

A Study on Responsible AI in the Military Domain

Seyong Kim*, Jeonghoon Park**, Junsang Kim***

*Lt. Col., Battle Command Training Program Group, Republic of Korea Army, Daejeon, Korea
**Captain, Battle Command Training Program Group, Republic of Korea Army, Daejeon, Korea
***Associate Professor, Dept. of Cyber Science, Korea Naval Academy, Changwon, Korea

[Abstract]

In this paper, we survey international efforts to introduce responsible AI into the military domain and suggest the countermeasures and development direction of the Korean military based on these efforts. Since AI technology in the military domain is bound to have a huge impact because it is related to the dignity of human life and is utilized in battlefield environments between nations. Therefore, it is necessary to prevent conflicts and issues arising from the introduction of AI technology in the military domain, and international consultations are ongoing for this purpose. From the perspective of our military, we must develop policies related to responsible AI so that the development of AI technology is not restricted while complying with international consultations. In order to achieve this research purpose, this paper first summarizes domestic and international issues and issues related to responsible AI and investigates international efforts made to date. In addition, we present four strategies reflecting the characteristics of military AI and considering the conditions and circumstances of the Korean military: defense data governance, legal and institutional framework, systematic AI evaluation mechanism, and creation of an open ecosystem.

▶ **Key words:** Responsible AI, AI ethics, Defense AI, Data governance, AI reliability

[요 약]

본 논문에서는 책임 있는 AI를 군사 영역에 도입하기 위한 국제적인 노력에 대해 조사하고 이를 기반으로 한국군의 대응방안 및 발전 방향을 제시한다. 군사 영역의 AI 기술은 인간 생명의 존엄성과 관련되어 있고 국가 간 전장 환경에서 활용되기 때문에 영향력이 매우 클 수밖에 없다. 그래서 군사 영역의 AI 기술 도입에 따른 갈등과 쟁점의 예방이 필요한 것이고, 이를 위한 국제적 협의가 계속되고 있다. 우리 군 입장에서는 국제적인 협의를 준수하면서도 AI 기술 개발에 제약이 되지 않도록 책임 있는 AI에 관련된 정책들을 개발해야 한다. 본 논문에서는 이러한 연구목적을 달성하기 위해, 우선 책임 있는 AI와 관련된 국내외적 이슈와 쟁점을 정리하고, 지금까지 진행된 국제적 노력을 조사했다. 그리고 군사 영역 AI의 특성을 반영하고 한국군의 여건과 상황을 고려하여 국방 데이터 거버넌스, 법적 및 제도적 틀, 체계적인 AI 평가 메커니즘, 개방형 생태계 조성의 네 가지 전략을 제시하였다.

▶ **주제어:** 책임 있는 AI, AI 윤리, 국방 AI, 데이터 거버넌스, AI 신뢰성

- First Author: Seyong Kim, Corresponding Author: Junsang Kim
*Seyong Kim (threedragon@mnd.go.kr), Battle Command Training Program Group, Republic of Korea Army
**Jeonghoon Park (pjh861234@gmail.com), Battle Command Training Program Group, Republic of Korea Army
***Junsang Kim (junsang@navy.ac.kr), Dept. of Cyber Science, Korea Naval Academy
- Received: 2025. 03. 05, Revised: 2025. 05. 23, Accepted: 2025. 05. 23.

I. Introduction

최근 AI 기술발전과 더불어 그 활용 범위가 사회 전 분야에 걸쳐 확대되면서 인간의 삶의 편의성과 함께 생산성도 향상되고 있다. 그 활용 범위는 민간 분야뿐만 아니라 군사 영역에서도 급속하게 확산하고 있으며 최근에는 러시아-우크라이나 전쟁, 이스라엘-하마스 전쟁에서 그 효과성을 입증하고 있다.

AI는 그 특성상 기술적 한계와 인간의 악의적 오용, 통제력 상실 등 수많은 위험 요소들을 내포하고 있다. 예를 들어, AI 시스템의 오작동이나 편향된 판단은 의도하지 않은 결과를 초래할 수 있으며, 유해정보 생성이나 사이버 해킹, 딥페이크 기술은 사회적 혼란과 안보 위협을 초래할 수 있다. 또한, 자율복제 AI, 초지능, 무기체계의 자율화는 기술적 통제력 상실로 이어질 수 있어, 군사적으로 활용할 시 돌이킬 수 없는 위협을 초래할 가능성이 있다.

AI의 정밀도와 정확도는 계속해서 발전하고 있지만, 아직 인간 수준의 판단력과 독립성을 가지지는 못한 실정이다. 그러므로 AI 기술이 적용된 무기체계가 군사적 활동에 개입하는 상황에서 발생하는 윤리적 문제들은 그 파급력이 매우 클 수밖에 없다. 예를 들면, AI 기반 경계감시시스템은 개인들의 프라이버시 침해 등의 부작용을 초래할 수 있고, 러-우 전쟁에서 사용된 ClearView 社의 AI 안면인식 기술은 민간인에 대한 무분별한 개인정보 사용 관련 법적 분쟁이 있다[1]. 또한, 이스라엘-하마스 전쟁에서 활용되고 있는 ‘라벤더’는 적군 식별 정확도가 90%에 이르지만, 오인식에 따른 민간인 피해 관련 논란이 존재한다[2].

군사 영역에서 AI 기술은 이제 인간을 단순히 보조하는 역할을 넘어 인간의 생사를 결정하는 수준까지 도달하고 있다. 그러므로 전장 상황에서 AI의 결정에 따라 전투를 수행했을 시, AI의 결정이 안전하고 신뢰할 수 있는지와 그 결과에 대한 최종 책임에 대한 국내외 논의가 다양하게 진행되고 있다. 본 연구에서는 이러한 AI 기술이 군사적으로 개발되고 활용되는 과정에서 발생하는 이슈 및 쟁점과 이를 해결하기 위한 국제적 노력과 시사점을 조사한 후 한 국군의 대응 및 발전 방향을 제안한다.

본 논문의 구성은 아래와 같다. 먼저 2장에서는 군사 영역에서의 책임 있는 AI 활용 관련 동향을 조사하고 3장에서는 주요 이슈와 쟁점을 분석한다. 4장에서는 주요 이슈와 쟁점에 대한 한국군 대응 방향을 제시하고 5장은 논문의 의의와 향후 연구 방향 등을 제시하며 결론을 맺는다.

II. Analysis of Responsible AI(RAI) Trends in the Military Domain

1. Definition and Basic Principles of RAI

AI가 도출하는 결과에 관한 실질적이고 구체화 된 우려는 알파고와 이세돌의 딥마인드 챌린지 이후 꾸준히 제기되어 왔다. 2017년 미국 캘리포니아 아실로마(Asilomar)에서 개최된 컨퍼런스에서 AI 개발과 목적, 윤리와 가치, 장기이슈 등을 다룬 총 23개의 준칙으로 이루어진 아실로마 AI 원칙(Asilomar AI Principles)이 그 시작이라 할 수 있다.

