국방 AI 자율 소프트웨어 정책에 관한 고찰: 사례 기반 분석과 정책과제

이다정, 권익현, 김동성*

국립금오공과대학교, 국립금오공과대학교, 국립금오공과대학교

djlee@kumoh.ac.kr, inkwonaf@kumoh.ac.kr, *dskim@kumoh.ac.kr

A Study on Defense AI Autonomous Software Policy: Case-Based Analysis and Policy Implications

Lee Da Jung*, Kwon Ik Hyun, Kim Dong Seong

요 약

본 연구는 국방 분야에서 인공지능(AI) 자율 소프트웨어의 발전과 적용 현황을 분석하고, 실제 사례를 바탕으로 문제점과 정책과제를 도출하였다. 최근 우크라이나-러시아 전쟁에서 드론 및 자율무기체계의 활용은 AI 소프트웨어가 현대전 양상을 근본적으로 변화시키고 있음을 보여준다. 그러나 동시에 민간인 피해, 책임 소재 불분명, 사이버 공격 취약성 등 심각한 문제가 발생하고 있다. 본 연구는 미국 DARPA, 이스라엘 무인 드론 사례, 한국 국방과학기술혁신 기본계획을 검토하여 윤리·법제 기반, 기술적 안전성, 국제 협력, 민군 생태계 조성 측면에서 정책 제언을 제시하였다.

Ⅰ. 서 론

4차 산업혁명과 함께 인공지능(AI)의 발전은 군사 영역에 빠르게 확산되고 있다. 특히 AI기반 자율 소프트웨어는 실시간 상황 인식과 자율적 의사결정을 통해 무인기, 전차, 지휘통제 체계 등 다양한 분야에서 전투 효율성을 향상시키고 있다[1][2]. 하지만 자율율무기체계(LAWS, Lethal Autonomous Systems)의 등장으로 국제사회는 윤리·법적 논란과 책임성 문제를 직면하게 되었다. 한국 역시 국방 AI 추진 기본계획을 수립했으나[9], 아직 법제적·윤리적 체계는 미비한 수준이다[10][11]. 이에 본 연구는 최근 사례와 문헌을 분석하여 국방 AI 자율 소프트웨어 정책의 문제점을 분석하고, 한국이 나아가야 할 정책 과제를 제시하고자 한다.

Ⅱ. 연구 방법론

2.1 문헌 분석

최근 5년간 발표된 논문과 정책 보고서를 검토하여 AI 무기체계 신뢰성, 국방 AI 보증 체계, 보안 프레임 워크, 인권 및 윤리 문제, 국제 규범, 국방 AI 신뢰성, 군사적 적용, 국방부 정책 자료, 한국국방연구원 보고서를 검토하였다.

2.2 사례 분석

실제 전장에서 활용된 ▲우크라이나 전쟁의 드론 사례, ▲DARPA Mosaic Warfare, ▲이스라엘 하르피 드론을 중심으로 분석하였다.

Ⅲ. 사례 분석

3.1 우크라이나-러시아 전쟁 사례

우크라이나 전쟁(2022~)은 AI 기반 드론이 실제 전투에서 대규모로 활용된 첫 사례이다. 양측은 자율 소프트웨어를 통해 표적 탐지·공격을 수행했으나, 오작동으로 민간시설 피해가 발생했다. 이는 국제인도법 위반 논란을 불러왔고, 자율무기 사용에 대한 국제 규범 필요성이 대두되었다.

3.2 미국 DARPA의 Mosaic Warfare

미국 DARPA는 'Mosaic Warfare'를 통해 무인기·센서·자율 소프트웨어를 결합하여 네트워크 중심 전투를 실현하고 있다. 그러나 모의훈련에서 적대적 사이버 공격으로 네트워크 전체가 마비된 사례가 보고되었으며 이는 자율 소프트웨어가 지닌 보안 취약성을 드러낸다.

3.3 이스라엘 하르피 드론

이스라엘은 표적 탐지 후 자동 자폭하는 하르피 드론을 실전에 투입하였다. 나고르노-카라바흐 전쟁에서 전술적으로 성공했지만, 민간인 피해 가능성과 표적 판별 오류 문제가 제기되었다.

Ⅳ. 문제점 및 이슈 도출

국방 AI 자율 소프트웨어의 발전은 다음과 같은 문제가 제기된다.

첫째, 윤리·법적 문제로 민간인 피해 발생 시 책임 주체가 불명확하며, 국제인도법과 충돌한다[4][5][11]. 둘째, 기술적 취약성으로 적대적 공격으로 인한 인식 오류와 사이버 공격으로 인한 체계 마비 가능성이 크다[2][3]. 셋째, 정책·제도의 미비로 한국의 정책은 기술개발 위주에 머물러 있으며 법적·윤리적 기준이 미비하다[9][10]. 마지막으로 산업 생태계의 한계이다. 국내는 방산 대기업 중심의 폐쇄적 구조로 스타트업·학계의 참여와 민군 협력이 부족한 실정이다[7][8].

