog of targeting’, 해군전쟁대학 리뷰, 71권, 3호(2018), 66쪽 참조; 및 Nisser, J., ‘군사 교리 구현: 이론적 모델’, 비교전략, vol. 40, no. 3(2021).

12 Boulanin 및 Verbruggen(주 1), 24-27쪽; ICRC, ‘자율무기체계에 대한 ICRC 입장’(주 1); 및 Taddeo, M. 및 Blanchard, A., ‘자율무기체계 정의에 대한 비교 분석’, 과학 및 공학 윤리, vol. 28(2022).

13 예를 들어 유엔, CCW, 치명적 자율무기체계 분야의 신기술에 대한 정부 전문가 그룹(LAWS에 대한 GGE), 2023년 회기 보고서, CCW/GGE.1/2023/CRP.2, 2023년 5월 6일, para. 18 참조.

14 Sauer, F., ‘기계 학습 및 자율 시스템의 군사적 응용’, ed. V. Boulanin, 인공지능이 전략적 안정성과 핵 위험에 미치는 영향: 제1권 유럽-대서양 관점(SIPRI: 스톡홀름, 2019년 5월), 86-87쪽.

의사결정지원체계(DSS)는 의사결정을 지원하기 위해 정보를 제공하도록 설계된 전산화 도구입니다.¹⁵ 이를 통해 정보 필터링과 플래그 지정 등 관련 정보를 표시(해석)하고, 적군 이동과 같은 시나리오를 예측(예상)하며, 최적의 행동 방침이나 목표 지정하는 등의 행동을 권고(지시)합니다.¹⁶ 이러한 기능은 누적적이지 않으며, 의사결정지원체계는 이들 중 일부 혹은 여러 기능을 선택적으로 수행할 수 있습니다. 의사결정지원체계는 소위 전문가 시스템에서 사용되는 형식 논리에 기반한 규칙 기반(rules-based) 프로그래밍을 포함해 다양한 데이터 처리 모델을 통해 구축될 수 있습니다.¹⁷ 이러한 체계 자체는 새로운 개념이 아니지만, 머신러닝·딥러닝 그리고 여러 데이터 소스를 처리·집계할 수 있다는 대규모 언어 모델(LLM) 등 최신 인공지능 기술 사용이 보고되면서 군

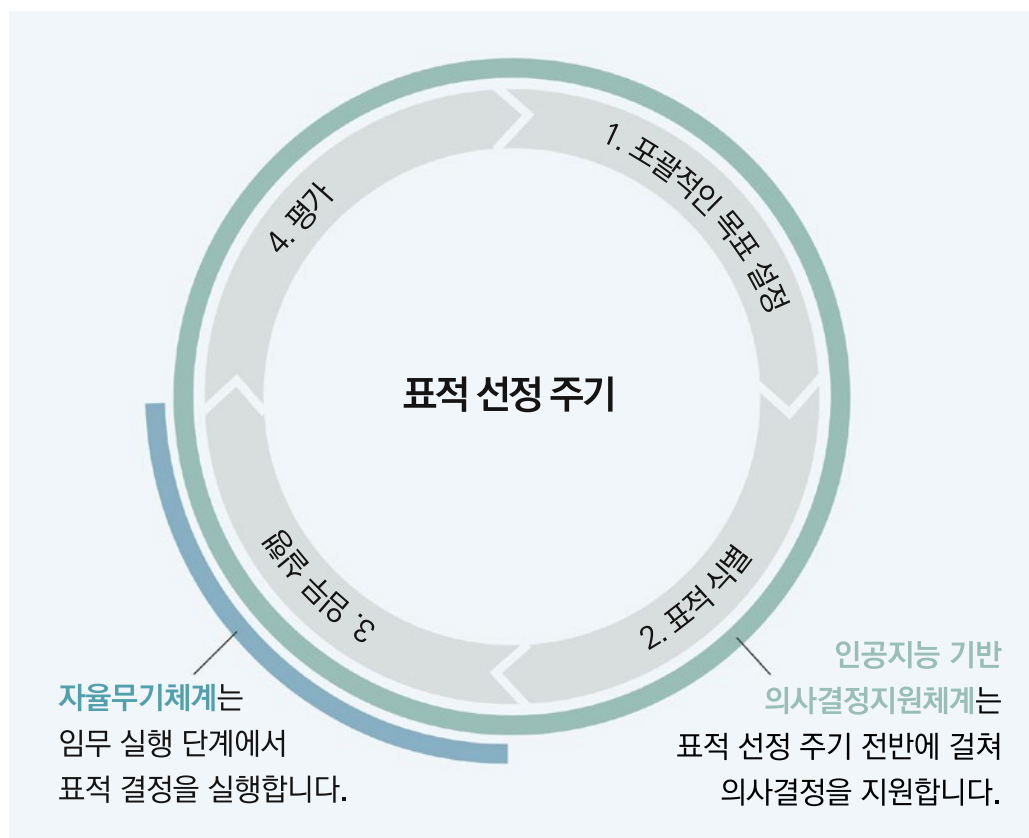


그림 2.1. 표적 선정 주기의 각 단계에서 자율무기체계와 인공지능 기반 의사결정지원체계의 범위

15 Bohanec, M., ‘의사결정지원’, D. Mladenici의 편집, 데이터 마이닝 및 의사결정지원: 통합과 협업(Springer Science+Business Media: 뉴욕, 2003); Şuşnea, E., ‘군사 행동에서의 의사결정지원체계: 필요성, 가능성 및 제약’, Journal of Defense Resources Management, 3권, 2호(2012); Holland Michel, ‘결정, 결정, 결정’(주 6), 13쪽; Phillips-Wren, G., ‘의사결정지원체계의 AI 도구: 리뷰’, International Journal on Artificial Intelligence Tools, 21권, 2호.

16 ICRC 및 제네바 아카데미, 무력 분쟁 시 무력 사용에 관한 군사적 의사결정에서 AI 및 관련 기술에 대한 전문가 자문 보고서(ICRC: 제네바, 2024년 3월), 9쪽; 및 Gadepally, V. N. 외, ‘국방부 및 정보기관을 위한 추천 시스템’, Lincoln Laboratory Journal, vol. 22, no. 1(2016).

17 ⁷ Şuşnea(각주 15); 및 Buchanan, B. G. 및 Smith, R. G., ‘전문가 시스템의 기초’, Annual Review of Computer Science, vol. 3, no. 1(1988).

9 자율무기체계와 인공지능 기반 의사결정지원체계의 군사 표적 선정

사적 활용 가능성을 둘러싼 국제 논쟁이 본격화되었습니다.¹⁸ 인공지능 기반 의사결정지원체계는 인공지능 분야의 현대 기술로 구현된 의사결정지원체계의 하위 범주입니다. 인공지능 기반 의사결정지원체계의 역할은 표적 선정 과정에서 공격 개시를 평가하는 데 필요한 정보를 구성·필터링·제시함으로써 사용자와 전장 사이의 관계 일부를 중재하는 것입니다.

무력 사용 결정에 미치는 영향

자율무기체계와 인공지능 기반 의사결정지원체계는 모두 무력 사용에서 인간의 역할에 영향을 미치며, 부분적으로는 무력 사용 결정 과정을 변화시킵니다. 무력 적용으로 이어지는 의사결정 과정은 단순하거나 선형적이지 않으며, 국방 조직 내 다양한 관행과 프로세스 전반에 걸쳐 여러 의사결정권자가 관여하는 복잡하고 상호의존적인 결정 과정입니다.¹⁹ 이러한 의사결정 과정에 기계 자율성을 도입하면 복잡성이 가중되고, 이에 따라 인간의 역할과 책임이 변화함으로써 의사결정 방식에도 영향을 미칩니다. 예를 들어, 자율무기체계와 인공지능 기반 의사결정지원체계 모두 설계 및 개발 단계를 포함하여 배치 이전의 생애 주기 단계에서 일반화된 표적 유형이나 모델과 관련된 의사결정을 요구할 수 있습니다. 그러나 인공지능 기반 의사결정지원체계는 무력 적용에 대한 인간의 의사결정을 직접 형성하고 영향을 미칠 수 있다는 점에서 차이가 있습니다. 예컨대 표적 우선순위를 설정하는 인공지능 기반 의사결정지원체계는 잠재적 표적의 성격과 사용자가 어디에 주목할지를 결정하는 데 영향을 미칩니다.

표적 선정 주기의 범위

자율무기체계와 인공지능 기반 의사결정지원체계의 차이는 각각이 적용되는 표적 선정 주기의 범위에 있습니다. 자율무기체계는 표적 선정 주기의 임무 실행 단계(3)에서 제한된 기능, 즉 표적 식별, 선택 및 교전 기능을 수행하는 반면, 인공지능 기반 의사결정지원체계는 표적 선정 주기 전반에 걸쳐 보다 광범위하게 활용됩니다(그림 2.1).

1단계에서 고차원 작전 의도를 실질적으로 지원하기 위한 인공지능 기반 의사결정지원체계의 활용은 아직 현실화되지 않았습니다.²⁰ 그러나 위게임 등 전략적 수준의 인공지능 기반 의사결정지원체계 응용은 표적 선정에서 고차원 작전 목표 설정에 간접적으로 영향을 미칠 가능성이 있습니다.²¹ 2단계, 3단계, 4단계는 현재, 가까운 시점, 그리고

18 King, A., '디지털 표적 선정: 인공지능, 데이터 및 군사 정보', *Journal of Global Security Studies*, vol. 9, no. 2(2024년 6월).

19 Ekelhof, M., 'AI는 전장을 변화시키고 있지만, 아마도 당신이 생각하는 방식과는 다를 것이다: 표적법의 실용화와 군사 작전에서 AI 사용 증가에 대한 분석', Geiß 및 Lahmann 편집(각주 6).

20 Vestner, T., '전략에서 명령으로: 인공지능을 활용한 군사 작전 준비 및 수행', Geiß 및 Lahmann 편집(각주 6).

21 Knack, A., Balakrishnan, N. 및 Clancy, T., '전략적 경계에 AI 적용', *신흥 기술 및 보안 센터(CETAS) 연구 보고서*, 2025년 3월.

단기적 미래의 인공지능 기반 의사결정지원체계의 실제 활용 가능성이 더 높게 평가됩니다.

