



Received: 2026/03/04
Revised: 2026/03/10
Accepted: 2026/03/24
Published: 2026/03/31

***Corresponding Author:**

Jong Lark Son
271, Jaun-ro, Youseong-gu, Daejeon,
Republic of Korea
Tel: +82-42-878-2422
E-mail: sone7174@naver.com

해군 AI 기반 무기체계 신뢰성 강화를 위한 시험평가 발전 방안에 관한 연구

Abstract

본 연구는 미래 다영역전장 환경에서 AI 무기체계 운용 선진국인 미국의 AI T&E와 RAI Toolkit, 영국의 FATE 등의 사례를 참고하여 시험평가 기준과 안정적인 평가절차 등 AI 기반 무기체계 신뢰성 강화를 위한 시험평가 발전방안을 제시하였다. 먼저 AI 기반 무기체계 특성을 고려한 AI 모델, 주체계에 대한 단계적 시험평가 수행 등 프레임워크를 확립해야 하며, 또한 시험평가에 필요한 물리적·가상적 인프라를 구축, 시험평가 인력의 전문성 강화, AI 기반 무기체계의 윤리적·법적 위험 완화 조치와 책임성 확보 등을 제안하였다.

This paper presents a test evaluation development plan to strengthen the reliability of artificial intelligence (AI)-based weapon systems, such as test evaluation standards and stable evaluation procedures, with reference to cases such as AI T&E in the United States, RAI Toolkit, and FATE in the UK in the context of future multi-domain operational environment. First, a framework must be established, such as an AI model that considers AI-based weapon system characteristics and a phased test evaluation of the main system. Additionally, we propose establishing the physical and virtual infrastructure necessary for test evaluation, strengthening the expertise of test evaluation personnel, and implementing ethical and legal risk mitigation measures and securing accountability for AI-based weapon systems.

Keywords

인공지능(AI), AI 기반 무기체계(AI-Based Weapon System), 시험평가(Testing & Evaluation), 시험평가 프레임워크(T&E Framework)

Development Plan of a Test Evaluation to Strengthen the Reliability of Navy AI-Based Weapon Systems

손종락*

해군 중령/해군대학 해양전략·전략학처 무기체계교관

Jong Lark Son*

CDR, ROK Navy/Instructor of Naval Weapon System. Dept. of Maritime Strategy & Force Studies, ROK Navy College

1. 서론

인도-태평양과 한반도를 둘러싼 대내·외 안보환경은 급속도로 변화하고 있다. 2026년 1월 미국 국방부가 공개한 국가방위전략서(2026 미 NDS)는 NATO, 일본, 한국 등 동맹국들의 안보 무임승차를 지적하며 동맹의 성격을 ‘보호’에서 ‘거래’로 전환함으로써 기존 국제 안보질서를 흔들고 있다[1]. 특히, 한·미동맹은 한국의 대북 억제 주도 와 동맹 부담분담(예: GDP 5% 지출) 논의가 핵심으로 제기되며, 북한의 재래식 전력에 대한 자주국방을 강조하고 있다[2]. 2026년 1월 미국의 베네수엘라 마두로 대통령 생포작전, 3월 핵 협상 결렬에 따른 이스라엘-미국-이란 전쟁은 4년간 지속되고 있는 러시아-우크라이나전쟁(이하 러-우전쟁)과 2025년 종식된 이스라엘-하마스분쟁(이하 이-하분쟁), 이란 핵시설 폭격 등 전 지구적으로 전쟁의 일상화를 보이고 있다. 한편, 북한은 지속적인 핵·미사일 위협 고도화와 최신형 구축함, 잠수함, 순항미사일, 무인기 등 재래식 전력도 강화하고 있다. 반면 대한민국은 K 방산 수출, 이지스구축함 및 KF-21 확보, 미국 함정 MRO 사업 등을 내실있게 추진하고 있으나, 인구절벽에 따른 병역자원 감소로 군사대비태세 저하가 현실화되고 있다.

이와 같은 대내·외 안보환경 속에서 우리 군은 AI 기반의 첨단전력 확보, 유·무인복합체계 구축에 역량을 집중하고 있다. 이-하분

쟁, 러-우전쟁은 본격적으로 AI 무기체계가 전쟁의 양상을 좌우하는 게임체인저이자 핵심요소로 주목받고 있다. 각국은 드론을 활용하여 감시·정찰·타격 등 주요 작전임무를 수행하고 있으며, Clearview AI, Gospel, Lavender & Where's Daddy 등 첨단 AI 기술을 활용하여 전장의 주도권을 장악하기 위해 국가역량을 집중하고 있다. 북한도 러-우전쟁 파병을 통해 무인전력 확보와 운용능력 향상을 추진하고 있을 것으로 예상된다. 대한민국은 2025년 9월 안규백 국방장관이 '50만 드론전사 양성'을 선언하며 장병 누구나 드론을 다룰 수 있는 시대를 열어 군 전투력 강화와 민간 산업 활성화를 동시에 추진하겠다고 선언했다[2].