구체적인 해외정책사례는 다음과 같다.

구분	해외 정책 사례	시사점/ 적용과제
윤리·법제	-미국 국방부 DoD Directive 3000.09 (2023 개정): 모든 자율무기체계는 인간의 의미 있는 통제(MHC) 원칙을 준수해야 함 - UN GGE(2021~): 치명적 자율무기(LAWS)에 대해 국제인도법 준수 필요성 논의	- 한국 법령에 MHC 원칙 명문화 - 국방 AI 윤리위원회 설립 및 심의 제도화
기술적 안전성	- DARPA "AI Next Campaign" (2018~): XAI(Explainable Artificial Intelligence) 프로그램 운영, 설명 가능한 의사결정 모델 개발 - NATO 2021 AI 전략: 모든 군사 AI 시스템은 보안성·신뢰성 검증을 의무화	- 국방 무기체계 XAI 의무 적용 - 사이버 보안 모듈 및 적대적 공격 대응 알고리즘 국방 규격화
국제 규범	- EU AI Act (2024): 고위험군 AI(군사·안보 포함)에 대한 엄격한 규제 및 설명가능성 요구 - Stop Killer Robots 캠페인(2025): 전 세계 90여 개국 참여, 자율무기 금지·규제 조약 제안	- UN LAWS 협상에서 한국의 중재자 역할 수행 - 한미일·NATO 공동 선언 추진
민군 생태	- 이스라엘 IAI·Elbit Systems: 스타트업 및 민간 기업과 공동으로 드론·AI 무기체계 개발.	- 민군 공동 R&D 프로그램 확대

계	- 미국 DIU(Defense Innovation Unit): 민간 스타트업과 군 협력 가속화. - EDA(2025 White Paper): 유럽 내 국방 AI 신뢰성 검증 연구 지원.	- 국방 데이터 개방 및 AI 전문 인재 양성

V. 정책 과제

구분	도출 근거(사례·문헌)	정책 과제
윤리·법제	-민간인 피해 책임 불분명 -HRW Report[4][11]	-MHC 원칙 법제화 -국방 AI 윤리위원회 설립 및 심의 제도화
기술적 안전성	-DARPA Mosaic Warfare사이버공 [2] -Kapusta Framework [2][3]	-설명 가능 AI(XAI)의무 적용 -적대적 공격 대응 알고리즘 및 보안모듈 탑재
국제 규범	-UN LAWS 협상 지연[6] -Stop Killer Robots Report [5]	-UN 협상 적극 참여 및 중재 역할 수행 -한미일·NATO 공동 선언 추진
민군 생태계	-국방 AI 추진 기본계획 한계[9] -EDA White Paper[7][8]	-방산-스타트업-학계 공동 R&D 확대 -국방 데이터 개방 및 AI 인재 양성

본 연구에서는 다음과 같이 네 가지 구체적 정책 과제를 제시한다. 첫째, 윤리적 법제 기반의 확립이다. 의미 있는 인간 통제(Meaningful Human Control, MHC)'원칙을 「국방 AI 운영지침」과 국방 관련 법률에 명문화하고 국방부 산하 국방 AI윤리위원회를 신설하여 무기체계의 개발 단계별 사전 심의, 운용 중 발생할 수 있는 윤리적 문제에 대한 평가를 제도화하는 것이다. 이를 통해 국제인도법 준수를 강화하고 한국이 국제사회에서 책임 있는 AI 무기 사용국으로서의 신뢰를 확보할 수 있다. 둘째, 기술적 안전성 강화로 모든 AI 기반 무기체계에 설명 가능한 인공지능(XAI)을 적용하여 알고리즘의 의사결정 과정을 투명하게 검증하는 것이다. 적대적 공격에 대응할 수 있는 보안 알고리즘과 사이버 침해 대응 모듈을 국방 규격(MIL-STD)로 제정하는 것과 국방 R&D 단계에서부터 "레드팀(Red Team)" 검증 체계를 운영하여 실전 환경에서 발생 가능한 해킹·교란 시나리오를 사전에 실험하는 것이다. 셋째, 국제 규범 참여 및 선도가 필요하다. 한국이 UN 협상에서 중재자적 입장을 취해 조건부 규제 +MHC 의무화라는 절충안을 제안할 수 있다. 한미일 및 NATO와 협력하여 자율무기 운용 원칙 공동 선언을 추진하는 등 국제 규범 형성 과정에서 규범 선도국으로 자리매김 할 수 있도록 하고 ASEAN, EU 등과의 방산 협력 채널을 통해 양자 및 다자 협약체결을 선도함으로써 AI 군사기술을 둘러싼 국제 경쟁에서의 협상력을 확보할 수 있게 될 것이다. 마지막으로는 민군 융합 생태계 조성이다. 방산기업-스타트업-대학 간 공동 연구개발 프로그램을 정부 매칭펀드 방식으로 확대하고, 국방 데이터셋을 공유 가능 수준에서 민군 공동 활용할 수 있는 개방형 국방 데이터 플랫폼을 구축하는 것이 필요하며 또한 국방 특화형 AI 인재 양성 트랙(군사학+AI 융합 교육) 및 산학연군 연계를 통한 전문 인력의 파이프라인을 확보함으로써 국가 전체 민군 AI 생태계 발전의 확장과 민군 동반성장을 촉진할 수 있다.