가장 앞선 국가 단위의 결과물은 영국 케임브리지대학의 실존 위기연구센터(Centre for the Study of Existential Risk)의 해악적 AI 보고서, 유럽연합의 신뢰할 수 있는 AI 윤리 가이드라인이 대표적이다. 해당 문건들은 AI 기술혁신 초기에 수행된 연구들로 AI 기반 기술의 설명 가능성에 주목하였다. 설명 가능성이란 AI 기술이 도출해 낸 결과가 어떤 절차와 원인에 의한 것인지 인간에게 이해할 수 있는 수준으로 설명할 수 있어야 한다는 것으로서, 인간이 AI의 결과에 대한 신뢰 여부를 판단할 수 있는 최소한의 근거를 보장하자는 취지였다. 그러나 미국 Amazon의 채용추천 AI에서 불거진 성차별 논란, 유니티티드 헬스그룹의 치료 우선순위 결정 알고리즘에서의 인종차별 등의 문제가 드러나면서 논의의 범위가 인간의 AI 도출 결과의 판단 권리 보장에서 AI 솔루션 자체의 윤리기준 마련으로 확장되었다. 현재에 이르러서는 이러한 논의를 AI의 신뢰성에 관련된 사안으로 바라보는 것이 일반적이며, 미국 국립표준기술 연구소의 ‘The American AI Initiative(2019)’, ISO/IEC의 ‘Overview of trustworthiness in AI(2020)’에서 정의한 AI 신뢰성에 대한 개념이 주로 활용되고 있다.

RAI는 AI 기술 활용의 윤리적, 사회적 및 법적 영향을 다루기 위해 도입된 개념으로 AI 기술이 포함된 응용체계를 개발, 운영, 배포하는 과정이 윤리적이고 투명하며 사용자 개인정보와 사회적 가치를 존중하는 방향으로 이뤄지는 것을 의미한다. RAI를 실현하기 위해서는 윤리성, 투명성, 공정성, 개인정보 보호 및 보안, 책임성의 원칙을 고려해야 하며 AI 기술이 인간의 가치와 복지에 부합하는 방식으로 적용되는지 검토해야 한다[3].

2. Trends in International Organizations

2.1 Convention on Certain Conventional Weapons Group of Governmental Experts (CCW GGE)

국제 협의체에서는 그동안 자율살상 무기체계(Lethal Autonomous Weapon Systems, LAWS) 관련 논의가 가장 활발하게 진행됐으며, 신형 안보 위협 관점에서 AI, 우주, 양자 등이 산발적으로 논의되던 중 최근 들어 AI에 관한 관심이 폭증하는 추세이다. 2013년에는 유엔(UN)의 특정 재래식 무기 금지조약(CCW)에서 LAWS 사용의 합법성에 대한 첫 논의를 시작했다. 2014년부터는 LAWS에 대한 윤리적이고 국제법적인 문제에 대해 논의해 오고 있고 2017년부터는 공식 정부 전문가 그룹회의(GGE)를 개최하여 국제적 합의를 도출해 나가고 있다. 2019년에는 LAWS의 개발이나 사용금지를 주장하는 국제인권단체나 3세계 국가들과 다른 결론을 도출하였는데, 군사 선진국들이 AI 기술을 활용한다면 전투원뿐만 아니라 민간인에 대해서도 피해를 줄일 수 있으므로 국제법상으로 LAWS를 금지하는 것을 배제한 것이다. 다만 LAWS로 인해 발생할 수 있는 부작용에 대해서는 인정을 하면서 이를 보완하기 위한 11가지 이행 원칙에 합의하였다. 주요 원칙으로는 LAWS 관련 기술을 사용할 때 국제인도법 및 적용 가능한 국제법적 의무 이행 여부를 고려하는 것과 LAWS 논의 시 CCW가 무기 사용의 인도주의적 요소와 안보적 요소를 모두 고려하는 적절한 틀을 제공하는 것이다[4].

CCW 조약의 당사국들은 지금까지 10년에 가까운 논의 끝에 LAWS를 사용하는 군대가 전쟁법을 준수할 수 있도록 '작전 기간과 지리적 범위, 작전 규모를 제한'할 필요가 있다는 부분에는 동의했다. 법적 구속력은 없지만, 적어도 죽음을 초래하는 어떤 과정에서는 인간이 언제 어디서나 그 역할을 해야 한다는 점도 명시했다. 다만 이러한 노력에도 국가 간 이견(법적 구속력 있는 새로운 협약 필요성, 자율무기 기술에 대한 적극적 규제 필요성 등)으로 LAWS 사용에 대한 국제적인 합의도출에 난항을 겪고 있다.

2.2 UN Security Council Briefing Meeting

2023년 영국 주최로 개최된 안보리 브리핑 공식회의(인공지능 : 국제평화와 안전에 대한 기회와 위협)는 AI를 단독으로 다룬 최초의 안보리 공식회의로, AI의 국제평화와 안전에 대한 잠재적 영향과 AI의 안전하고 책임 있는 사용 등을 논의하였다. 이후 영국은 AI 규제 계획을 포함한 백서를 발간하고 AI 관련 정상회의 개최 계획을 발표했으며 안보리 브리핑도 개최하였다.

해당 회의에서 유엔은 국제 AI 거버넌스와 규범을 논의

할 수 있는 최적의 기구이며, AI의 발전을 민간에 일임하는 것은 곤란하고, 유엔이 AI 거버넌스 체계 수립에 중심이 되어야 한다고 강조하였다. 세부적으로는 AI 기술이 로봇 등 군사적으로 활용되는 경우와 함께, 사이버 공격이나 가짜정보 확산 등에 이용될 가능성에 대한 논의가 중점적으로 진행되었다. 이사국들은 AI의 군사적 이용은 분쟁을 심화시키는 것이 아니라, 평화를 유지하는 기능을 수행해야 하며 이를 위해 AI 기술 이용에 특별한 주의가 필요하다는 것을 동의하였다. 또한, 자율살상무기에 대한 프레임워크의 필요성을 제기하고, 인간이 AI 기반 무기체계에 대한 최종 결정의 주체가 되어야 한다는 것을 합의하였다[5].

2.3 EU AI Act

유럽연합(EU)이 최종적으로 합의한 AI 규제법인 "EU AI Act"는 세계 최초로 포괄적인 AI 규제를 포함하였다. 그러나 프랑스 등의 강력한 요구에 따라 군사 및 안보 목적의 AI 사용에 대해 외부 용역업체를 포함한 광범위한 예외가 인정되었다. 이는 법 집행 당국의 적합성 평가를 거치지 않은 고위험 AI 시스템을 허용하는 내용을 포함한다[6].

군사 영역 AI의 예외를 적용한 구체적인 사유는 다음과 같다. 우선 군사 영역에서의 AI는 국가의 안보와 전략적 이익에 직접 연결되어 있고 군사기술은 국가 간 경쟁에서 중요한 역할을 하기 때문이다. 따라서 군사 영역 AI 적용은 국가의 전략적 이익을 고려해야 한다. 또한, 군사 영역에서 AI 기술은 일반 분야보다 더 빠르게 발전하고 있기 때문이다. 군사기관들이 자체적으로 연구개발을 수행하고 기술 경쟁력을 유지하기 위해서는 규제의 완화가 필요하다. 그래서 유엔은 군사 영역에서 AI 사용을 규제하는 것보다 국제적 협력을 우선순위에 두고 있다. 그래서 군사 영역에서 국제적인 표준과 협력을 유지하기 위해서는 규제를 유연하게 적용해야 한다. 이러한 이유로 유럽연합은 군사 영역을 AI 규제법에서 예외로 두었다.