이에 따라 우리 군은 AI·무인·로봇 등 첨단 과학기술 기반 'AI 과학기술 강군' 육성을 목표로 AI 기반 무기체계 전력화를 위해 노력하고 있다. 미래 전장의 주역으로 등장할 AI 기반 무기체계 도입은 신뢰성, 책임성, 안전성 등 윤리적 문제를 내포하고 있으며, 어떻게 시험평가(test & evaluation) 할 것인가에 대한 명제를 제시하고 있다[3].

구체적으로 기존의 '합격·불합격' 중심의 정략적 평가에서 능력과 임무 위주의 '가치 중심' 평가와 맥락·의도 기반 판단 검증, 신뢰성·적응성 평가, 인간-AI 협업 구조를 고려한 평가 등 다양한 요구사항이 존재한다.

그러나, AI 기반 무기체계 시험평가와 관련된 규정과 규정은 담론 차원에서 논의되고 있으며, 국방부 및 각 군의 AI·유·무인복합체계의 경쟁적인 도입과 병행하여 제도화되지 못하고 있는 현실이다. 육군은 다목적무인차량('24~'25년)과 AI 기반 전술차량용 지능형 RCWS(remote controlled weapon stations, '23~'25년) 개발사업과 연계된 시험평가 방안과 규정화 작업을 활발하게 추진하고 있으나, 해군은 상대적으로 미흡한 실정이다.

따라서 본 연구는 미래 다영역전장 환경에서 AI·무인체계 시험평가 기준과 안정적인 평가절차 등 AI 기반 무기체계 신뢰성 강화를 위한 시험평가 발전방안을 제시하고자 한다.

2. 시험평가(T&E) 이론적 고찰

2.1 시험평가의 정의, 역할

시험평가는 무기체계 개발 및 획득과정의 한 분야

로 해당 무기체계에 대한 성능, 기술, 품질, 운용관리 측면에서 제반 요구조건 및 개발목표의 충족 여부와 운용적합성, 효율성 등을 확인 검증하는 절차이다[3].

시험(test)은 개발 및 운용 측면에서 대상 체계의 객관적 성능을 검증하고 평가하기 위한 기초자료를 획득하는 과정을 의미한다. 평가(evaluation)는 시험을 통해 수집된 자료 또는 시뮬레이션 자료, 연구자료 등 기타 수단을 활용하여 획득한 자료를 근거로 평가기준과 비교 분석하는 과정이다[3].

시험평가는 일반적으로 개발시험평가(DT&T, development test & evaluation)와 운용시험평가(OT&E, operation test & evaluation)로 구분되어 수행된다. 개발시험평가는 연구개발주관기관이 표준화된 시험환경에서 요구성능을 측정하고 설계상 중요한 문제점을 식별하고 보완하여 기술적으로 개발목표가 충족되었는지 검증하는 것이다. 운용시험평가는 소요군의 운용자에 의해 무기체계가 운용되는 작전환경 또는 유사한 환경에서 작전요구성능(ROC)의 충족과 군 운용의 적합성, 효율성, 안정성 등을 확인하고 전력화 지원요소 등에 대한 적합성을 시험평가하는 것이다[3].

2.2 AI 기반 무기체계 시험평가 개선 필요성

일반적인 무기체계 시험평가는 시제품에 대하여 작전운용성능의 충족성, 합동성·상호운용성, 군 운용의 적합성, 전력화지원요소의 실용성을 확인하여 '전투용 적합' 또는 '전투용 부적합' 판정 자료를 제공하기 위해 소요군(해군: 전평단 시험평가처)이 주관하여 수행한다.

일반 무기체계의 시험평가 수행은 무기체계를 운용하고 통제하는 주체가 사람이기 때문에 실제 운용 환경 하에서 대상 무기체계를 야전에서 직접 탑승하여 조작과 기동을 하면서 작전지역의 극한 환경 조건에서 무기체계의 운용능력을 비롯한 각종 센서의 능력, 최대속력, 기동능력 등을 직접 확인한다[4].

반면 AI 무기체계는 다양한 시험평가 제한사항이 있다. 첫째, AI의 비결정적인 특성으로 명확한 시험기준 수립이 제한되어 평가기준이 부재할 수 있다. 둘째, 실제 작전환경과 상황을 반영하지 못한 비현실적인 시나리오로 인해 시험의 신뢰성이 저하될 수 있다. 셋째, 학습데이터의 편향성과 대표성 부족으로 실제

환경에서 오판 가능성이 존재한다. 넷째, 같은 조건에서도 결과가 달라져 반복검증이 제한되며 시험평가 결과 예측 불확실성이 증대된다. 다섯째, 정량 평가의 한계로 정량적 지표만으로 AI의 상황판단(의도, 정황)을 평가하기가 어렵다. 여섯째, 인간과 AI 협업 구조 평가체계 구축이 제한되며, 일곱째, 실제 전장환경의 복잡성이 시험에 반영되지 않는다. 끝으로, AI의 도성 판단 및 설명 가능성이 부족하여 AI 판단의 의도, 책임을 설명하고 검증하기가 어렵다[5].