인공지능 기반 의사결정지원체계는 2단계에서 이미 표적 지정과 우선순위 결정에 사용되고 있는 것으로 알려져 있습니다. 한 가지 예로, 데이터 포인트 간 상관관계를 추론해 표적 목록을 생성하는 방식이 있습니다.²² 예를 들어, 특정 개인의 생활 패턴 데이터가 의심되는 표적 인물의 데이터와 상관관계를 보일 경우, 해당 개인이 잠재적 표적으로 지정될 수 있습니다. 또한 군사 교리와 국제인도법(IHL)을 포함한 데이터 세트를 기반으로 훈련된 인공지능 챗봇을 광고함으로써 군사 표적 파괴에 적합한 군수품 선택을 용이하게 한 미국 기술 스타트업의 사례도 보고되었습니다.²³ 일부 논평가들은 인공지능 기반 의사결정지원체계가 2단계에서 중요 인프라 매핑이나 부수적 피해 추정치를 강화해 민간인 피해를 한층 완화하는 데 기여할 수 있다고 제안합니다.²⁴

인공지능 기반 의사결정지원체계는 표적 선정 주기의 임무 실행 단계(3)에서도 활용될 수 있습니다. 여기에는 표적 선택과 추적을 지원하기 위한 인공지능 활용이 포함되며, 즉 표적의 지리적 위치를 고정하고 최신 위치를 파악하는 기능이 있습니다. 예를 들어, 이스라엘의 Where's Daddy?(‘아빠 어디 있어?') 체계는 휴대전화 위치 데이터를 사용하여 미리 지정된 표적이 특정 위치에 있을 때 운용자에게 알림을 전송하는 것으로 알려져 있습니다.²⁵ 또한 인공지능 기반 의사결정지원체계는 행동 방침 분석, 무기 체계의 가용성 및 적합성 표시 등 자원 최적화를 통해 임무 실행을 지원할 수 있습니다. 마지막으로, 인공지능 기반 의사결정지원체계는 전투 피해 평가 보고 및 사후 보고서를 포함하여 4단계에서 공격 사후 평가를 지원하는 데 사용될 수 있습니다.²⁶

표적 결정 실행에서의 역할

자율무기체계와 인공지능 기반 의사결정지원체계는 표적 선정 단계의 범위뿐 아니라 표적 결정 실행 단계에서도 서로 다른 역할을 합니다. 자율무기체계는 인간 운용자의 개입 없이 표적을 식별·선정·공격합니다. 즉, 자율무기체계는 사전에 구성된 일반화된 인간 의사결정을 기반으로 표적 결정을 실행합니다. 이러한 일반화된 결정에는 공격 과정에서 자율무기체계가 구체화하는 표적 유형이 포함될 수 있습니다. 반면, 인공지능 기반

22 NATO, ‘NATO, AI 기반 전투 시스템 도입’, 보도자료, 2025년 4월 14일.

23 Scale AI, ‘방위를 위한 미세 조정된 LLM’, [n.d.]; Biddle, S., ‘메타 기반 군사 챗봇, 공습에 대해 “쓸모없는” 조언 제공으로 광고’, The Intercept, 2024년 11월 24일; 및 Vincent, B., ‘Scale AI, 국가 안보 사용자를 위한 “Defense Llama” 대규모 언어 모델 공개’, DefenseScoop, 2024년 11월 4일.

24 Greipl, A. R., ‘도시전에서의 인공지능: 민간인 보호 강화 기회?’, Military Law and the Law of War Review, vol. 61, no. 2(2023); 및 Tucker, P., ‘특수부대원들은 AI가 전투에서 민간인 사망을 줄일 수 있기를 바란다’, Defense One, 2024년 8월 26일.

25 Abraham(각주 5)

26 Hsu, J., ‘전투 피해 탐지기는 전쟁 중 지원 단체의 신속한 대응에 도움이 될 수 있다’, New Scientist, 2024년 5월 31일; Tucker, P., ‘AI가 러시아 미사일로 인한 우크라이나 피해를 어떻게 예측할 수 있을까’, Defense One, 2023년 1월 9일.

의사결정지원체계의 결과는 특정 표적 결정을 내리는 사용자가 해석하고 실행합니다. 즉, 인공지능 기반 의사결정지원체계는 인간 사용자의 표적 결정 실행을 지원합니다. 이러한 차이는 두 체계가 각각 충족하려는 고유한 운용상의 요구를 반영합니다. 자율무기체계는 작전 범위, 지속성 및 속도를 향상시키도록 설계되었습니다.²⁷ 예를 들어, 지속적인 통신 링크 없이 작동할 수 있는 자율무기체계는 통신이 지연 혹은 제한되거나 군인의 접근이 위험한 환경에서도 작전 범위를 확장할 수 있습니다.

반면, 인공지능 기반 의사결정지원체계는 ‘인간 의사결정의 품질을 향상시키고 속도를 높이며 일관성을 강화함으로써’ 인간 의사결정자를 지원하도록 설계되었습니다.²⁸ 이는 인간 분석가가 인지 능력을 초과하는 방대한 양의 데이터를 일일이 해석해야 하는 부담을 줄이는 방식으로 작동합니다.²⁹ 예를 들어, 행동 방침 분석 과정에서 사용자에게 더 다양한 선제적 대응 옵션을 제시하고, 특정 대상과 관련된 정보를 필터링하고 표시(flagging)합니다. 이러한 기능은 사용자가 정보를 수동으로 검토할 필요성을 줄이고, 상위 수준의 판단을 수행할 여지를 제공합니다.

정책적 합의

정책적 관점에서 자율무기체계와 인공지능 기반 의사결정지원체계는 모두 무력 사용 결정에 영향을 미친다는 공통점을 갖지만, 두 가지 중요한 차이점이 존재합니다. 자율무기체계는 인공지능 기반 의사결정지원체계보다 좁은 범위의 표적 선정 작업에 사용되며, 표적 선정 결정을 직접 수행합니다. 반면 인공지능 기반 의사결정지원체계는 인간의 판단을 지원하는 역할을 합니다. 그러나 실질적으로 두 체계가 늘 명확히 구분되지는 않습니다. 각 체계의 기능은 본질적 속성보다는 구성, 교전 규칙, 작전 지휘 방식 등 운용 맥락에 따라 결정되기 때문입니다. 예를 들어, 미국의 MIM-104 패트리엇 미사일 체계는 수동 모드와 ‘자동 발사’ 모드를 모두 갖추고 있습니다. 수동 모드에서는 탐지 및 표적 교전 작업이 인간 운용자의 통제를 받지만, ‘자동 발사’ 모드에서는 시스템이 인간 운용자와 무관하게 선택된 표적과 자동으로 교전합니다.³⁰

패트리엇과 같은 체계는 특히 임무 실행 단계에서 자율무기체계와 인공지능 기반 의사결정지원체계 간 범주적 구분의 효용성이 재검토될 필요가 있습니다. 이러한 체계에 대한 정책 논의는 두 체계 간 범주 구분이 고정적 개념이 아니라, 사용 사례에 따라 달라질 수 있다는 점을 인정할 필요가 있습니다. 예컨대, 정책 논의는 인간 사용자가 체계와 상호작용하는 시점과 방식, 특히 표적 선정 주기 동안 인간과 기계 간의 표적 선정 작업이나 기능 분포가 어떻게 변화하는지에 초점을 둘 수 있습니다.

27 Boulanin and Verbruggen(각주 1), 61-63쪽; ICRC, ‘자율무기체계에 대한 ICRC의 입장’(각주 1).

28 Holland Michel, Decisions, Decisions, Decisions(각주 6), 18쪽.

29 Weinbaum, C. and Shanahan, J. N. T., ‘데이터 중심 시대의 정보’, Joint Force Quarterly, vol. 90, no. 3(2018), 5-6쪽.

30 Scharre, P., Army of None: Autonomous Weapons and the Future of War(W. W. Norton & Company: 런던, 2018), 9장.

보다 일반적으로 볼 때, 표적 선정 관행에 인공지능을 통합하려면 다음 세 가지 요소에 주의를 기울여야 합니다. (a) 체계가 자율무기체계 또는 인공지능 기반 의사결정지원체계로 작동하는 조건, (b) 체계의 어포던스(affordance)—즉, 무기 체계와의 연결성을 포함해 가능한 동작 범위를 결정하는 하드웨어 및 소프트웨어 사양, (c) 이러한 역할을 규정하는 승인 임계값(thresholds)과 통제 구조입니다. 이러한 분석은 운용적 현실에 보다 근거를 두어야 합니다. 이러한 맥락에서 제3장은 자율무기체계와 인공지능 기반 의사결정지원체계가 초래할 수 있는 의도하지 않은 피해의 비교 위험을 다룹니다.

3장. 의도하지 않은 피해 위험

자율무기체계와 인공지능 기반 의사결정지원체계는 모두 의도하지 않은 피해, 특히 민간인 피해 위험을 동반합니다. 이 장에서는 각 체계에서 이러한 위험이 어디서, 그리고 어떤 방식으로 발생하는지를 비교합니다. 기술적 한계가 체계의 신뢰성과 인간-기계 상호작용에 미치는 영향을 중점적으로 살펴보고, 다른 상황적 변수를 함께 고려합니다.

군사 분야에서 인공지능의 한계는 이미 잘 알려져 있습니다. 첫째는 ‘블랙박스(black box)’ 문제입니다. 이는 체계의 불투명성과 함께, 체계가 언제 어떻게 실패할지를 평가할 방법이 (현재) 부족하다는 점을 의미합니다.³¹ 그 결과, 의도하지 않은 동작의 원인을 명확히 규명하기 어렵습니다. 둘째, 인공지능 체계는 신뢰성 있는 작동을 위해 양질의 데이터(관련성 있고 완전하며 정확한 데이터)가 필요합니다. 그러나 전장 환경에서는 데이터 품질이 다양한 요인에 의해 쉽게 손상될 수 있습니다.³² 특히 배치 시점의 데이터 오염 등 적대적 공격에 취약합니다.³³ 인공지능 체계는 이 밖에도 설계된 조건이나 환경과 다른 상황에서 사용될 경우 실패하기 쉽습니다.³⁴ 또한 내재적 편향으로 인해 성별이나 인종에 따라 차별적 영향을 미칠 가능성도 있습니다.³⁵ 인공지능 기반 의사결정지원체계와 인공지능을 통합한 자율무기체계는 정의상 이러한 기술적 한계에 공통적으로 취약합니다.

신뢰성 문제

자율무기체계와 인공지능 기반 의사결정지원체계가 야기하는 위험의 핵심은 기술적 신뢰성에 있습니다.³⁶ 신뢰성이란 주어진 환경에서 일정 기간 동안 체계가 의도한 대로 작동할 것이라는 특성을 의미합니다.³⁷ 모든 기술은 디지털이든 아니든 신뢰성 문제를 내포하며, 실패하거나 의도하지 않은 방식으로 작동할 수 있습니다.

신뢰성은 더욱이 인공지능을 통합하는 체계에서 핵심적입니다. 인공지능 체계는 앞서 언급한 기술적 한계로 인해 인간 사용자에게 예측 불가능하거나 거의 감지되지 않는

31 Bunch, K. et al., 강화 학습 AI 시스템의 위험 평가: 기술을 넘어서(RAND Corporation: 캘리포니아주 산타모니카, 2024); AI Now Institute(주 5); Panwar, R. S., Qiang, L. 및 Shanahan, J. N. T.(편집), ‘군사 인공지능 테스트 및 평가 모델 실습’, INHR 및 Center for a New American Security, 2024년 12월.

32 Holland Michel, A., ‘블랙박스의 잠금 해제: 군사 AI의 예측 가능성 및 이해 가능성’, 유엔 군축 연구소, 2020, 7쪽; Menthe, L. et al., 전투원을 위한 인공지능의 한계 이해: 제1권—요약(RAND Corporation: 캘리포니아주 산타모니카, 2024).

33 Hoffman, W. 및 Kim, H. M., ‘군사적 결정 우위를 위한 인공지능의 위험 감소’, CSET 정책 간략서, 2023년 3월; Uesato, J. 등, ‘적대적 위험 및 약한 공격에 대한 평가의 위험’, 기계 학습 연구 논문집, vol. 80(2018).

34 Taddeo, M. 등, ‘국가 안보를 위한 인공지능: 예측 가능성 문제’, CETAS 연구 보고서, 2022년 9월.

35 Blanchard, A. 및 Bruun, L., ‘군사 인공지능의 편향’, SIPRI 배경 보고서, 2024년 12월.

36 Holland Michel, Decisions, Decisions, Decisions(주 6), 18쪽; 및 Probasco 등(주 6).

37 국제표준화기구, ‘신뢰성 - 어휘’, ISO/IEC TS 5723:2022, 2022년 7월.