2.4 REAIM(RAI in the Military Domain Summit)

2023년 UN에서의 AI 논의와는 별개로, 한국과 네덜란드 정부는 「AI의 책임 있는 군사적 이용에 관한 고위급 회의」를 최초 공동 개최하고, 60개국 참여 하 「행동계획(Call to Action)」을 발표하였다. 이는 법적 구속력은 없는 선언으로, 'AI의 책임 있는 군사적 사용' 관련 25개 고려사항을 제시하였다. 해당 문서는 미·중을 포함, 60개국이 참여했다는 긍정적 평가와 함께, 구체적 규범 마련은 하지 못한 평이한 내용으로 작성되었다는 부정적 평가가 공존하고 있다. 해당 선언의 전문에서는 AI의 군사적 이용

에 기회와 위험이 공존함을 지적하고, AI의 책임 있는 군사적 이용의 중요성을 강조했다. 본문에서는 AI의 군사적 이용의 기회·위험에 대한 이해 제고의 필요성 인식, AI 이용 결정에 대한 인간책임 보장 노력, 위험평가의 필요성 인식, 유관자 간 교환·정보 공유의 중요성 강조, 국제법에 합치하는 데이터 수집·이용 원칙 확인, AI 전체 수명주기에 걸친 세심한 주의, 보호장치 마련 및 인간 감시의 필요성 강조 등을 다루었다. 행동계획에서는 모든 이해관계자의 책임 부담 촉구, 다중 이해관계자들을 포함한 논의 지속 전념, AI의 군사적 이용 관련 정책·전략을 개발하도록 국가들을 초청, 민간 부문의 책임 있는 군사적 AI 이용 증진 촉구 등의 내용으로 작성되었다[7].

2차 회의는 2024년 9월, 한국에서 개최되었으며 결과문서로 '행동을 위한 청사진(Blueprint for Action)'을 채택했다. 총 61개국이 동참한 이번 채택문서는 군사 영역 AI 규범 마련을 위한 청사진을 제시하는 선언적인 성격의 문서로서 AI의 책임 있는 이용에 대한 구체적인 내용을 담고 있다. 동 문서에는 국제법 준수, 적절한 수준의 인간 통제 유지, AI에 대한 신뢰도 증진, AI에 대한 설명 가능성 개선 등 책임 있는 AI 이용에 필요한 원칙과 거버넌스 발전 방향 등이 제시됐다[8].

2.5 Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy

미국은 기술혁신을 저해할 수 있는 과도한 규제는 지양한다는 기초와 더불어 현존하는 국제법의 틀 속에서 AI의 책임 있는 군사적 이용을 공약하는 내용의 정치적 선언문을 마련하였다. 또한, 각국의 참여를 독려하기 위해 1차 REAIM 고위급 회의에서 「AI와 자율성의 책임 있는 군사적 이용에 관한 정치적 선언」 초안을 발표하고 주요국(한국 포함) 대상 의견 수렴을 거쳐 23년 11월 「AI와 자율성의 책임 있는 군사적 이용에 관한 정치적 선언」을 공식 발표하였다[7]. 해당 발표에는 총 51개국이 서명(한·미·영·프·독·일·호·캐 등 주요 서방국가 참여, 중·러 미참여)하였으며, 자발적 이행 중심의 연성규범 형태(법적 구속력 無, 정치적 구속력 有)로 참여국 확대를 위해 포괄적이며 광범위한 개념과 용어를 제시하였다[9].

이후 3개의 실무그룹(감독, 보장성, 책임성)을 조직하여 정기토의를 통해 그 성과물을 공유해 나갈 예정이다. 감독(Oversight) 그룹은 책임 있는 군사적 AI 역량 및 자율성 체계의 개발, 배치, 이용에 대한 국가 정책, 제도, 지도원칙 등 수립 및 이행(정치적 선언 A, B, C 조항 연관)을 담당하고 있다. 책임성(Accountability) 그룹은 AI의 개발·

배치·이용을 감독하거나 운용하는 인원의 책임성(Responsibility and Accountability) 이행 및 증진을 위한 정책 및 시행안(정치적 선언 E, F, G 조항 연관)을 다루고 있다. 또한 보장성(Assurance) 그룹은 총 수명주기에 걸쳐 군사적 AI 역량 및 자율성 체계의 신뢰성(Reliability)을 확보하고 위험을 감소시키는 설계·개발·배치·이용 정책 및 시행안(정치적 선언 D, H, I, J 연관) 분야를 담당하고 있으며 이러한 실무그룹들은 참여국 간 토의를 통해 합의점을 도출해 나가고 있다[10].

3. Trends in Major Countries

3.1 United States

미 국방부는 2018년 AI 전략을 발표하였고, 5대 추진전략으로 ① 인간 중심의 AI 적용 및 미국/동맹국 보호, ② 공유데이터 기반 클라우드 적용, ③ AI 인재육성, ④ 산·학 및 국제동맹들과의 협력체계 구축, ⑤ 국방 AI 윤리 및 안전성 선도를 제시하였다. 이를 달성하기 위해 비전과 핵심 원칙을 정립하여 AI를 법적이고 윤리적 방법으로 사용함으로써 자신들의 가치를 증진하고, 이를 위해 학계, 산업계, 국제적 동맹의 지도자들로부터 의견을 수렴하여 군사 영역의 안전과 AI 윤리의 발전을 위해 협의해 나가고 있다. 탄력적이고 강력하며, 신뢰할 수 있는 AI 기술개발을 위한 투자와 더불어 설명 가능한 AI 개발과 AI의 시험, 평가, 검증 및 인증에 대해서는 선제적으로 대응할 것이며, 상황인식, 결정지원 관련 기술을 고도화하여 AI에 의한 비고의적 손상이나 부수적인 피해를 감소시킬 수 있도록 지원할 예정이다[11].

미국 국방혁신위원회(Defense Innovation Board: DIB)에서는 전투원과 비전투원 모두에게 적용될 수 있는 윤리적이고 책임성 있는 AI 사용 원칙을 논의하였고 구체적인 윤리원칙을 수립하였다. 여러 번의 공청회와 전문가 자문회의, 고위급 실무회의 등을 통해 미 국방부의 AI 윤리원칙(책임성, 공정성, 추적 가능성, 신뢰성, 통제 가능성)을 공표하였다. 이러한 윤리원칙을 이행하기 위해 정책과 전략, 이행계획 등을 수립하고 적극적으로 추진하고 있다[12].

3.2 United Kingdom

영국은 AI 능력 글로벌 3위(1위 미국, 2위 중국) 목표로 특히 자율 무기체계 개발에 초점을 두고 있다. 영국은 2022년 국방 AI 추진전략을 발표하면서 부록으로 국방 AI 윤리원칙(인간 중심, 책임성, 이해 가능성, 신뢰성, 편향성 및 피해 최소화)을 구상하고 제정하였는데, Table. 1.과 같이 NATO의 책임 있는 AI 활용원칙과 매우 유사하다.

또한, 군사 영역에서 AI 도입 시 총 수명주기관리를 위한 정책을 수립(Dependable AI Joint Service Publication) 및 추진하고 있으며, AI의 잠재적 위험성 등 평가 및 투명한 기록관리를 위해 책임 있는 AI 위험평가 프레임워크를 만들어서 활용하고 있다[13][14].