VI. 결론

본 연구는 최근 전쟁 사례와 학술·정책 보고서를 토대로 국방 AI 자율 소프트웨어 활용에 제기되는 문제점과 정책 과제를 도출하였다. 그 결과 윤리·법제적 공백, 기술적 취약성, 국제 규범의 부재, 민군 생태계의 한계라는 핵심 문제가 확인되었다. 구체적으로 ▲MHC(Meaningful Human Control) 원칙 법제화, ▲XAI(Explainable Artificial Intelligence) 및 보안 모듈 의무화, ▲UN 및 다자 협상에서의 한국의 중재자적 역할 강화, ▲민군 융합 생태계 구축이라는 정책 과제를 제안하였다. 이러한 정책 과제는 단순한 기술적 대응을 넘어, 법·윤리적 정당성 -> 기술적 신뢰성 -> 국제적 정합성 -> 산업적 지속가능성이라는 연속적인 구조 속에서 추진되어야 한다. 특히, 한국은 국방 AI 정책을 통해 안보 강화와 동시에 산업 경쟁력 확보, 그리고 국제사회에서의 규범 선도국 위상을 달성할 수 있을 것이다.

따라서 본 연구의 함의는 국방 AI 자율 소프트웨어 정책이 더 이상 선택이 아닌 국가 안보와 산업, 국제 규범을 통합하는 전략적 필수 과제라는 점을 명확히 하는 데 있다. 그러나 주로 문헌 및 사례 분석에 기반하였기 때문에 실제 실행 단계에서의 실증적 계량적 검증은 한계가 있다. 따라서 다음과 같은 구체적 연구가 필요하다. ①법제도적 시뮬레이션 연구, ②XAI 기반 군사 시스템 평가 연구로 실제 무기체계(드론, 무인지상차량 등)에 XAI 를 적용했을 때 설명 가능성이 얼마나 확보되는지에 대한 실증연구, ③사이버 보안 취약성 실험연구, ④국제 규범 형성 과정 분석 연구, ⑤민군 융합 생태계 효과 분석 연구를 통해 실제 공통 R&D 프로젝트의 기술혁신과 산업 파급효과에 어떤 영향을 미치는지 효과 분석 연구가 필요하다. 이러한 연구는 제안된 정책의 실행 가능성과 성과를 실증적으로 평가하여 정책적 정합성을 강화해야 할 것이다.

ACKNOWLEDGMENT

이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원-지역지능화혁신인재양성사업의 지원을 받아 수행된 연구임(IITP-2025-RS-2020-II201612).
이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원-대학ICT연구센터(ITRC)의 지원을 받아 수행된 연구임(IITP-2025-RS-2024-00438430).

참 고 문 헌

[1] Cools, K., & Maathuis, C., "Trust or Bust: Ensuring Trustworthiness in Autonomous Weapon Systems," arXiv preprint, 2024, pp. 1-15.

[2] Kapusta, A. S., Jin, D., Teague, P. M., Houston, R. A., Elliott, J. B., Park, G. Y., & Holdren, S. S., "A Framework for the Assurance of AI-Enabled Systems," arXiv preprint, 2025, pp. 1-22.

[3] Tallam, K., "Engineering Risk-Aware, Security-by-Design Frameworks for Assurance of Large-Scale Autonomous AI Models," arXiv preprint, 2025, pp. 10-35.

[4] Human Rights Watch & Harvard Law School International Human Rights Clinic, "A Hazard to Human Rights: Autonomous Weapons Systems and Digital Decision-making," Human Rights Watch Report, 2025, pp. 5-28.

[5] Stop Killer Robots, "Policy Brief: Autonomous Weapons Systems – Key Issues and Path to a Treaty," Policy Brief, 2025, pp. 1-12.

[6] Lieber Institute, "Artificial Intelligence in Armed Conflict: The Current State of International Law," CyCon 2025 Series, 2025, pp. 30-52.

[7] European Defence Agency, "Trustworthiness for AI in Defence," EDA White Paper, 2025, pp. 1-18.

[8] Trends Research, "Governing Lethal Autonomous Weapons in a New Era of Warfare," Trends Research Article, 2025, pp. 40-60.

[9] 국방부, 「국방과학기술혁신 기본계획, 2023-2037」, 국방부 정책자료, 2020, pp. 3-25.

[10] 한국국방연구원(KIDA), 「AI와 국방전략의 연계: 미국의 국방혁신구상을 통해 본 한국의 미래전 대비」, 연구보고서, 2025, pp. 1-6.

[11] 한국국방연구원(KIDA), 「자율살상무기 국제 규범 형성에 관한 연구」, 미래국사학회, 2023, pp. 53-73.