방식으로, 혹은 둘 모두의 방식으로 실패할 수 있습니다. 이로 인해 결함은 종종 심각한 피해가 발생한 이후에야 발견되기도 합니다.³⁸

인간-기계 상호작용의 차이

직접적 또는 간접적 피해 발생 양상

자율무기체계는 액추에이터(actuator)와 같은 물리적 구성요소를 통해 인간 운용자와 독립적으로 목표 결정을 실행합니다. 따라서 ‘현실’에서 발생하는 위험은 직접적인 위험으로 볼 수 있습니다. 예를 들어, 자율무기체계가 위협이 아닌 대상(예: 민간인)을 위협(예: 전투원)으로 오인하는 경우, 이를 이른바 ‘거짓 긍정(false positive)’이라 합니다. 이 경우 공격을 차단하지 않는 이상 체계는 위협이 아닌 대상을 직접적으로 공격할 수 있습니다.

이러한 이유로 자율무기체계의 위험 완화 조치는 의도하지 않은 동작 문제의 관리에 초점을 둡니다.³⁹ 위험 완화 방식은 두 가지로 구분됩니다. 첫째, 대상 유형, 운용 기간, 배치 지역 범위 등 통제 매개변수가 체계 설계에 포함되도록 하여 배치 전이나 임무 실행 단계부터 자율무기체계를 관리하는 방식입니다.⁴⁰ 이러한 조치에는 자율무기체계의 감사 가능성, 인증 절차, 교전 규칙의 동적 업데이트가 포함됩니다.⁴¹ 둘째, 실시간 킬 스위치(kill switch)와 같은 운용자 개입형 감독 통제 절차를 적용하는 방식입니다.⁴²

반면, 인공지능 기반 의사결정지원체계는 인간의 의사결정을 지원하도록 설계되었기 때문에 현실 위험은 원칙상 간접적입니다. 즉, 위험이 실제 피해가 되려면 인간이 인공지능 기반 의사결정지원체계가 제공하는 정보(결과)를 해석하고 그에 따라 행동해야 합니다. 예를 들어, 이 체계가 부정확하거나 불완전하거나, 왜곡된 혹은 허위 정보를 인간에게 제공하는 경우, 예를 들면 위협이 아닌 대상을 위협으로 오인하는 경우, 인간 사용자가 이 정보를 사실로 받아들이고, 행동할 때 피해가 발생합니다. 따라서 인공지능 기반 의사결정지원체계의 위험 완화는 사용자가 부정확한 결과에 대해 이의를 제기하거나 수정하거나 무시할 수 있는 능력을 확보하는 데 중점을 두어야만 합니다.

38 Ryan, M., ‘AI에서 우리는 신뢰한다: 윤리, 인공지능 및 신뢰성’, 과학 및 공학 윤리, 제26권, 제5호(2020).

39 Blanchard, A. et al., ‘자율무기체계에 대한 위험 기반 규제 접근 방식’, Digital Society, vol. 4, no. 23(2025).

40 Boulanin, V. et al., ‘무기 체계의 자율성 한계: 인간 통제의 실질적 요소 식별’, SIPRI/ICRC, 2020년 6월.

41 Cummings, M. L., ‘치명적 자율무기: 의미 있는 인간 통제인가, 의미 있는 인간 인증인가?’, IEEE Technology and Society Magazine, vol. 38, no. 4(2019); Kwik, J. et al., ‘군사적 인공지능 통제: 교전 규칙 및 군사 지시 활용’, T. M. C. Asser Institute Policy Brief No. 2025-01, 2025년 2월 3일; 및 Spayne, P. et al., ‘안전하게 운용하기: 인공지능을 포함하는 치명적 자율무기체계에 대한 “안전한 운용”과 “안전하게 운용”의 개념 병합’, Defence Studies, vol. 25, no. 1(2025).

42 Verdiesen, I., Santoni de Sio, F. 및 Dignum, V., ‘자율무기체계에 대한 책임과 통제: 포괄적인 인간 감독을 위한 프레임워크’, Minds & Machines, vol. 31(2021); 및 Sharkey, N., ‘로봇 무기에 대한 인간 감독 통제 원칙을 향하여’, Politica & Società, vol. 3, no. 2(2014).

인간의 판단에 미치는 영향

자율무기체계는 인간 운용자와 지리적·시간적·인식적 측면에서 거리를 두고 표적 선정 결정을 내립니다. 따라서 사용자는 해당 체계가 운용 맥락상 미칠 수 있는 영향을 사전에 평가해야 합니다. 이러한 평가는 체계 사용의 합법성과 윤리성을 판단하는 데 핵심적인 역할을 합니다. 그러나 이러한 평가는 언제나 상당한 불확실성이 수반됩니다. 사용자가 체계의 배치 과정에서 발생 가능한 모든 상황을 예측하지 못했거나, 예측할 필요가 없었을 경우 자율무기체계의 사용은 의도하지 않은 피해로 이어질 수 있습니다.

인공지능 기반 의사결정지원체계는 사용자와 전장 간의 관계 일부를 매개하므로, 자율무기체계와는 다른 유형의 위험 시나리오가 있습니다.⁴³ 이 체계는 공격 개시를 평가하는 데 도움이 될 수 있는 정보를 구성·필터링·제공합니다. 그러나 부정확하거나 불완전하거나 왜곡된 혹은 허위 정보를 제시할 경우, 결과적으로 인간의 의사결정 과정을 저해할 수 있습니다. 예를 들어, 인공지능 기반 의사결정지원체계가 민간인이나 민간 시설의 존재 관련 정보를 잘못 제공하는 경우 사용자는 민간인 및 민간 시설 피해 위험을 잘못 판단하도록 만들 수 있습니다.

이러한 위험은 자동화 편향(automation bias)으로 인해 더욱 심화됩니다. 여기서 자동화 편향이란 ‘인간이 자신의 판단 능력보다 자동화된 시스템의 결과를 더 확신하는 것’을 의미합니다.⁴⁴ 나아가 이러한 위험은 인간의 오류나 기술적 한계보다는 설계 단계의 선택에서 기인할 수 있음을 인식해야 합니다. 결과적으로 인공지능 기반 의사결정지원체계는 복잡성과 잡음을 걸러내어 전장을 단순화된 형태로 제시하며, 이를 통해 표적 선정 관행을 지원하는데 적합한 기능에 초점이 맞추어져 있습니다.⁴⁵

인공지능 기반 의사결정지원체계는 인간이 표적 결정을 내리는 속도를 높임으로써 의도하지 않은 피해의 실체화에 기여할 수도 있습니다. 속도는 의사결정 지원을 위해 인공지능을 군사 분야에 도입하는 핵심 전략적 동인이지만, 동시에 위험의 주요 원천이기도 합니다.⁴⁶ 속도 증가는 더 많은 공격이 빠른 속도로 연속적으로 발생하거나 대상에 대한 더 많거나 나은 정보를 얻기 전에 발생할 수 있음을 의미합니다. 즉, 속도는 인간이 인공지능 기반 의사결정지원체계 결과를 감독하는 능력에 영향을 미칠 수 있습니다. 이는 예를 들어, 민간인을 단기간에 더 많은 공격에 노출시키거나 공격의 존재를 발견하기 전에 노출시키는 등 피해가 실체화될 가능성을 높입니다.

속도 편향(speed bias)과 자동화 편향이라는 두 가지 우려는 인공지능 기반 의사결정지원체계가 인간 의사결정자를 ‘지원’하는 대신 의사결정 과정에 지나치게 크고 과도한

43 Boulanin, ‘AI 기반 군사적 의사결정의 위험과 이점’(주 6).

44 Bode, I., ‘군사 영역에서의 인간-기계 상호작용 및 인간 행위’, 국제 거버넌스 혁신 센터(CIGI), 정책 간략 보고서 193호, 2025년 1월, 6쪽.

45 Barrett-Taylor, R., ‘전장의 디지털 표현의 한계’, CETAS 전문가 분석, 2024년 6월.

46 Schwarz, E., ‘책임 있는 군사 AI 거버넌스의 (불)가능성’, ICRC 인도주의 법률 및 정책 블로그, 2024년 12월 12일; Bo, M. 및 Dorsey, J., ‘군사 AI 및 무력 분쟁법에 대한 심포지엄: 속도의 “필요성” - 규제되지 않은 AI 의사결정지원체계의 민간인 비용’, OpinioJuris, 2024년 4월 4일; Reynolds, I. J., ‘미국 군사적 사고에서의 속도와 전쟁: AI 기반 의사결정을 위한 조건 매핑’, Millennium(2025년 4월 1일).

영향을 미칠 위험을 야기합니다. 이는 인공지능 기반 의사결정지원체계가 의사결정 과정에 해를 끼쳐 최적이지 아닌 결과를 초래할 수 있음을 의미할 수 있습니다. 그러나 최악의 경우, 인공지능 기반 의사결정지원체계가 사실상 자율무기체계로 사용된다는 것을 의미합니다. 즉, 인간 운용자와 인공지능 기반 의사결정지원체계 간의 위계가 뒤바뀌어 운용자가 결과에 대해 명목상의 감독만 수행하고 체계 권고사항을 수동적으로 승인하게 됩니다. 그 결과, 인간이 인공지능 기반 의사결정지원체계에 대해 판단을 내릴 수 있는 공간이 줄어들거나 완전히 폐쇄될 수 있습니다.

맥락적 변수

실제로 자율무기체계 및 인공지능 기반 의사결정지원체계 사용 시 의도하지 않은 피해가 발생하는 것은 체계의 본질적인 속성뿐만 아니라 체계의 성능 및 사용 방식에 영향을 미치는 맥락적 변수에도 달려 있습니다. 여기에는 운용 조건, 체계 어포던스(affordance) 및 통제 구조라는 세 가지 요인이 관련되며, 이 세 가지 요인은 모두 의도하지 않은 피해의 위험에 양적 및 질적 영향을 미칩니다. 이러한 요인은 위험이 발생할 수 있는 경로를 늘리거나 위험이 발생할 수 있는 경로의 마찰을 최소화하는 데, 주로 인간이 (의미 있는) 감독을 수행할 수 있는 범위를 줄임으로써 이루어집니다.

운용 조건

운용 조건은 체계가 사용되는 환경적 특성입니다. 체계 성능과 사용자의 위험 예측 능력에 영향을 미치는 변수에는 환경이 사전에 알려진 정도, 관측 가능한 정도, 어수선한 정도(cluttered), 동적 또는 정적인 정도 등이 있습니다. 예를 들어, 민간인이나 민간 시설물이 (비교적) 없는 환경에서 자율무기체계 또는 인공지능 기반 의사결정지원체계를 사용하면 민간인이나 민간 시설물에 대한 피해 위험이 줄어듭니다.

체계 어포던스(affordance)

체계 어포던스에는 체계의 소프트웨어, 하드웨어 및 다른 체계와의 연결성(예: 인공지능 기반 의사결정지원체계가 독립형 체계로 사용되는지 또는 무기 플랫폼에 연결되는지)이 포함되지만 이에 국한되지는 않습니다. 예를 들어, 일부 실험 프로그램에서는 인공지능 기반 의사결정지원체계가 실시간 사격 통제 체계에 연결되어 인공지능이 인간 사용자에게 무기 플랫폼을 자동으로 신호를 줄 수 있도록 합니다.⁴⁷ 무기 하드웨어와의

47 Northrop Grumman, 'Northrop Grumman, 전방 지역 방공에 최첨단 AI 기능 추가', 보도자료, 2024년 10월 7일; Zoldi, D., '모듈식 AI 및 ML 지원 시스템을 통한 불량 UAS 대응', Military Embedded Systems, 2022년 3월 2일; Insinna, V., '미 공군, 드론 조종사의 AI 두뇌에 대한 계약 체결, 하지만 세부 사항은 비밀 유지', Breaking Defense, 2024년 7월 29일; Airbus, 'Airbus와 Helsing, 유인 및 무인 군용 항공기의 팀 구성을 위한 인공지능 분야에서 협력', 보도자료, 2024년 6월 5일.