Table 1. UK MOD AI Ethical Principles Vs NATO Principles of Responsible Use[13][14]

UK MOD	NATO
Human-Centric	Lawfulness and Governability
Responsibility	Responsibility & Accountability
Understanding	Explainability & Traceability
Bias & Harm Mitigation	Bias Mitigation
Reliability	Reliability

3.3 China

중국은 제3차 일대일로 국제협력 정상포럼에서 글로벌 AI 거버넌스 이니셔티브를 발표하였다. 중국 역시 군사 강대국들이 AI의 군사적 사용에 대하여 신중하고 책임 있는 활동을 해야 함을 강조한다. 다만 인간 중심 같은 ‘보편가치’ 보다 주권국 간의 동등한 권리와 기회에 초점을 두고 있어 AI의 개발 및 이용에 동등한 권리를 가져야 함을 주장하고 있다[15]. 이러한 중국의 글로벌 AI 거버넌스 이니셔티브 발표를 통해 서방과 중국 간의 글로벌 AI 거버넌스를 둘러싼 새로운 경쟁이 본격화되고 있음을 알 수 있다.

중국 또한 AI 수명주기에 윤리·도덕을 접목하여 AI 관련 업계에 종사하는 자연인·법인·기타 관련 기관을 위한 윤리적 지침을 제공하고 책임 있는 AI 연구개발 및 발전을 도모하고 있다. 과학기술정보통신부 산하 ‘국가 차세대 AI 거버넌스 전문위원회’는 6가지 기본 윤리적 요구사항(인간 존엄성 보장, 프라이버시 및 데이터 보호, 공정성 및 차별 금지, 투명성 및 설명 가능성, 신뢰성 확보, 책임성 강화 및 윤리적 소양 향상)이 담긴 차세대 AI 윤리 규범을 2021년 9월 발표하였는데 이는 주요국들의 AI 윤리원칙과 유사하다[16]. 중국의 AI 윤리 규범은 정부 부처, 기업, 대학, 연구소, 학회 등 관련 기관 또한 동 규범 및 기관별 현황에 근거하여 구체적인 윤리 규범과 관련 조치를 수립하도록 명시하였다. 중국은 한국에서 개최된 2차 REAIM 고위급 회의 참석과 관련해 "모든 당사자와 협력해 AI가 인류에게 더 나은 혜택을 줄 수 있도록 촉진할 것"이라고 밝혔다. 하지만 REAIM 선언문에 서명을 하지 않았기 때문에 향후 움직임을 주시할 필요가 있다.

3.4 Canada

캐나다는 2030년까지 군사 영역에서 디지털 변혁을 이루기 위해 AI를 전면적으로 도입하는 것을 목표로 주요 노력을 기울이고 있다. 이를 실행하기 위해 Fig. 1과 같이 세부수행과제인 AI 통합 및 배치, 변화 관리, 윤리, 안전 및 신뢰성, 인재양성 및 교육/훈련, 글로벌 파트너십 강화를 설정하였으며 특히, 윤리, 안전 및 신뢰성을 가장 중요한 요소로 판단하고 있다.

캐나다는 군사 영역에서 AI를 사용할 때 발생할 수 있는 잠재적 위험을 최소화하고, AI가 공정하고 책임 있게 사용되도록 보장하기 위한 원칙과 지침을 개발하고 있다. 이는 AI의 개발 및 사용과정에서 공정성, 투명성, 책임성을 포함한 윤리적 고려사항을 통합하는 것을 포함한다. 또한, AI 사용의 안전성을 보장하기 위해 AI 시스템의 설계, 개발, 배포 및 운영유지 단계에서 엄격한 안전기준을 적용하여 AI의 결정이 일관성과 신뢰성을 제고할 수 있도록 관련 지침을 마련하고 있다. 특히, AI의 군사적 사용은 인간의 판단을 보완하는 도구로 사용하며, 사람의 생명을 다루는 중요한 결정은 여전히 인간의 통제하에 이루어지도록 원칙을 제시하고 있다[17].

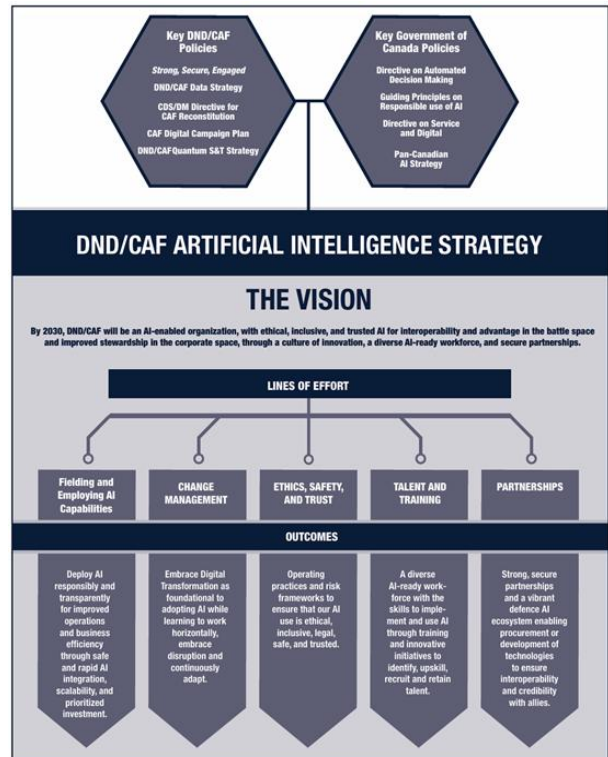


Fig. 1. Canadian Armed Forces AI Strategy[17]

3.5 Other Countries

네덜란드는 2023년 2월 제1차 REAIM 행사를 한국과 공동으로 개최하였으며, AI의 책임 있는 군사적 활용에 대한 관심이 많다. 다만 앞서 언급한 타 국가와는 다소 상반되게 AI가 인간의 통제를 벗어나지 않게 관리해야 한다는 측면에서 기술개발을 제약해야 한다는 입장이다. 또한, 전통적으로 이러한 국제적 논의나 레짐, 법제화 등에서 많은 활동에 적극적으로 참여하고 있다[18].

이스라엘은 최근 하마스와의 전쟁에서 AI 기술을 적용한 무기체계를 다수 활용하고 있으며, AI 기술이 국가 존립과 매우 밀접한 관계에 있다고 판단하고 있다. 그래서 국제사회의 군사적 AI 거버넌스 참여보다는 이스라엘군 자체 윤리지침 개발에 더욱 집중하고 있다[19].

독일은 AI 지휘결심체계 고도화 수행을 위해 DAIO(Defense AI Observatory) 조직을 신설하여 국방 AI 응용과 분석을 지원하고 있으며, 책임 있는 AI 추진을 위해 AI 기반 무기체계 도입 시 국제법(제네바협약 36조)을 준수하고, 연방정부 데이터 윤리위원회 운영과 NATO 규범 제정에 적극적으로 참여하고 있다[20].

4. Domestic Trends

한국은 2019년 12월 발표된 ‘AI 국가전략’을 통해 AI 윤리기준 및 실천방안 마련을 공론화함으로써 관련 논의를 국가 전체의 과제로 확대하였다. 그 이후 2020년 12월 국내 차원의 AI 윤리기준을 제시하였고, 2021년 5월에 이르러 관련 제도화 방향이 담긴 ‘신뢰할 수 있는 AI 실현 전략’을 발표하였다. 2021년 자료에서 특징적인 부분은 국내 차원의 AI 신뢰성에 관한 개념을 명확하게 제시했다는 점이다. 해당 자료는 AI 신뢰성을 “AI가 내포한 위험과 기술적 한계를 해결하고 활용·확산 과정에서의 위험·부작용을 방지하기 위한 가치 기준”으로 정의하였다. 특히 ‘내포된 위험’ 및 ‘기술적 한계’에 대한 언급을 통해 국제적인 AI 신뢰성 논의의 추세가 반영되었음을 알 수 있는데, 추진 전략의 세부 내용에 신뢰 확보 체계가 명시되었고 검증 및 인증체계에 대한 구상 또한 포함되었으므로 향후 국내 기업이 활용할 수 있는 인증체계가 도입될 것으로 예상할 수 있다. 실제 한국정보통신기술협회는 상기 전략에 대응한 AI 신뢰성 평가모델을 개발해 국내 기업들의 AI 모델과 AI 시스템에 대해 신뢰성 인증(Certification of AI Trustworthiness)을 2024년부터 시행하고 있다.

4.1 Progress on AI-related Legislation

국내 AI 산업진흥과 AI 신뢰 기반 조성을 위한 가장 ‘AI 기본법’은 21대 국회에서 활발히 논의가 이루어지다가 회

기 종료로 자동 폐기되었다. 22대 국회가 시작하면서 다시 활발한 입법 활동이 추진되다가 지난 2024년 12월 ‘인공지능 발전과 신뢰 기반 조성 등에 관한 기본법’(이하 AI 기본법)이 국회를 통과하여 2026년 1월 중 시행예정이며, 동법이 신속하게 시장에 안착할 수 있도록 하위법령·지침 마련 등 후속 조치도 조기 추진(2025년 상반기)할 계획이다.

AI 기본법은 AI에 관한 국가 차원의 거버넌스 체계를 정립하고, AI 산업을 체계적으로 육성하며, AI로 인해 발생할 수 있는 위험을 사전에 방지하기 위한 내용으로 구성돼 있다. 주요 내용은 3년마다 AI 기본계획을 수립, 국가 AI 위원회의 설치 및 운영, 국가 AI 센터 운영, AI 안전연구소 운영, 전문인력 양성·확보, AI 윤리원칙을 제정·공표, 고위험영역 AI 해당 여부에 대한 검토, 고위험영역 AI 제품 또는 서비스 제공 시 사전고지 의무화, 생성형 AI에 대한 고지 등의 내용이 포함되어 있다.

4.2 Trends in the Korean Military Domain

최근 들어 군사 영역에서의 RAI에 관한 관심이 높아지고 있다. 그동안 군사 영역에서 AI 적용에 관한 관심이 커지면서 관련 사업들의 계속해서 추진되고 있었으나, RAI라는 주제는 AI의 저변확대 이후라는 인식이 지배적이었다. 하지만 이후 국방부가 2024년 REAIM 행사를 주관하여 관련 분야의 국제적인 위상을 높이고 있고, 미국 주도의 협의체에도 적극적으로 참여하고 있다. 또한, 미 CDAO(Chief Digital and Artificial Intelligence Office)의 RAI팀과 협의 및 공동 연구를 추진하고 있으며 관련 정책연구를 여러 차례 발주하였다.

최근 3년 이내 수행된 군사 영역에서 책임 있는 AI 관련 정책연구는 Fig. 2와 같다. 해당 연구들은 큰 틀에서의 방향과 그림을 제시하는 수준이고, 현존하는 AI 신뢰성 기술과의 연결 방향 및 구체적인 합의 기준 등이 제시되지는 않았다. 그러므로 향후 AI 신뢰성 확보를 위한 기술적 요구사항과 정책적 고려사항에 관해 추가 연구를 진행할 필요가 있다[21].

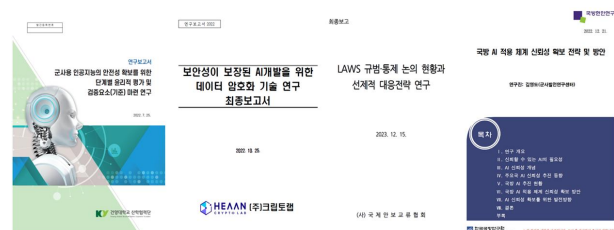


Fig. 2. Researches on AI Ethics and Trustworthiness in Military Domain[21]

III. Key Issues and Challenges Related to RAI in the Military Domain of the ROK Armed Forces

1. Difficulty in Establishing RAI Governance

군사 영역에 책임 있는 AI 도입을 위해 많은 연구 및 노력하고 있지만, 아직 법령 및 제도상 AI 도입을 위한 절차 반영이 미흡한 현실이다. 국방전력발전업무훈령 및 국방정보화업무훈령 상에는 AI 기반 전력 획득을 위한 내용이 부실하고, 일부 훈령 상에 학습용 데이터 관련 표준화 지침, AI 적용 무기체계 개발 시 데이터 확보 및 관리계획 정도만 반영이 되어 있으며, AI 기반 전력의 신뢰성, 안전성, 윤리원칙 준수 여부를 확인할 수 있는 정책 및 제도는 아직 마련되어 있지 않다. 국방부 전력정책국에서는 AI의 효율적 적용 및 활용(MLOps 중심)을 위한 국방획득체계 개선을 추진하고 있고 방위사업청은 AI를 포함한 SW 획득 시 Agile 기법 적용을 추진하는 등 국방 AI 획득제도 개선을 추진 중이나, 책임 있는 AI 즉, RAI에 대한 구체적인 언급은 되고 있지 않다. 민·군을 막론하고 AI 및 자율기술의 개발과 활용에 대한 중요성은 인정하고 있으나 국방 AI 윤리지침이 정립되어 있지 않은 상황에서 책임 있는 AI를 논의하기 부적절하다는 의견과 책임 있는 AI에 대한 명확한 정의의 부재로 인해 적용 대상 범위에 대한 이견이 존재한다. 특히, 책임 있는 AI는 기술혁신을 저해하는 규제적인 성격이 강하다는 의견과 AI 기반 전력의 안전과 신뢰성을 확보할 수 있다는 의견이 대립하고 있다.

또한, AI는 하나의 소프트웨어 기술로 특정체계로 분류하기가 어려움에도 불구하고 국방획득체계는 무기체계 및 전력지원체계(정보체계 포함)로 구분되어 획득절차가 달라 일관된 지침 및 통제에 어려움이 있다. 예를 들어 AI 기반 객체 인식 기능이 적용된 CCTV 도입 시 GOP 과학화경계시스템은 무기체계로 분류되고 해·강안 과학화경계시스템은 전력지원체계로 분류되어 획득 방법 및 절차가 상이한 등 일관성이 부족하다. 유사한 AI 기능이 적용되더라도 AI 획득절차, 시험평가, 운영유지 방식이 상이한 것이다. 국방부 내부적으로는 AI 및 자율무기에 대한 주요업무 주무부서가 분리되어 있다. 지능정보화국은 국방부 내 AI업무를 총괄하는 부서로 AI정책 및 국제협력 업무를 담당하고 국방정책기획국은 군비통제측면에서 자율살상무기에 대한 대응 및 RAI관련 국제레짐(International Regime) 업무를 담당하고 있다. 전력정책국은 AI 기반 전력 획득 정책을 수립하고 시험평가 업무를 총괄하며 첨단전력기획국은 AI 유무인 복합전투체계(MUM-T) 도입 및 국방 R&D 업

무를 수행한다. 국방획득체계의 이원화와 AI 도입을 위한 국방부 내 주무부서의 분리는 군사 영역의 RAI 관련 거버넌스 정립의 장애 요소로 작용하고 있다.