이러한 직접적인 연결은 인공지능 기반 의사결정지원체계가 도출한 결과와 위험 발생 간의 마찰을 줄입니다.

통제 구조

통제 구조는 표적 선정 주기에서 자율무기체계와 인공지능 기반 의사결정지원체계의 활용 방식을 결정합니다. 여기에는 누가, 어떤 조건에서, 어떤 수준의 자동화로 이러한 체계 사용을 승인할 수 있는지를 결정하는 조직 및 기술적 구성이 포함됩니다.⁴⁸ 통제 구조는 표적 선정 시 자동화 체계에 대한 감독을 보장하는 데 필수적인 요소입니다. 예를 들어, 이스라엘이 가자 작전에서 사용하는 합소라(‘복음’) 체계는 집계된 정보 입력을 기반으로 표적을 추천하는 인공지능을 내장하고 있는 것으로 알려졌습니다. 공식적인 공격 승인은 인간 운용자에게 있지만, 이 체계는 표적 추천과 표적 교전 사이의 마찰을 최소화하는 방식으로 신속한 표적 추천 주기를 지향하는 것으로 알려졌습니다. 잘 조정된 통제 구조가 없다면, 이러한 체계는 공식적으로는 통제가 인간에게 있더라도 표적 결정이 기능적으로 자동화되는 쪽으로 기울어질 수 있습니다.

통제 구조에는 사용자가 체계의 성능을 테스트, 평가 및 보장할 수 있도록 하는 프로세스와 절차가 포함될 수 있습니다.⁴⁹ 또한 데이터 거버넌스 관행 및 조직 작업 흐름과 같이 체계 한계에 대응하기 위한 사용 매개변수 및 위험 완화 조치를 설정하는 것도 포함될 수 있습니다.⁵⁰ 인공지능 기반 의사결정지원체계의 경우, 조직 작업 흐름에는 표적 선정 주기의 관련 단계에서 결과를 검증하고, 더 광범위한 정보 평가에서 결과에 가중치를 부여하며 적절하게 전달할 책임을 할당하는 것이 포함될 수 있습니다.⁵¹

정책적 함의

자율무기체계와 인공지능 기반 의사결정지원체계의 신뢰성 문제는 겹치는 부분이 있지만, 두 체계가 인간의 의사결정과 상호작용하는 방식이 다르기 때문에 의도하지 않은 피해의 발생 방식도 서로 다르게 나타납니다. 이러한 차이점을 이해하면, 위험 완화 전략이 어디에서 일치하는지 혹은 차이가 나는지 평가하는 데 도움이 될 수 있습니다. 또한, 환경 조건, 체계 어포던스, 통제 구조라는 세 가지 주요 맥락적 요소는 두 체계 모두

48 Devitt, S. K., ‘의미 있는 인간 명령: 자율무기체계에 대한 도덕적, 법적 책임을 가능하게 하는 방법으로서의 사전 통제 지시’, Geiß 및 Lahmann 편집(주 6); Boulanin, V. 및 Lewis, D. A., ‘군사 분야에서 AI 개발 및 사용에 대한 책임 있는 의존’, 윤리 및 정보 기술, 제25권, 제1호(2023); Bo, M., Bruun, L. 및 Boulanin, V., 자율무기체계 개발 및 사용에 있어서 인간의 책임 유지: 자율무기체계와 관련된 국제인도법 위반에 대한 책임에 관하여(SIPRI: 스톡홀름, 2022년 10월).

49 미국 국방부, 최고 디지털 및 인공지능 사무국(CDAO), 인공지능 모델의 테스트 및 평가: 테스트 및 평가 전략에서 고려해야 할 사항(CDAO: 워싱턴 DC, 2024년 4월).

50 Afina, Y. 및 Grand-Clément, S., ‘바이트와 전투: 책임 있는 군사 AI에 대한 데이터 거버넌스 포함’, CIGI 논문 번호 308, 2024년 10월; Vogel, K. M. 외, ‘AI가 정보 분석에 미치는 영향: 협업, 알고리즘 투명성, 책임성 및 관리 문제 해결’, Intelligence and National Security, vol. 36, no. 6(2021).

51 Hughes, M. 외, ‘AI와 전략적 의사결정: AI 강화 정보에서 신뢰와 불확실성 전달’, CETAS 연구 보고서, 2024년 4월.

에서 위험이 어떻게 실체화되는지 결정하는 데 핵심적인 역할을 합니다. 특히, 인공지능 기반 의사결정지원체계와 관련된 향후 정책은 이러한 요소들을 중점적으로 관리하여 의도하지 않은 피해의 위험을 더 효과적으로 줄이거나 통제하는 데 초점을 맞출 필요가 있습니다.

4장. 법적 측면

각국은 자율무기체계와 인공지능 기반 의사결정지원체계의 개발 및 사용이 국제인도법(IHL)을 포함한 국제법을 준수해야 한다는 데에 공감대를 형성하고 있습니다.⁵² 자율무기체계 관련 정책 논의에서 국제인도법 준수는 핵심 쟁점으로 다루어져 왔으며, 각국은 자율무기체계가 국제인도법에 제기하는 과제에 대한 심층적 이해를 발전시켜 왔습니다.⁵³ 본 장에서는 인공지능 기반 의사결정지원체계가 제기하는 법적 문제가 자율무기체계의 문제와 어느 정도 중복되는지, 또한 자율무기체계 정책 논의에서 얻은 통찰이 인공지능 기반 의사결정지원체계에 관한 정책적 노력에 어느 정도 영향을 미칠 수 있는지를 살펴봅니다.

국제인도법이 적대행위 수행에 있어 인간에게 요구하는 것과 기계에게 허용하는 것

표적 선정에 사용되는 자율무기체계와 인공지능 기반 의사결정지원체계는 모두 공격 수행 시 구별, 비례성, 예방의 원칙을 포함한 일련의 국제인도법 규칙을 준수해야 합니다.⁵⁴ 또한 두 기술 모두 이러한 규칙이 본질적으로 인간에게 요구하는 것과 기계에 허용하는 것에 대해 의문을 제기합니다.

이러한 해석상의 질문은 자율무기체계 논의의 핵심이었습니다. 국제인도법은 적대행위 수행을 규제하는 규칙을 준수하기 위해 필요한 평가적 결정과 판단 측면에서 인간이 자동화 기술에 의존하는 것을 금지하지 않는다는 점이 널리 인정되고 있습니다.⁵⁵ 그러나 국제인도법이 그러한 의존을 허용하는 범위가 무제한이 아니라는 점 또한 일반적으로 수용되고 있습니다. 이는 표적 결정 과정에서 인간 주체성이 반드시 유지되어야 하기 때문입니다. 비록 용어에 대한 논쟁은 있지만, 여기서 ‘인간 주체성’은 법적 의무 이행과 책무성 확보를 위해 인간이 모든 무기체계의 동작과 영향을 합리적으로 예측(foresee)하고 통제(control)할 수 있는 능력을 유지해야 한다는 포괄적 관점을 반영합니다.⁵⁶

52 CCW, LAWS에 관한 GGE, 2019년 회기 보고서, CCW/GGE.1/2019/3, 2019년 9월 25일, 부록 IV, ‘지침 원칙’; 유엔 총회 결의안 79/239(주 4); 및 REAIM(주 4).

53 예를 들어 CCW, LAWS에 관한 GGE, CCW/GGE.1/2023/CRP.2(주 13)를 참조하십시오.

54 1949년 제네바 협약에 추가되고 국제적 무력 충돌의 희생자 보호(API)에 관한 의정서, 1977년 12월 12일 서명 개방, 1978년 12월 7일 발효, 제41조, 제48조, 제51조, 제52조 및 제57조; 및 ICRC, ‘규칙’, 관습 IHL 데이터베이스, [n.d.], 규칙 1, 6, 7, 13, 14, 15, 16, 17, 18, 19 및 47.

55 예를 들어 CCW, LAWS에 대한 GGE, ‘무기 체계에서 자율성 사용에 대한 IHL 구현’, 미국이 제출한 작업 문서, CCW/GGE.1/2019/WP.5, 2019년 3월 28일; CCW, LAWS에 대한 GGE, ‘치명적 자율무기체계(LAWS) 분야의 신기술에 IHL을 적용하는 것에 대한 GGE 문서에 대한 영국의 제안’, 영국의 제안, 2022년 3월; ICRC, ‘군사 분야의 인공지능에 대한 유엔 사무총장 제출’(주 7) 참조; Bruun, L., Bo, M. 및 Goussac, N., 자율무기체계 개발 및 사용에 있어서 국제인도법 준수: IHL은 무엇을 허용하고, 금지하고, 요구하는가?(SIPRI: 스톡홀름, 2023년 3월).

56 예를 들어 CCW, LAWS에 대한 GGE, CCW/GGE.1/2023/CRP.2(주 13); ICRC, ‘자율무기체계에 대한 ICRC 입장’(주 1); Bruun, L., 자율무기체계 규제에 대한 2계층 접근 방식: 경로 및 가능한 요소 식별(SIPRI: 스톡홀름, 2024년 8월), 19쪽 참조; Lewis, D. 및 Sweeny, H., ‘인 지적 주체성 행사: 무력 충돌에서 자연 및 인공지능에 관한 법적 틀’, 하버드 로스쿨 국제법 및 무력 충돌 프로그램, 법적 개념 논문, 2025년 1월.

인공지능 기반 의사결정지원체계 역시 인간이 적대행위 의무를 수행하면서 기계적 프로세스에 어느 정도 의존할 수 있는지 질문합니다. 따라서 자율무기체계 정책 논쟁에서도 출된 교훈은 인공지능 기반 의사결정지원체계 맥락에도 적용될 수 있습니다. 즉, 자율무기체계 논쟁에서 확립된 인간 주체성 유지의 중요성에 대한 전제는 인공지능 기반 의사결정지원체계에서 (비)허용되는 인간-기계 상호작용의 유형과 정도를 판단하는 기준이 될 수 있습니다. 국제적십자위원회(ICRC)와 여러 기관들은 이미 군사 분야의 인공지능 적용, 특히 인공지능 기반 의사결정지원체계에서 인간 주체성 보장의 중요성을 강조해 왔습니다.⁵⁷

법적 책임과 책무성

법적 결정을 내리는 책임과 책무성은 인간(국가 기관 및 개인)에게 있으며, 기계에 전가될 수 없습니다.⁵⁸ 이는 자율무기체계와 인공지능 기반 의사결정지원체계 모두에 적용됩니다. 그러나 자율무기체계가 국가의 유해 사건 조사 및 억제 능력(국제인도법상 요구 사항)을 저해하여, 결과적으로 책임성 확보 능력을 약화시킬 수 있다는 우려가 오랫동안 제기되어 왔습니다.⁵⁹ 이러한 우려는 주로 두 가지 요인에서 비롯됩니다. 첫째, ‘블랙박스’ 문제로, 복잡한 자동화 체계가 어떤 기준에 따라 결과를 산출하는지 이해하고 설명하기 어렵다는 점입니다.⁶⁰ 둘째, ‘다수 참여 문제’로, 자율무기체계의 행동이 설계·개발부터 사용에 이르기까지 체계의 여러 생애 주기 단계에서 다양한 행위자의 결정을 반영할 가능성이 크다는 점입니다.⁶¹

책임성과 관련된 유사한 문제는 인공지능 기반 의사결정지원체계에서도 발생할 수 있습니다. 체계는 기술적·조직적 추적성 문제를 일으킬 가능성이 높아, 분석·권고·예측 등 특정 산출물이 어떠한 근거에서 도출되었는지 이해하고 설명하기 어렵게 만들 수 있습니다.⁶² 적용 범위가 광범위하기 때문에, 이러한 추적성 문제는 표적 선정 주기 전반의 (초기) 여러 단계에서 다양한 형태로 나타날 수 있습니다. 예를 들어, 표적 선정 프로세스 초기 단계에서 인공지능이 생성한 산출물에 의존할 경우, 무력 행사 결정 과정의 불

57 ICRC, ‘군사 분야의 인공지능에 관한 유엔 사무총장에게 제출한 의견’(각주 7); Peace Movement Aotearoa 및 Stop Killer Robots(뉴질랜드 아오테아로아), ‘군사 분야의 AI에 관한 유엔 사무총장에게 제출한 의견’, 2025년 4월 11일; Bode I. et al., ‘군사 분야의 AI 의사결정지원체계에 관한 유엔 사무총장에게 제출한 의견’, 2025년 4월 9일.