2. Issues in Legislation of the AI Basic Act

현재 추진되고 있는 입법 법안들을 보면 EU AI 규제법과 같이 군사 영역에 관한 예외조항이 있는 법안은 1개에 불과하다. 만약 관련 법안들이 통과된다면 규제에 초점이 맞추어져 국방 AI 산업의 기술혁신에 제한이 생기는 문제점이 있다. 또한 “고위험영역 AI” 범주에 군사용 무기체계의 포함 여부에 대한 이슈가 있다. 법령상 명확하게 군사용 무기체계에 대한 직접적인 언급은 없으나, “국민에게 영향을 미치는 국가기관 등의 의사결정”도 고위험영역 AI에 포함된다고 언급하고 있어 군사용 무기체계가 이 범주에 포함될 우려가 있다. 만약 고위험영역 AI로 분류 시 관련 내용의 고지 및 책무(위험관리방안 수립·운영, 이용자 보호 방안의 수립·운영, 고위험영역 AI에 대한 사람의 관리·감독, 신뢰성 확보조치 관련 문서작성 및 보관 등) 등으로 인한 많은 제약사항이 발생하고, 군사 영역에서 RAI 적용 관련 국제적 논의가 진행되고 있는 상황에서 해당 법률과 상충하는 이슈가 발생할 수 있다.

3. Opinions on the Degree of Participation in International AI Regimes

현재 군사 영역에서 RAI의 활용에 대한 국제레짐 및 법제화의 동참에 대해서는 여러 의견이 제시되고 있다. 먼저 적극적으로 동참하자는 찬성론자들의 의견은 AI의 군사적 활용 관련 국제적 규제가 가시화되는 상황에서 글로벌 RAI 환경조성에 주도적 역할을 수행하여 미래 우리의 군사력 건설에 이득이 되고 부정적 파급효과가 발생하지 않도록 사전에 방향성 형성에 관여하자고 주장한다. 또한, AI 관련 안전성과 신뢰성을 확보한 AI 기반 K-방산 무기체계가 세계를 리드할 수 있는 기회의 장으로 만들자는 의견을 제시하고 있다. 반면에 국제레짐 및 법제화에 신중하게 동참해야 한다는 신중론자는 선진국 대비 AI 기술 수준이 낮은 한국의 참여가 AI 기술발전에도 영향을 받을 수 있으며, 특히 군사 영역 RAI 관련 전담조직 및 전문가가 부재한 상황이기 때문에 신중한 접근이 필요하다는 의견을 제시하고 있다. 또한, 기존의 CCWG 자율살상그룹에서 논의되고 있는 바를 준수한다면 별도의 국제레짐화는 불필요하다는 의견도 있다. 결과적으로 우리 국익을 극대화하는 측면에서 우리 군의 AI 기반 전력 개발·활용에 부정적 영향은 최소화하고 우리의 기술 수준과 역량을 발전시키

기 위한 협력의 기회로 활용함과 동시에 국제적 요구에 부응하되 국익에 따른 전략적 관여와 실익의 확보를 할 수 있는 균형적 입장이 필요하다.

4. RAI Evaluation and Certification Issues

군사 영역에서의 RAI 평가 관련 가장 큰 문제는 세부적인 평가 기준 및 전담기관이 부재하다는 것이다. 또한, AI 특성상 AI 자체가 평가대상이 되기도 하지만 평가대상이 되는 무기체계의 일부분이 AI인 경우도 있는 등 여러 상황이 존재하기 때문에 기존방식의 평가 기준 등을 적용하기가 매우 어려운 게 사실이다. 또한, AI는 대부분 학습을 통해 성능이 구현되는데 이러한 학습은 양질의 데이터를 필수로 요구하게 된다. 그러나 군사 영역은 그 특성상 데이터의 수집 및 저장, 활용이 보안 등의 문제로 어려운 것도 사실이다. 그러나 이러한 문제점은 대부분 민간에서도 형태와 상황은 다르지만 고민하는 부분이 많으므로 적극적인 대외 교류 등을 통해 상당 부분 극복할 수 있다.

군사 영역의 AI 무기체계가 활용되는 작전(전투)환경의 다양하고 복잡한 특성은 AI의 기술적 한계가 나타나는 상황을 초래할 수 있고 이러한 문제가 발생하면 책임소재에 대한 이슈가 발생할 수 있다는 점도 이러한 평가 기준 등을 조기에 정립하지 못하는 이유 중 하나로 볼 수 있다. AI 무기체계의 시험평가 발전방안('22년, 합참)에서는 무기체계를 AI 무기체계와 일반 무기체계로 구분해야 하며 AI 무기체계의 시험평가를 신뢰성 개념과 책임 있는 AI 개념을 적용하여 개선할 필요가 있으며 표준화된 평가절차를 마련하고 인증제도의 발전이 필요하다고 제시하였다. 이를 위해 전담기관 및 조직 구성이 요구되며 이는 RAI 임무 수행을 위한 전문인력 확보 및 양성을 모두 포함하는 것이다. 이러한 평가 기준과 전담조직 등이 구성된다면 자연스럽게 지금 추진 중인 RAI 평가 및 인증에 대한 이슈는 어느 정도 해소될 것으로 판단된다.

IV. Direction for Response by the ROK Armed Forces

AI의 윤리적 문제는 광범위의 기술윤리 문제와 깊이 연결되어 있다. 데이터 편향성, 책임 소재 불분명, 일자리 감소 우려 등 AI로 인한 윤리적 쟁점들은 개인정보 침해, 디지털 격차 심화와 같은 기술 전반의 윤리적 문제와 맥락을 같이 하므로 AI 윤리 문제 해결은 기술윤리라는 큰 틀 안에서 함께 논의되고 모색되어야 한다[22]. 이러한 윤리 문

제에 대한 통합적 검토와 앞서 살펴본 추진 동향, 주요쟁점 및 이슈의 분석, 그리고 국방부에서 발주한 「안전하고 신뢰할 수 있는 AI 활용 국방 가이드라인」 정책연구를 통해 수립된 군내 관계자 및 관련 기관 전문가들(국방부, 각 군 본부, 방사청, 국방과학연구소, 한국국방연구원, 방산업체 관계자 35명)의 의견을 바탕으로 4가지 측면에서의 한국군의 대응 방향을 제시하였다.

1. Establishing AI-Data Governance in Defense

군사 영역에서 책임 있는 AI 구현을 위해서는 첫 번째로 군사 영역 데이터 수집 및 분석과 AI 적용에 관련된 거버넌스 정립이 필요하다. 이를 위해서는 군사 영역에서 신뢰할 수 있는 AI 개발 및 활용을 위한 RAI 정책 및 전략을 수립하여 체계적으로 추진할 수 있는 발판을 마련해야 한다. 또한, 군사 영역 RAI 정책 및 전략 수립 시 AI 시스템의 설계, 개발, 시험평가, 운영유지 등 생애 주기 전반에 걸쳐 AI의 안전성, 신뢰성, 윤리원칙 준수 여부 등에 대해 모니터링할 수 있는 시스템과 책임 체계를 구축하여야 한다.

두 번째는 또한 여러 국방 관련 기관에 흩어져있는 RAI 관련 조직 간의 임무를 재정비하고 이를 제어할 수 있는 추진체계를 정립하여야 한다. 앞서 언급한 바와 같이 국방 획득체계의 이원화된 고질적인 문제를 해결하는 것은 AI 전수명주기에 걸쳐 안전하고 책임 있는 AI의 국방 활용을 위해서는 꼭 필요한 과제이다. 현재 어느 기관(부서)에서도 주도하지 못하고 있는 상황에서 관련 기관/부서 간의 임무와 역할 등을 고려하여 추진체계를 정립해야 한다. 추진체계 정립 전에는 국방부에서 운영하는 국방 AI 위원회를 활용하는 방안도 고려해야 한다.

세 번째는 군사 영역 AI 윤리원칙을 제정해야 한다. 군사 영역에서 AI 윤리원칙이 부재하다는 것은 책임 있는 AI를 군사 영역에 적용하면서 기준점이 없다는 것과 같다. 미국 및 영국 국방부에서도 RAI를 추진하기 위해 기본적으로 AI 윤리원칙을 수립하였으며 이를 바탕으로 신뢰성, 책임성 등을 확보하기 위한 구체적인 방안들을 추진하고 있다. 미 국방부의 경우 CDAO의 RAI팀에서 RAI Toolkit을 작성하여 운영하고 있으며, NATO 및 주요국과 함께 이를 발전시켜 나가고 있다.

마지막으로, 군사 영역 RAI 관련 업무를 조정 통제할 수 있는 전담조직을 설립하거나, 국방부 차원의 전담부서를 신설하여 운영하여야 한다. 신설된 전담부서(조직)는 군사 영역 RAI 정책의 수립지원, 구체적 계획의 수립 및 실행, 정책 이행 과정에서 발생하는 문제점 및 개선 사항의 수집과 반영, 안전하고 신뢰할 수 있는 국방 AI 활용

지침의 준수 여부의 평가 및 감시 등의 업무를 수행해야 한다.

2. Removing Unnecessary Barriers through Legal and Institutional Reforms

책임있는 AI의 요구사항을 무기체계의 개발과 획득과정에서 반영하는 것은 일종의 규제가 추가되는 것이므로 사업 추진에 제약사항이 발생하고 기술개발과 도입과정의 지연을 발생시킨다. 이를 보완하기 위해서는 기술개발과 획득에 관련된 기존 법령 및 제도를 정비하여야 한다.

첫째, AI 기술이 적용된 국방 전 영역(무기체계, 전력지원체계) 획득절차를 간소화하고 유연하게 만드는 것이 필요하다. 또한, 민간에서 AI 도입을 위해 활용하고 있는 DevOps 또는 MLOps 등의 도입과 Agile 방식의 획득 및 운영체계를 군사 영역에서도 적극적으로 도입하여 절차적인 부분을 개선해야 한다. 이를 통해 최신 AI 기술이 적용된 전력의 신속한 확보가 가능하다.

둘째, 신규 개발 및 도입된 AI 기술을 빠르게 적용 및 테스트할 수 있는 국방 규제 샌드박스 제도 도입이 필요하다. 군사 영역 AI 관련 법령 개정, 제도 마련과 같은 규제 정비는 오랜 기간이 소요되므로, 국방 AI 분야의 개발 및 운영에 규제 샌드박스를 적용하여 민간의 최신 AI 기술이 군사 영역에 선제적이고 도전적으로 적용될 수 있는 제도적 기반을 마련해야 한다.

셋째, 안전하고 신뢰할 수 있는 AI를 군사 영역에서 개발하고 활용하기 위한 RAI 가이드라인을 작성 및 배포하여 기준점을 제시해야 한다. 군사 영역은 민간과 다른 환경에서 운영되는 특성이 있기에 별도로 가이드라인이 필요하다.

넷째, 군사 영역에서 AI의 설계, 개발, 시험평가, 운영유지 등 생애 주기 전 과정에서 지켜야 할 원칙과 기준을 제시하여야 하며 이를 확인할 수 있는 한국형 RAI Toolkit을 개발하고 제도화하여야 한다. 앞서 제시한 가이드라인을 기반으로 사용자의 목적에 맞게 활용할 수 있는 Toolkit을 개발하여 제공한다면, 각군, 관련 기관 및 기업에서 추진하는 군사 영역의 AI 관련 사업에 RAI를 손쉽게 적용할 수 있다.

3. Establishing a Systematic Foundation

군사 영역에 활용 가능한 AI 서비스를 발굴 및 사업화를 위한 기반체계를 구축하기 위해서는 우선 국방의 중장기 발전전략과 현재의 기술 수준 등을 고려한 국방 AI 데이터 로드맵을 수립해야 한다. 특히 미래전의 작전개념과

장기무기체계 발전계획을 뒷받침할 수 있는 AI 기술개발을 위한 양질의 데이터를 확보하기 위한 발전 로드맵이 필요하다.

둘째, AI 기술발전수준을 고려하여 국방 분야 AI 사업을 기획하고 이를 단계적으로 적용해 나가야 한다. 현재 민간에서 시범 적용 및 활용하고 있는 TRL(Technology Readiness Levels) 5 이상의 AI 기술은 국방 영역에 선제적으로 적용하여 성과를 확대해 나가고, 미래 전력에 필요한 AI 기술(TRL 4 이하)은 연구개발을 통해 국방 영역에 적용할 수 있도록 핵심 사업으로 분류하여 발전시켜야 한다.

셋째, 군사 보안 문제와 데이터 확보의 어려움을 해결하기 위하여 국방부 내부 네트워크에서 AI 자원을 제공하는 클라우드 기반의 플랫폼을 구축하여야 한다. 군사 영역에 AI가 체계적으로 개발 및 운영되기 이전까지는 일관된 정책과 프로세스, 표준화 등을 고려하여 국방부 주도로 전군을 지원해 줄 수 있는 플랫폼을 구축하여 전략적으로 활용하여야 하며 점차적으로 미래 작전영역을 고려하여 엣지 클라우드 형태로의 발전을 고려해 나가야 한다.

넷째, AI 과학기술 강군을 구현하기 위해서 군 AI 핵심 인력을 조기에 양성 및 확보하여야 한다. 국방 전 영역에서 AI의 적용 및 확산이 가속화되면, 이를 지속적으로 개발하고 관리할 수 있는 인력이 필요하므로, 미래 국방의 핵심전력으로써 데이터·AI 인재를 선제적으로 확보하고 운영할 수 있도록 해야 한다.

4. Creating an Open Ecosystem for Defense AI

RAI 업무는 국방부 내·외부 전문가들의 집단지성을 모아서 발전시켜야 한다. 이를 위해 먼저 군사 영역 RAI 관련 국내 민·관·산·학·연 등과 연계된 생태계를 구축해야 한다. 이를 통해 RAI 기술의 발전과 실용화를 촉진하며, 민간의 우수한 아이디어와 기술을 군사 영역에 도입할 수 있는 기반을 마련할 수 있으며, 군사 영역 RAI 분야의 국제적 이슈에 대해 국내 대응체계를 구축하는 데 큰 역할을 할 것이다.