58 CCW, LAWS에 대한 GGE, ‘지침 원칙’(주 52); 유엔 총회 결의안 79/239(주 4); 및 REAIM(주 4).

59 API 의정서(주 54), 제85조 및 제146조. 자율무기체계 및 책임에 대해서는 Bo, M., ‘ICC 규정에 따른 민간인 공격 전쟁 범죄의 mens rea에 비추어 본 자율무기 및 책임 격차’, *Journal of International Criminal Justice*, vol. 2, no. 19(2021); Crootoof, R., ‘전쟁 불법 행위’, *New York University Law Review*, vol. 97, no. 4(2022); 및 Geiß, R., ‘자율무기체계 사용에 대한 국가 통제: 위험 관리 및 국가 책임’, 편집 R. Bartels 등, 군사 작전 및 국제법에 따른 통제 개념(T. M. C. Asser Press: 헤이그, 2021)을 참조하십시오.

60 예를 들어 ICRC, ICRC의 자율무기체계에 대한 입장 및 배경 문서(ICRC: 제네바, 2021년 5월), 7쪽; CCW, LAWS에 대한 GGE, ‘팔레스타인의 자율무기체계에 대한 규범적 및 운영적 프레임워크 제안’, 팔레스타인이 제출한 작업 문서, CCW/GGE.1/2023/WP.2, 2023년 3월 3일.

61 Bo, Bruun 및 Boulanin(주 48).

62 Bo(주 59); Holland Michel, 결정, 결정, 결정(주 6), 57쪽.

투명성이 높아질 수 있습니다. 그러나 인공지능 기반 의사결정지원체계가 인간의 책임 추궁 능력에 어떻게 영향을 미치는지는 전문가와 정책 입안자들이 특히 주의를 기울여야 하는 부분입니다.

법적 평가에서의 역할

자율무기체계에서는 사용자가 체계를 활성화하기 전에 구별 및 비례성에 관한 법적 평가를 수행해야 합니다. 이에 따라 주요한 우려는 특정 공격과 관련해 수행된 이러한 평가가 공격이 진행되는 동안에도 계속 유효한지, 다시 말해 사용자가 자율무기체계 활성화 시 국제인도법⁶³ 준수를 보장하도록 그 작동과 영향을 합리적으로 예측하고 통제할 수 있는지에 관한 것입니다. 특히 역동적인 환경이나 장기간에 걸친 자율무기체계 운용 상황에서는, 활성화 전에 이루어진 법적 평가가 더 이상 유효하지 않게 될 가능성에 대한 우려가 큼니다.

반면, 인공지능 기반 의사결정지원체계는 이러한 동작과 효과에 대한 합리적 예측 및 통제 능력에 동일한 우려를 직접적으로 제기하지는 않습니다. 자율무기체계와 달리, 인공지능 기반 의사결정지원체계는 법적 평가를 ‘실행’하지 않고, 사용자의 판단을 보조하도록 설계되어 있으며 그 동작은 원칙적으로 항상 인간의 감독하에 이루어집니다. 그러나 이것이 인공지능 기반 의사결정지원체계가 적대 행위 규칙 준수와 관련하여 아무런 문제가 없다는 뜻은 아닙니다. 오히려 이 체계는 인간 사용자가 법적 평가를 수행할 때 인공지능의 결과에 어느 정도까지 의존할 수 있는지에 관한 새로운 의문을 제기합니다.⁶⁴ 예를 들어, 확률적 방식에 기반한 표적 선정 조언에 의존할 경우 공격과 관련된 많은 국제인도법 의무의 맥락적·예방적 성격이 훼손되는지의 문제입니다. 이는 자율무기체계 논의에서는 상대적으로 덜 두드러졌지만, 인공지능 기반 의사결정지원체계의 맥락에서 국제인도법 준수를 보장하는 방법을 논의할 때 반드시 검토되어야 할 쟁점입니다. 국제적십자위원회(ICRC)와 관련 분야 전문가들은 인공지능 기반 의사결정지원체계의 사용이 국제인도법 준수를 위해 필요한 맥락 및 가치 기반의 인간 판단 능력을 저하시킬 수 있다고 경고했습니다.⁶⁵ 앞서 3장에서 논의한 바와 같이 또 다른 우려는 인공지능 기반 의사결정지원체계의 사용으로 인해 이 체계에 대한 속도·거리·편향 증가로 사

63 ICRC, ICRC의 자율무기체계에 대한 입장 및 배경 문서(각주 60), 7쪽; Woodcock, T. K., ‘인간/기계(학습) 상호작용, 인간 주체성 및 국제인도법 비례성 기준’, *Global Society*, vol. 38, no. 1(2023).

64 ICRC 및 제네바 아카데미(주 16), 10쪽; Woodcock(주 63); Dorsey, J. 및 Bo, M., ‘공동 표적 선정 주기에서의 AI 지원 의사결정지원체계: 법적 과제, 위험 및 인간적 차원’, *국제법 연구*(출간 예정).

65 예를 들어 ICRC 및 제네바 아카데미(주 16), 17쪽; Dorsey 및 Bo(주 64); Holland Michel, *Decisions, Decisions, Decisions*(주 6); Nadibaidze, Bode 및 Zhang(주 6); Scharre(주 30); Boutin, B. et al., ‘군대에서 인공지능의 글로벌 거버넌스에 대한 DILEMA 성명’, 군사 인공지능(DILEMA) 프로젝트에 국제법 및 윤리 설계, Asser Institute, 2023년 1월; 및 Stewart, R. 및 Hinds, G., ‘전쟁 알고리즘: 무력 충돌에서의 의사결정에서 인공지능 활용’, ICRC 인도주의 법 및 정책 블로그, 2023년 10월 24일.

용자가 자신의 판단을 적극적으로 행사하기보다는 체계의 권고를 수동적으로 승인하는 역할에 그치는 결과를 낳을 수 있다는 점입니다.⁶⁶

제36조 검토에 대한 합의

인공지능 기반 의사결정지원체계가 법적 평가를 ‘실행’하는 것이 아니라 ‘정보’를 제공한다는 점은, 법적 검토 의무에도 영향을 미칠 수 있습니다. 제네바 협약 제1추가의정서 제36조에 따르면, 분쟁 당사국은 모든 신형 무기, 전투 수단 및 전투 방법에 대해 그 사용이 일부 또는 모든 상황에서 합법적인지 여부를 검토해야 합니다.⁶⁷ 자율무기체계는 무기 체계로서 명백히 제36조 검토 대상에 해당합니다. 그러나 인공지능 기반 의사결정지원체계, 특히 그 핵심 소프트웨어와 같은 하위 구성요소가 어느 수준까지 검토 대상에 포함되는지는 불분명합니다.⁶⁸ 국제적십자위원회(ICRC)는 무기 플랫폼에 통합된 인공지능 기반 의사결정지원체계나, 무기 체계의 사용 방식에 영향을 미치는 인공지능 기반 의사결정지원체계는 각각 전투의 ‘수단’과 ‘방법’에 해당하므로 법적 검토의 대상이 되어야 한다고 주장했습니다.⁶⁹ 그러나 이는 여전히 국가 및 전문가들 사이에서 체계적으로 정립되어야 할 과제로 남아 있습니다.

정책적 합의

인공지능 기반 의사결정지원체계와 관련하여 국제인도법(IHL) 준수를 위한 노력은 일정 부분 자율무기체계 논쟁에서 비롯됩니다. 자율무기체계와 인공지능 기반 의사결정지원체계는 모두 적대 행위 수행을 규율하는 동일한 규칙의 적용을 받으며, 이들의 사용은 이러한 규칙 준수가 인간에게 무엇을 요구하고, 기계가 표적 결정에서 무엇을 허용받는지에 대한 질문을 야기합니다. 자율무기체계에 관한 책임성 논의의 상당 부분은 인공지능 기반 의사결정지원체계 논의에도 유사하게 적용될 수 있습니다. 예컨대, 책임 있는 지휘·통제 체계 구축, 감사 추적 시스템, 유해 사건에 대한 조사 메커니즘 등은 두 체계 모두에 필요하며, 이미 널리 공감대를 얻고 있는 조치입니다.⁷⁰

그러나 이처럼 중첩되는 쟁점이 있음에도, 자율무기체계는 표적 선정 결정을 직접 실행하는 반면, 인공지능 기반 의사결정지원체계는 그러한 결정에 정보를 제공한다는 점에서, 후자의 경우 특히 인간이 법적 평가를 위해 인공지능 결과에 얼마나 의존할 수 있는가 하는 고유한 과제가 존재합니다.

66 Dorsey 및 Bo(각주 64); 및 Zhou, W. 및 Greipl, A. R., ‘군사적 의사결정에서의 인공지능: 인간을 대체하지 않고 지원’, ICRC 인도주의 법 및 정책 블로그, 2024년 8월 29일.

67 API(각주 54), 제36조.

68 예를 들어 CCW, LAWS에 대한 GGE, ‘지침 원칙’(각주 52); 및 Klonowska, K., ‘제36조: AI 의사결정지원체계 및 기타 신흥 전쟁 기술 검토’, Gill, T. D. 등 편집, 국제인도법 연감, 제11권, 226~227쪽, 23(헤이그: T. M. C. Asser Press, 2020) 참조.

69 ICRC, ‘군사 분야의 인공지능에 관한 유엔 사무총장에게 제출한 의견’(각주 7).

70 예를 들어 CCW, GGE on LAWS, ‘Rolling text’, 2024년 11월 8일 참조.

더 나아가, 인공지능 기반 의사결정지원체계가 무력 적용 외에도 다양한 군사·비군사적 작업에 활용될 수 있다는 점은, 합법적 사용에 대한 논의가 단순히 적대 행위 수행에 관한 국제인도법(예: 제네바 협약 및 포로·민간인 보호 의무)에만 머물지 않고, 그 밖의 국제법적 쟁점까지 폭넓게 다뤄질 필요가 있음을 시사합니다.⁷¹ 예를 들어, 인공지능 기반 의사결정지원체계의 도입이 촉진하는 감시 및 데이터 수집의 확대는, 이러한 관행이 국제인권법과 어떻게 맞닿는지에 대한 면밀한 검토를 요구합니다. 다른 국제법 체계도 자율무기체계의 개발 및 사용과 관련이 있지만, 국제인도법에 대한 관심에 비해 크게 저평가되어 왔습니다. 그러나 인공지능 기반 의사결정지원체계의 군사적 적용 범위가 확대되고 다양한 정책 포럼에서 그 중요성이 커짐에 따라, 군사 분야에서 인공지능에 대한 의존도가 높아짐에 따른 국제법 체계 전반의 함의를 탐구하는 일이 시급해지고 있습니다.

71 전쟁포로의 처우에 관한 제네바협약(III), 1949년 8월 12일 서명 개방, 1950년 10월 21일 발효; 그리고 전시 민간인 보호에 관한 제네바협약(IV), 1949년 8월 12일 서명 개방, 1950년 10월 21일 발효.