둘째, 현재 주요 국가를 중심으로 자국의 AI 기술을 보호하기 위한 국제적 움직임이 강화되고 있는 상황에서, 다각도의 국제협력을 통한 공동 대응체계를 구축해 나가고 한국군이 주도적인 역할을 할 수 있도록 선제적으로 대응해 나가야 한다. 특히, 우방 국가인 미국과 한-미 RAI 협의체를 구성하여 양국 간의 기술 교류와 공동 연구를 통해 상호 발전방안을 모색할 필요가 있다. 현재 미 국방부에서 추진 중인 RAI Toolkit의 2, 3단계 개발에 공동으로 참여하여 상호 협력할 수 있는 기반을 마련하여 AI 분야에서

우방국으로의 협력을 공고히 할 필요가 있다. 또한, RAI 관련 국제협력 회의에 적극적으로 참여하여 AI의 개발 및 활용에 부정적 영향은 최소화하면서 국익을 극대화할 수 있는 방향으로 유도해야 한다. 이를 통해 글로벌 RAI 환경 조성에 주도적인 역할을 하여 K-방산의 해외 진출에 기여할 필요가 있다.

V. Conclusions

본 연구는 군사 영역에서의 책임 있는 AI의 도입과 활용이 AI 기술 발전과 함께 필수적으로 요구되는 시대적 과제를 먼저 설명하고, 국내·외에서의 군사 영역 RAI 관련 동향을 살펴보았다. 또한, RAI를 군사 영역에 도입하는 과정에서 생기는 이슈/쟁점들과 이를 고려한 한국군의 대응 및 발전 방향도 함께 제시하였다.

군사 영역에서 AI는 인간 생명의 존엄성과 관련되어 있고, 국제적 합의가 요구되는 등의 특수성으로 인해 기술적 신뢰성과 투명성을 담보하는 명확한 추진전략 확보가 필수적이다. 또한, 책임 있는 AI에 관련된 정책은 윤리적 문제를 고려하기 때문에 결국 국방 영역의 AI 혁신의 범위를 제약할 수밖에 없다. 본 논문에서는 이러한 문제를 고려하여 국방 AI-데이터 거버넌스 정립, 법령 및 제도 정비, 체계적 기반 체계 구축, 개방형 생태계 조성의 4가지 측면에서 책임 있는 국방 AI 활용 전략을 제시하였다. 제안한 전략을 우리 군의 정책에 반영한다면 이러한 제약을 최소화하면서도 국제 요구 수준에 적합한 AI 원칙을 수립할 수 있다. 결국, 우리 군의 AI 역량 개발과 활용에 미치는 부정적 영향은 최소화하고, 국제적 요구에는 능동적으로 대응할 수 있는 정책적 제안이라고 판단된다.

향후 본 논문에서 제시한 방안들이 구체적 논의와 정책적 실행으로 이어지기를 기대하며, 최근 입법된 AI 기본법을 기반으로 2025년 개최된 AI 행동 정상회의(AI Action Summit)와 국제 AI 표준 서밋 등에서 논의되는 결과들을 바탕으로 발전된 정책을 제안하고자 한다.

REFERENCES

- [1] European Data Protection Board. "Dutch Supervisory Authority Imposes a Fine on Clearview Because of Illegal Data Collection for Facial Recognition." EDPB, 5 Mar. 2024, www.edpb.europa.eu/news/national-news/2024/dutch-supervisory-authority-imposes-fine-clearview-because-illegal-data_en.
- [2] Beaumont, Peter. "Israel's Military Uses AI System to Select Targets for Airstrikes in Gaza." *The Guardian*, 3 Apr. 2024.
- [3] PwC. "Response to the Recent U.S. Executive Order on 'Responsible AI'." Samil PwC Monthly Newsletter, Dec. 2023, p. 12.
- [4] UN Office at Geneva. Report of the 2019 Session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems. CCW/GGE.1/2019/3, 25 Sept. 2019.
- [5] UN Security Council. "Briefing Artificial Intelligence: Opportunities and Risks for International Peace and Security." Jul. 2023.
- [6] Yoon, Jung-Hyun, and Eun-Jung Cho. "The Significance and Implications of the Passage of the EU AI Regulation Act." *INSS Issue Brief*, vol. 527, Mar. 2024, pp. 5-7.
- [7] Netherlands Ministry of Foreign Affairs. "Call to Action on Responsible Use of AI in the Military Domain." 2023.
- [8] ROK Ministry of Foreign Affairs. "Outcome of Responsible AI in Military Domain (REAIM) Summit 2024." Sep. 2024.
- [9] U.S. Department of State. "Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy." *State.gov*, www.state.gov/political-declaration-on-responsible-military-use-of-artificial-intel-ligence-and-autonomy-2.
- [10] Ganschow, Jan. "First Plenary Meeting of the PD on Responsible Military Use of AI and Autonomy." Mar. 2024.
- [11] U.S. Department of Defense. "DoD Data, Analytics, and Artificial Intelligence Adoption Strategy." Jun. 2023, pp. 8-10.
- [12] U.S. Department of Defense. *Ethical Principles for Artificial Intelligence*. 24 Feb. 2020.
- [13] Moore-Bick, Chris. "The UK Approach Implementing Responsible AI in Defence." Mar. 2024.
- [14] Cyberspace Administration of China. "Global AI Governance Initiative." Oct. 2023.
- [15] Choi, Kye-Young. "Global AI Governance Discussions and International Cooperation." *KISDI Premium Report*, Jun. 2023.
- [16] Department of National Defence and Canadian Armed Forces. "Artificial-Intelligence Strategy." Mar. 2024.
- [17] NATO. *Principles of Responsible Use for Artificial Intelligence in Defence*. 23 Oct. 2023.
- [18] People's Republic of China. *Global AI Governance Initiative*. 18 Oct. 2023.
- [19] National New Generation Artificial Intelligence Governance Professional Committee. *Ethical Norms for the New Generation Artificial Intelligence*. Ministry of Science and Technology of the People's Republic of China, 25 Sept. 2021.
- [20] Department of National Defence, Canada. *The Department of National Defence and Canadian Armed Forces Artificial Intelligence Strategy*. 2021.
- [21] Ministerie van Defensie. "Defensie Strategie Data Science en AI

2023-2027." Aug. 2023, pp. 3-15.

- [22] Raska, Michael. "Israeli Forces Display Power of AI, but It's a Double-Edged Sword." The Straits Times.
- [23] Borchert, Heiko, et al. "Master and Servant: Defense AI in Germany." The Very Long Game: 25 Case Studies on the Global State of Defense AI, Springer Nature Switzerland, 2024, pp. 195-216. DOI: 10.1007/978-3-031-58649-1_9.
- [24] Kim, Baekjung, and Seokjae Jeong. A Study on the Improvement of the Test and Evaluation System for Defense Weapon Systems. Journal of the Korea Institute of Defense Technology, vol. 5, no. 2, 2023, pp. 1-9.
- [25] Floridi, Luciano, et al. "AI for Human Flourishing: Beyond Principles." AI and Society, vol. 33, no. 4, 2018, pp. 689-706

Authors



Seyong Kim received the M.S degrees in Operation Analysis from Korea National Defense University, Korea, in 2009. Mr. Kim is currently a Lt. Col. in the Republic of Korea Army. He is interested in AI, big

data, cloud computing, and M&S.



Jeonghoon Park received the M.S. degrees in statistics from Purdue University, US, in 2023. Mr. Park is currently a Captain in the Republic of Korea Army. He is interested in AI, Machine Learning, big data, statistics,

and M&S.



Junsang Kim received the B.S., M.S. and Ph.D. degrees in Computer Science and Engineering from Hanyang University, Korea, in 2003, 2005 and 2017, respectively. Dr. Kim is currently an associate professor of

Department of Cyber Science at Korea Naval Academy. He is interested in AI, big data, cloud computing, and blockchain.