5장. 정책 대응

자율무기체계는 10년 넘게 국가, 지역 및 국제 정책 논쟁의 대상이 되어 왔습니다.⁷² 이 논쟁을 위한 가장 중요한 국제 포럼은 2016년에 설립된 유엔 특정재래식무기금지협약 치명적 자율무기체계 관련 정부 전문가 그룹으로, 여기서 각국은 자율무기체계에 대한 적절한 규제 대응 방안을 논의합니다. 2023년 이후 자율무기체계에 대한 다자간 논의를 위한 두 가지 추가 경로가 생겼는데, 유엔 총회 제1위원회와 인공지능의 책임 있는 군사적 이용에 관한 고위급회의(REAIM)입니다. 간단히 말해, 자율무기체계에 대한 활발한 국제 다자간 논쟁이 이루어지고 있습니다.

반면, 군사 표적 선정에 사용되는 인공지능 기반 의사결정지원체계는 국제 정책 논의에서 비교적 새로운 주제로, 자율무기체계만큼 정책 입안자들에게 친숙하지 않습니다. 예를 들어, 현재 인공지능 기반 의사결정지원체계에 대한 규제 대응 방안을 논의하는 전담 다자간 포럼은 없습니다. 그럼에도 불구하고, 많은 정책 입안자와 시민사회 단체는 인공지능 기반 의사결정지원체계가 제기하는 위험을 고려하여 다자간 대응 방안을 마련할 것을 촉구해 왔습니다.

이러한 대조는 중요한 정책적 질문을 제기합니다. 즉, 국제 정책 논쟁은 인공지능 기반 의사결정지원체계를 어떻게 다루어야 할까요? 즉, 인공지능 기반 의사결정지원체계와 자율무기체계는 유사한 정책 대응을 필요로 할까요, 아니면 다른 정책 대응을 필요로 할까요? 그리고 이러한 체계를 별도로 고려해야 할까요, 아니면 함께 고려해야 할까요?

이 장의 목적은 이러한 질문에 직면한 정책 입안자들을 지원하는 것입니다. 정책 입안자들이 활용할 수 있는 세 가지 접근 방식은 간략하게 다음과 같습니다. (a) 자율무기체계에 대한 기존 다자간 노력에 인공지능 기반 의사결정지원체계를 구체적으로 포함시키기, (b) 인공지능 기반 의사결정지원체계에 전념하는 새로운 프로세스를 수립하기, 또는 (c) 인공지능 기반 의사결정지원체계에 대한 구체적인 접근 방식을 취하지 않습니다. 이 장에서는 이전 장에서 강조된 자율무기체계와 인공지능 기반 의사결정지원체계의 유사점과 차이점을 활용하여 각 방안의 실현 가능성과 예상되는 취약점을 평가합니다.

접근 방식 1: 자율무기체계 관련 기존 다자간 논의에 인공지능 기반 의사결정지원체계를 구체적으로 포함시키기

정책 입안자들이 활용할 수 있는 한 가지 접근 방식은 치명적 자율무기체계 정부 전문가 그룹(GGE on LAWS)과 같이 자율무기체계에 전념하는 기존의 다자간 협의 트랙에 인공지능 기반 의사결정지원체계를 포함시키는 것입니다. 인공지능 기반 의사결정지원체계와 자율무기체계가 제기하는 문제에는 상당한 중복이 존재하므로, 이러한 접근 방식에는 명확한 정당성이 있습니다. 예를 들어 군사 표적 선정에 사용되는 인공지능 기반

⁷² Blanchard, A. et al., '자율무기체계 정책 논쟁의 딜레마'(주 2).

의사결정지원체계는 인간의 의미 있는 개입, 민간인 및 민간 시설에 대한 위협 등 자율 무기체계에 대한 다자간 논의에서 국가들이 최우선적으로 제기해 온 쟁점들을 그대로 드러냅니다.⁷³ 국제 사회가 오랫동안 자율무기체계에 집중해 온 것은 인공지능 기반 의사결정지원체계에 대응할 수 있는 기회로 볼 수 있습니다. 10년간의 정책 논의를 통해 복잡한 자동화 체계와 인간-기계 상호작용의 어려움을 포함하여 군사 표적 선정 자동화 등 풍부한 지식이 축적되었습니다. 실제로 자율무기체계 정책 논의에서 얻은 오랜 통찰, 즉 기술을 표적 선정 과정에 통합하면 실제 무력 사용과는 별개의 고유한 과제가 발생한다는 점은 인공지능 기반 의사결정지원체계의 복잡성에 대응하는 데 중요한 기반이 됩니다. 더욱이 자율무기체계와 관련된 많은 도구, 기술 및 프로세스는 특정 형태의 인공지능 기반 의사결정지원체계에도 관련될 수 있습니다. 더 나아가 자율무기체계 관련 규범 개발 과정에서 활용된 많은 기술적 도구, 정책적 절차, 협의 메커니즘 역시 특정 형태의 인공지능 기반 의사결정지원체계에도 적용 가능하므로, 기존 논의 구조를 확장하는 실질적 토대로 기능할 수 있습니다.

따라서 자율무기체계 전용 프로세스에 인공지능 기반 의사결정지원체계를 포함시키는 것은 기존의 비용과 자원을 효율적으로 활용하고, 정책 입안자들이 자율무기체계 관련 다자간 논의를 통해 축적한 전문성을 보다 즉각적으로 활용할 수 있도록 함으로써 상당한 이점을 제공할 수 있습니다.

그러나 이러한 접근 방식에는 두 가지 주요 한계가 존재합니다. 첫째, 자율무기체계에 관한 현재의 다자간 노력이 유지하고 있는 기술적 중립성은 인공지능 활용이 제기하는 구체적인 문제들이 간과될 수 있다는 뜻입니다. 둘째, 자율무기체계와 인공지능 기반 의사결정지원체계의 적용 범위 차이는 논의 주제를 제한하거나 희석시켜, 관련 노력을 저해할 수 있습니다. 이러한 한계점에 대해서는 아래에 자세히 설명되어 있습니다.

기술적 중립성

기존 자율무기체계 다자간 포럼, 특히 치명적 자율무기체계 관련 정부 전문가 그룹은 자율무기체계를 뒷받침할 수 있는 기술 유형에 대해 대체로 중립적인 입장을 유지해 왔습니다. 이는 현재 추진 중인 이니셔티브들이 표적 살상에 드론을 사용하는 사례의 증가와, 인공지능을 포함한 다양한 기술의 발전으로 인간의 감독 없이 표적 살상이 가능해질 것이라는 우려 속에서 형성되었음을 보여줍니다.⁷⁴ 비록 치명적 자율무기체계 관련 정부 전문가 그룹에서 인공지능과 데이터 문제가 점점 더 중요하게 다뤄지고 있지만, 해당 그룹의 권한은 여전히 ‘치명적 자율무기체계 분야의 신기술’로 엄격히 한정되

73 유엔, 총회, 제79차 회기, ‘치명적 자율무기체계’, 사무총장 보고서, A/79/88, 2024년 7월 1일.

74 예를 들어, 자율무기체계에 대한 현대 다자간 이니셔티브를 활성화한 보고서는 인공지능을 거의 언급하지 않고 대신 로봇 공학의 발전에 초점을 맞춥니다. 유엔, 총회, 인권 이사회, 제23차 회기, ‘사법 외, 약식 또는 자의적 처형에 대한 특별 보고관 보고서’, A/HRC/23/47, 2013년 4월 9일.

어 있습니다.⁷⁵ 다시 말해, 인간의 개입 없이 표적을 식별, 선정, 공격하는 체계에 초점이 있습니다.

군사 표적 선정에 활용되는 인공지능 기반 의사결정지원체계에 대한 현재의 우려는, 인공지능과 관련된 데이터 수집 및 처리 기술이 사회 전반에 미치는 심오한 영향에 대한 보다 광범위한 논의의 일환으로 제기된 것입니다. 즉, 최근 인공지능 기반 의사결정지원 체계가 현대 무력 분쟁에서 사용되면서 주목받게된 인공지능의 안전하고 책임 있는 사용을 보장하려는 국제적 정책 논의와 맞물려 있습니다.

그러나 자율무기체계 관련 다자간 포럼이 기술적으로 중립적인 입장을 유지하는 한, 이러한 포럼에서 인공지능 기반 의사결정지원체계를 다룬다는 것은 이를 글로벌 정책 의제에 올려 놓았던 사회적 우려와 분리하여 다루게 됨을 의미 합니다. 즉, 인공지능과 같은 특정 기술을 논의에서 배제한다면, 의료나 법 집행 등 안전이 핵심적인 다른 분야에서 인공지능 거버넌스와 관련된 통찰과 발전의 이점을 누릴 기회를 놓칠 수 있습니다. 예를 들어, 인공지능 기반 의사결정지원체계가 민간 인공지능 거버넌스 경로에 포함된 다면, 특히 이중 용도의 감시 환경에서 데이터 사용에 대한 인권 기반의 제약과 국제인도법 중심의 거버넌스를 조화시키기 위한 노력이 필요하게 될 것입니다.⁷⁶ 이는 또한 인공지능 기반 의사결정지원체계에 대한 국제 정책 대응이, 군사 영역에서의 인공지능의 다른 잠재적 활용 분야와 분리될 수 있음을 의미할 수 있습니다.

적용 범위

자율무기체계에 대한 기존 다자간 이니셔티브는 주로 자동화된 무력 적용으로 인해 제기되는 인도주의적 우려에 의해 추진되어 왔습니다. 앞서 설명한 바와 같이, 정부 전문가 그룹은 치명적 자율무기체계 중 특히 군사 표적 선정에 사용되는 체계인 자율무기체계에 초점을 두었습니다. 2장에서 언급했듯이 인공지능 기반 의사결정지원체계는 전투 관리, 공급망 물류, 분쟁 및 불안정 예측, 인도주의적 서비스를 포함하여 군사 작전의 전략적, 작전적, 전술적 수준 전반에 걸쳐 훨씬 더 광범위한 응용 프로그램을 포괄합니다. 본 보고서가 중점적으로 다루는 군사 표적 선정 범위 내에서도, 인공지능 기반 의사결정지원체계는 자율무기체계보다 훨씬 더 광범위한 기능과 작업을 포괄합니다.

자율무기체계에 전념하는 논의에 인공지능 기반 의사결정지원체계를 포함시키는 것은, 자율무기체계와 밀접하게 연관된 인공지능 기반 의사결정지원체계의 하위 체계 일부를 제한적으로 다루거나, 혹은 자율무기체계 전용 포럼에서 논의 범위를 확장해야 함을 의미합니다. 전자의 접근은 인공지능 기반 의사결정지원체계에 대한 매우 제한된 다자간 대응으로 이어질 수 있으며, 후자의 접근은 논의가 피상적이 되거나 세부적으로 다루기

75 예를 들어 CCW, LAWS에 대한 GGE, ‘롤링 텍스트’(주 70)를 참조하십시오.

76 예를 들어 Morley, J., et al., ‘의료를 위한 데이터 및 인공지능 관리: 국제적 이해 개발’, JMIR 형성 연구, vol. 6, no. 1(2022)를 참조하십시오.

위해 상당한 시간과 자원을 요구할 수 있습니다. 후자의 경우 기존 절차와 자원을 활용하여 얻은 효율성이 상쇄될 가능성이 높습니다.

접근 방식 2: 인공지능 기반 의사결정지원체계 전담 프로세스 구축

두 번째 접근 방식은 인공지능 기반 의사결정지원체계를 다루는 전담 포럼을 구축하는 것입니다. 이는 총회 제1위원회의 의제에 군사 분야 인공지능을 포함하거나, 인공지능의 책임 있는 군사적 이용에 관한 고위급회의(REAIM)에 이를 연계하는 것처럼, 기존 군사 분야 인공지능 전담 프로세스의 일부로 인공지능 기반 의사결정지원체계에 집중하는 방식일 수 있습니다. 또는 기존 포럼 내에 인공지능 기반 의사결정지원체계를 전담하는 별도의 프로세스를 구축하는 방식도 가능합니다. 예를 들어, 치명적 자율무기체계 관련 정부 전문가 그룹(GGE)의 후원 아래 전문가 실무그룹을 구성하는 방식이 이에 해당합니다.

인공지능 기반 의사결정지원체계와 자율무기체계에 관한 쟁점에는 상당한 차이가 존재합니다. 두 체계의 적용 범위, 전제하는 인간-기계 상호작용의 형태, 그리고 각각이 야기하는 법적 문제(2~4장 참조)가 다르기 때문입니다. 이러한 점에서, 이 접근 방식은 두 가지 경로 모두를 정당화할 수 있습니다. 인공지능 기반 의사결정지원체계 전담 프로세스는 이러한 문제를 해결하고, 자율무기체계 전담 프로세스가 포괄하기 어려운 광범위한 사안을 탐구할 수 있는 기회를 제공합니다. 이를 통해 정책 입안자들은 다른 분야에서 축적된 인공지능 거버넌스 관련 통찰과 발전을 통합하여 인공지능 기반 의사결정지원체계의 폭넓은 적용을 다룰 수 있습니다. 또한 치명적 자율무기체계 관련 정부 전문가 그룹(GGE) 논의에서 진전을 제약해 온 기존의 틀을 넘어설 가능성도 있습니다. 그러나 이 접근 방식에는 최소 두 가지 과제가 존재합니다. 첫째, 추가적인 자원과 전문성이 요구된다는 점, 둘째, 중복되거나 상충하는 거버넌스 체계가 형성될 위험입니다.

논쟁의 기준을 설정하고 전문성을 개발하는 데 필요한 노력, 시간, 자원

새로운 프로세스에서 논의의 기준에 대한 합의에 도달하기까지 상당한 시간이 필요합니다. 인공지능 기반 의사결정지원체계는 그 유형과 용도가 매우 다양하므로, 광범위한 관점을 취할지 혹은 고위험 적용 분야에 집중할지를 결정해야 합니다. 인공지능 기반 의사결정지원체계에 대한 전담 거버넌스 노력은 기존의 정책 통찰과 자율무기체계 논의에서 축적된 전문성을 토대로 이루어져야 합니다. 그러나 이 전문성을 인공지능 기반 의사결정지원체계의 특성에 맞게 적용하고 해석하는 방법에 대한 추가 연구가 필요합니다. 또한 자율무기체계 논의에서 충분히 다뤄지지 않았던 인공지능 기반 의사결정지원체계 관련 측면에 대한 새로운 연구가 뒷받침되어야 합니다. 이를 테면 인공지능과 관련된 데이터 수집 및 처리 기술 활용, 그리고 그에 따른 인권적 고려사항 등이 포함될 수 있습니다.

중복되면서도 모순되는 거버넌스 체제의 위험

2장에서 제안한 바와 같이, 자율무기체계와 인공지능 기반 의사결정지원체계의 구분은 단순한 범주적 구획보다 훨씬 더 모호합니다. 적어도 임무 수행 시점에서는 두 체계 간의 구분이 기술 자체의 본질적 속성보다는 사용 조건, 제공 능력, 통제 구조 등에 따라 달라집니다. 따라서 정책 논의에서 동일한 범주적 구분을 적용하는 것은, 예컨대 치명적 자율무기체계에 대한 정부 전문가 그룹이 자율무기체계에 대해 기술 중립적 접근 방식을 채택하는 것처럼, 특정 정책 선택의 반영이라 할 수 있습니다. 인공지능 기반 의사결정지원체계와 자율무기체계는 실제로 중복될 가능성이 있기 때문에, 각 체계에 대해 독립적으로 운영되는 거버넌스 프로세스 간에도 겹치는 영역이 존재할 수 있습니다. 이러한 중복은 각각의 노력 간 비효율을 초래하거나 거버넌스 구조의 일관성을 훼손할 위험이 있습니다. 따라서 자율무기체계와 인공지능 기반 의사결정지원체계에 대한 별도의 논의 과정은 진행 상황에 대한 전문성과 통찰을 공유함으로써, 정책적 대응의 일관성과 대칭성을 확보하고 중복적·모순적 결과를 피하는 것을 목표로 해야 합니다.

접근 방식 3: 인공지능 기반 의사결정지원체계에 대한 별도 접근 방식 취하지 않기

세 번째 방안은 국가가 인공지능 기반 의사결정지원체계에 대한 별도 접근 방식을 취하지 않는 것입니다. 이는 인공지능의 책임 있는 군사적 이용에 관한 고위급회의(REAIM)를 포함하여 더 포괄적인 틀 안에서 인공지능 기반 의사결정지원체계를 다루며, 체계 중심의 접근을 피하는 것을 의미합니다. 이러한 방식에서 국가는 유엔 총회의 군사 분야 인공지능 결의안, 치명적 자율무기체계(LAWS) 관련 정부 전문가 그룹(GGE), 군사 분야의 책임 있는 인공지능 이니셔티브 등 기존의 다자적 논의 과정 속에서 인공지능 기반 의사결정지원체계에 대한 명시적인 언급을 생략할 수 있습니다. 그 결과, 이러한 프로세스의 결과가 인공지능 기반 의사결정지원체계의 개발 및 활용에 결국 영향을 미칠 것으로 예상됩니다. 이 접근 방식의 장점은 높은 유연성과 함께 군사 분야에서 인공지능이 적용되는 광범위한 영역을 포괄적으로 논의할 수 있다는 점입니다. 반면, 인공지능 기반 의사결정지원체계를 명시적으로 다루지 않기 때문에 해당 결과의 인공지능 기반 의사결정지원체계와의 관련성이 불분명해질 가능성이 있으며, 체계가 야기할 수 있는 특정 위험을 완화하기 위한 세부적 조치가 충분히 마련되지 못할 수 있습니다.

6장. 주요 결과 및 권고사항

앞서 제시한 자율무기체계와 인공지능 기반 의사결정지원체계의 비교·분석 결과는 특성화(2장), 의도하지 않은 피해 위험(3장), 법적 측면(4장), 정책 대응(5장)에서 도출되었으며, 표 6.1에 요약되어 있습니다. 두 체계는 몇 가지 공통점을 지니지만(특히 일부 운용 맥락에서 중첩될 수 있다는 점에서, 범주적 구분은 정책적 선택의 문제임을 유념해야 합니다), 보다 본질적인 차이점들도 존재하며, 이에 따라 두 체계를 함께 혹은 개별적으로 다루는 방식은 서로 다른 함의를 지닙니다. 본 장에서는 비교 분석을 통해 도출된 두 가지 주요 결과와, 정책 입안자들이 향후 관련 논의를 심화하기 위한 세 가지 핵심 권고사항을 제시합니다.

주요 결과

자율무기체계와 인공지능 기반 의사결정지원체계는 모두 무력 사용에 대한 인간의 개입에 영향을 미칩니다.

자율무기체계와 인공지능 기반 의사결정지원체계는 모두 무력 사용에 대한 의사결정 방식을 변화시킴으로써, 인간의 역할을 새롭게 정의하는 잠재력을 가지고 있습니다. 무력 행사로 이어지는 의사결정 과정은 본래 다단계적이며, 다양한 행위자에 분산되어 있습니다. 여기에 자율무기체계와 인공지능 기반 의사결정지원체계를 도입할 경우, 의사결정 전체 과정에서 인간 의사결정권자의 역할, 책임, 법적 의무에 대한 의문이 제기됩니다.

예를 들어, 자율무기체계와 인공지능 기반 의사결정지원체계는 모두 인간의 감독과 통제가 어떻게 유지될 수 있는지, 그리고 의도하지 않은 피해의 위험을 최소화할 수 있는 방법은 무엇인지에 대한 우려를 불러일으킵니다. 자율무기체계의 경우, 이러한 우려는 주로 사용자의 직접적인 입력 없이 원격으로 작동하는 체계와 관련이 있습니다. 반면 인공지능 기반 의사결정지원체계의 경우, 표적 선정 결정 속도의 증가와 사용자가 시스템의 결과를 이의 제기 없이 그대로 수용하거나, 수정하지 않고 의존할 가능성과 관련이 있습니다.

자율무기체계와 인공지능 기반 의사결정지원체계는 무력 행사에서 서로 다른 기능을 수행합니다.

자율무기체계와 인공지능 기반 의사결정지원체계가 표적 선정 결정에서 인간의 역할에 각각 어떠한 방식으로 영향을 미치는지를 이해하는 것은, 의도하지 않은 피해의 위험을 관리하고 합법적 사용을 보장하는 데 중요합니다(그림 6.1). 자율무기체계는 표적 결정을 ‘실행’하는 역할을 하므로, 자율적인 무력 행사가 법적 요건을 위반하거나 예기치 못한 결과를 초래할 우려가 있습니다. 반면, 인공지능 기반 의사결정지원체계는 표적 결

정에 ‘정보를 제공’하므로, 사용자에게 제공되는 데이터를 필터링, 선택, 분석하여 사용자가 해석하고 이에따라 행동하게 함으로써 사용자와 전장의 관계를 중재한다는 우려가 있습니다. 인공지능 기반 의사결정지원체계가 표적 결정에서 인간의 역할을 변화시키는 방식은, 어떤 의미에서 자율무기체계보다 더 복잡하고 광범위합니다. 인공지능 기반 의사결정지원체계가 표적 결정에 대한 판단에 자체에 영향을 미치기 때문입니다. 특히 인공지능 기반 의사결정지원체계가 전장에 대해 부정확하거나 불완전하거나 오해의 소지가 있거나 잘못된 정보를 제공하는 경우 더욱 심각한 우려를 낳을 수 있습니다

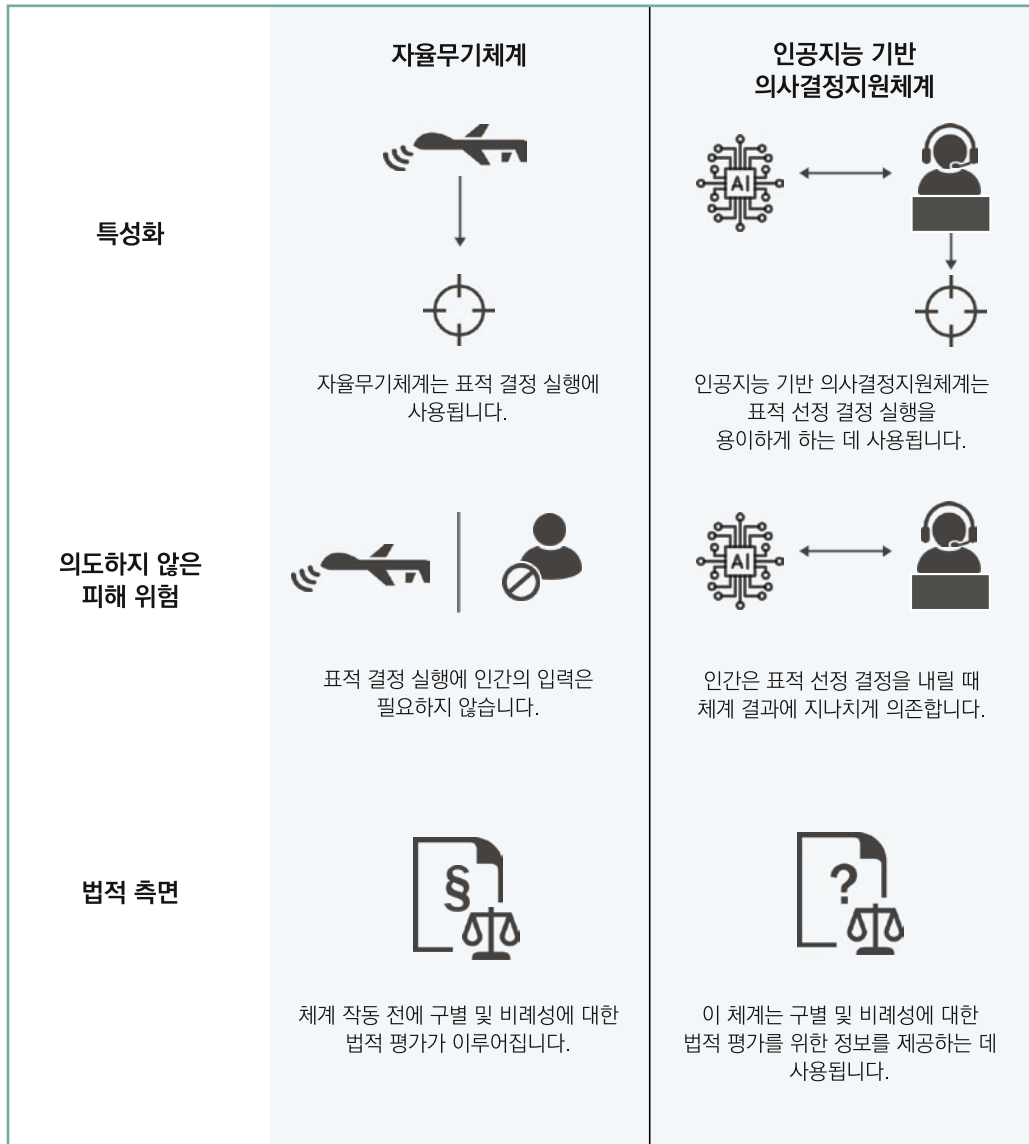


그림 6.1. 인간-기계 상호작용, 위험 경로 및 무력 사용에서 인공지능 기반 의사결정지원체계와 자율무기체계 간의 차이점

권고사항

인공지능 기반 의사결정지원체계 전담 다자간 프로세스 고려

현대 무력 분쟁에서 인공지능 기반 의사결정지원체계의 활용이 확산되고 있다는 보고에 따라, 일부 정책 입안자와 시민사회 단체는 이 체계에 대한 국제적 다자간 대응의 시급한 필요성을 제기했습니다. 이러한 요청에 대응하기 위해 국가는 최소 세 가지 방안을 검토할 수 있습니다.

- (a) 자율무기체계에 초점을 둔 기존 다자간 프로세스에 인공지능 기반 의사결정지원체계를 포함하는 방안,
- (b) 자율무기체계 프로세스 내 하위 절차로서 혹은 별도 전담 프로세스로서 이 체계를 다루는 새로운 구조를 수립하는 방안,
- (c) 현재의 자율무기체계 논의가 인공지능 기반 의사결정지원체계를 자연스럽게 포괄할 것이라는 가정 아래, 별도의 접근을 취하지 않는 방안.

각 방안은 정치적 실현 가능성과 정책적 비용·효익 문제를 수반합니다. 국가는 이러한 요청에 가장 현실적이고 균형 있는 방식으로 대응할 수 있는 전략을 모색할 필요가 있습니다.

자율무기체계 정책 전문성의 확장

국제사회가 자율무기체계 문제에 쏟아온 정책적 노력과 축적된 전문성은 인공지능 기반 의사결정지원체계 거버넌스에 귀중한 토대가 됩니다.

자율무기체계 관련 논의는 인간과 기계의 상호작용, 책임 이행, 위험 관리, 법적 의무 준수 등에서 풍부한 지식을 축적해 왔습니다. 이러한 교차 영역을 고려할 때, 인공지능 기반 의사결정지원체계에 대한 국제적 정책 논의는 자율무기체계 거버넌스에서 도출된 통찰을 적극 반영해야 합니다. 정책 입안자들은 이 통찰이 인공지능 기반 의사결정지원체계의 제도 설계 및 정책 개발에 어떻게 적용될 수 있을지 연구를 심화해야 합니다.

인공지능 기반 의사결정지원체계에 대한 이해 제고와 연구 강화

인공지능 기반 의사결정지원체계는 기존 자율무기체계 논의에서 충분히 다루어지지 않았던 새로운 쟁점과 윤리·법적 고려사항을 드러냅니다. 예를 들어, 인공지능과 관련된 데이터 수집 및 처리 기술의 활용과 이 시스템들에 의해 제기되는 인권적 고려사항에 관한 논의가 있습니다. 이 체계가 새로운 글로벌 정책 의제로 부상하고 있다는 사실은 각국이 인공지능을 군사적 의사결정 과정에 통합하는 방식과 그 한계를 재검토할 기회를 제공합니다. 따라서 인공지능 기반 의사결정지원체계는 인공지능을 통합한 의사결정 과정에서 인간의 주체성을 어떻게 유지하고 강화할 것인지에 관한 정책적 논의의 기회를 마련합니다. 국가들은 해결해야 할 주요 분야를 식별하고 연구를 의뢰해야 합니다.

표 6.1. 4가지 주요 영역에서 비교된 자율무기체계(AWS)와 인공지능 기반 의사결정지원체계(AI-DSS)

분야	유사성	차이점
유형	자율무기체계와 인공지능 기반 의사결정지원체계는 모두 표적 선정 결정 과정에서 인간의 역할과 개입 수준을 변화시킵니다.	<ol style="list-style-type: none"> 1. 표적 선정 주기의 범위: 자율무기체계는 임무 실행 단계에 국한되지만, 인공지능 기반 의사결정지원체계는 표적 선정 주기의 여러 단계에서 보다 광범위하게 사용됩니다. 2. 표적 선정 결정의 실행 역할: 자율무기체계는 표적 선정 결정을 스스로 실행하지만, 인공지능 기반 의사결정지원체계는 그 결정을 사용자가 실행할 수 있도록 지원합니다. 3. 인공지능 기반 의사결정지원체계는 군사 표적 선정 영역을 넘어, 다양한 군사 및 비군사 응용 분야로 확장됩니다.
의도치 않은 피해 위험	자율무기체계와 인공지능 기반 의사결정지원체계는 모두 인공지능의 기술적 한계로 인해 발생할 수 있는 신뢰성 문제를 내포합니다.	<ol style="list-style-type: none"> 1. 인간-기계 상호작용의 차이: 자율무기체계는 표적 선정 결정을 인간 개입 없이 실행하므로 위험이 직접적인 반면, 인공지능 기반 의사결정지원체계는 인간이 그 결과를 해석하고 행동해야 하므로 위험이 간접적입니다. 2. 인간 판단에 미치는 영향: 자율무기체계의 사용자는 체계 운용 맥락에서 예상되는 효과를 평가해야 하며, 이러한 과정에서 예견하지 못한 피해 위험이 발생할 수 있습니다. 반면 인공지능 기반 의사결정지원체계의 사용자는 체계가 제시하는 정보를 해석·수용·활용해야 하므로, 불량 데이터·결함 있는 분석 결과·자동화 편향으로 인한 위험이 존재합니다.
법적 측면	<ol style="list-style-type: none"> 1. 자율무기체계와 인공지능 기반 의사결정지원체계는 모두 적대 행위 수행 시 국제인도법(IHL)의 적용을 받습니다. 2. 두 체계 모두 기계 의존적 의사결정이 국제인도법 의무 이행에 미치는 영향과, 책임·책무성을 어떻게 보장할 것인지에 대한 의문을 제기합니다. 	<ol style="list-style-type: none"> 1. 법적 평가에서의 역할: 자율무기체계는 법적 평가를 '실행'하는 체계이므로, 사용자가 그 효과를 사전에 합리적으로 예측·통제할 수 있는지 여부가 핵심 쟁점입니다. 2. 반면 인공지능 기반 의사결정지원체계는 법적 평가 과정에서 인간이 인공지능의 결과를 어느 정도 신뢰·의존할 수 있는지에 대한 문제를 제기합니다. 3. 제36조 검토: 자율무기체계는 명백히 무기 기술에 해당하므로, 국제인도법 제36조상의 법적 검토 절차가 적용됩니다. 반면 인공지능 기반 의사결정지원체계가 동일한 검토 요건의 적용 대상이 되는지는 아직 명확하지 않습니다. 4. 국제인도법을 넘어: 인공지능 기반 의사결정지원체계는 인권법(예: 감시·데이터 수집)과 환경법(예: 자원 채굴과 생산의 지속가능성) 등 다양한 분야와 연결됩니다. 반면 자율무기체계에는 이러한 법적·윤리적 측면이 상대적으로 덜 탐구되어 있습니다.

33 자율무기체계와 인공지능 기반 의사결정지원체계의 군사 표적 선정

정책적 대응	<p>1. 자율무기체계와 인공지능 기반 의사결정지원체계는 모두 무력 사용에서 기계 자율성이 담당하는 역할에 대한 정책적 대응이 필요합니다.</p> <p>2. 자율무기체계가 인공지능 기능을 통합하고, 인공지능 기반 의사결정지원체계가 군사 표적 선정에 사용되는 한, 두 체계 모두 군사 의사결정과 표적 선정에서 인공지능 활용을 다루는 정책적 대응을 요구합니다.</p>	<p>1. 기술적 중립성: 자율무기체계에 대한 다자간 정책 노력은 인간 개입 없이 표적을 식별·선택·교전할 수 있는 무기에 대한 우려를 기반으로 하여 기술적으로 중립적 접근을 유지하고 있습니다. 반면 인공지능 기반 의사결정지원체계에 대한 논의는 군사적 응용을 넘어, 인공지능 자체 기술의 확산과 윤리적 함의에 대한 우려의 맥락에서 제기됩니다.</p> <p>2. 적용 범위: 자율무기체계는 군사 표적 선정에 한정되지만, 인공지능 기반 의사결정지원체계는 군사·민간을 아우르는 광범위한 응용 분야를 가지며, 이에 따라 더 포괄적인 정책 고려가 요구됩니다.</p>
--------	---	---

저자소개

알렉산더 블랜차드(Alexander Blanchard) 박사는 SIPRI 인공지능(AI) 거버넌스 프로그램의 선임연구원이다. 인공지능의 군사적 응용 분야의 개발, 사용 및 통제와 관련된 문제들을 위주로 연구하고 있다.

로라 브룬(Laura Bruun)은 SIPRI 인공지능 거버넌스 프로그램의 연구원이다. 새로운 군사 기술, 특히 자율무기체계와 군사 AI가 국제인도법(IHL)의 준수 및 해석에 미치는 영향을 중심으로 연구하고 있다. 브룬의 연구 주제에는 군사용 AI 시스템 내 편향이 IHL 준수에 미치는 영향과, 군사 AI의 개발 및 사용 과정에서 인간의 책임성을 어떻게 보장할 수 있는지가 있다.

번역: 이대훈 (피스모모 평화/교육 연구소 소장)

검수: 김가연 (피스모모 리서치랩 실장)



**STOCKHOLM INTERNATIONAL
PEACE RESEARCH INSTITUTE**

Signalistgatan 9
SE-169 72 Solna, Sweden
Telephone: +46 8 655 97 00
Email: sipri@sipri.org
Internet: www.sipri.org



**PEACEMOMO
피스모모**

서울특별시 서대문구 연희로 41다길 48-12, 1층
전화: 02-6351-0904
이메일: office@peacemomo.org
홈페이지: www.peacemomo.org