유인체계와 A-무인체계 시험평가 특성은 Table 1과 같다.

Table 1. Comparison of test evaluation characteristics [5]

Manned Weapon System	AI-unmanned Weapon System
<ul style="list-style-type: none"> High predictability of results Repeated test is possible under the same conditions Fixed conditions, clear features and performance criteria Applying Standardization Scenario Validation-focused on functionality Restricted environment, controlled variables Independent evaluation on a system-by-system basis 	<ul style="list-style-type: none"> Low predictability of outcomes Output different results under the same conditions 'Value-centered' judgment based on context and intent criteria Configuring context-based dynamic scenarios Reliability and Effectiveness Centered Configuring complex real-world battlefield environments Human-AI Collaboration Structure Evaluation Required

3. AI 기반 무기체계 시험평가 발전방안

‘AI의 특성은 기존 시험평가 방식의 근본적인 변화를 요구한다’는 미 국방부 CDAO AI T&E 담당자의 발언처럼 미래 다영역전장에서 안전하고 신뢰할 수 있는 AI 기반무기체계 운용을 위해서는 한반도 전장에 적합한 시험평가방안이 정립되어야 한다. 이를 위해 AI 무기 개발 선진국들의 사례를 살펴보고 우리의 발전방안을 고찰해 보겠다.

3.1 국외 동향

먼저 미 국방부(DoD)는 AI T&E framework¹⁾와

1) ① Operational T&E(OT&E): 작전환경 내 임무수행 적합성, 시나리오 기반 실효성 평가, ② Human systems integratio(HSI) T&E: 사람-AI 협업성, 사용성, 인지부담, 사용자 수용성 등 인간중심 평가, ③ System integration(SI) T&E: 센서, 통신, SW간 통합 연동성과 AI 기능의 상호 운용 검증, ④ AI model T&E: 데이터 품질,

RAI Toolkit을 통해 AI 기반 무기체계 시험평가 제한 사항에 대응하고 있다[6]. 첫째, RAI Toolkit의 SHIELD 원칙²⁾ 기반지침을 통해 평가기준을 제시함으로써 체계화가 가능토록 하였다[7]. 둘째, 데이터 편향 및 품질문제는 model(data) T&E에서 편향 탐지 및 다양성 확보 기준을 포함시켰다. 셋째, model T&E에 비결정성 대응을 위해 반복 가능한 시나리오 기반의 시험체계를 구축함으로써 반복시험 불가 및 결과 예측 불확실성을 해소하였다. 넷째, 정량 평가의 한계는 SHIELD ‘인간중심 설계’, ‘추적 가능성’ 원칙을 기반으로 수치 외 판단 근거와 의도성·맥락성 등을 정성적으로 지표화함으로써 보완하였다. 다섯째, 의도성 판단 및 설명가능성 부족은 RAI 전 영역에서 설명가능성·투명성·책임성을 내재화를 시도하였다[5].

NATO는 S&T trends 2023-2043에서 TEVV³⁾ 기반 임무중심 평가를 시험평가 핵심전략으로 제시하였다[5]. 먼저, 구체적으로 기능(function) 검증 중심에서 임무 수행 성공여부 중심으로 시험평가 목표를 전환하는 임무중심(mission centric) T&E를 강조하였다. 둘째, 실제 환경(real world)과 가상(synthetic) 환경의 결합을 통해 전장의 유사성을 확보하는 혼합(hybrid)환경 T&E를 제안하였다. 셋째, 반복 불가한 변수를 대응하기 위해 자동 시나리오 생성 및 루프화된 평가를 수행하는 자동화된 평가절차를 정립하고, 넷째, AI의 의사결정 추론 과정과 판단 로그를 분석해 판단 흐름 추적구조(decision trace) 기반의 책임성과 설명가능성을 확보하였다[8].

영국은 FATE(future advantage through evaluation)에서 국방의 미래 시험평가 체계를 위한 기반 프로그램을 제시하였다. 즉, 분리된 검증 활동이 아닌 지속적인 피드백 체계 구축과 복잡하고 경쟁적인 환경에서 운용될 첨단 적응형 군사 시스템 개발 지원을 위해 CLEAR 원칙 기반의 T&E 계획을 발표하였으

편향, 적합성을 모델 성능, 설명력, 비결정성 등 성능 검증

2) SHIELD는 제품수명주기에 따른 책임성의 6가지 원칙 및 실행절차로 ① S(set foundation, 기초설정), ② H(hone operationalizations, 운영화 구체화), ③ I(improve & innovate, 개선 및 혁신), ④ E(evaluate progress, 진척평가), ⑤ L(log for traceability, 추적 가능성 기록), ⑥ D(detect via continuous monitoring, 연속 모니터링을 통한 감지)

3) TEVV(test, evaluation, verification, and validation)는 AI 및 소프트웨어 시스템의 신뢰성, 안전성, 윤리성을 보장하기 위한 핵심 시험평가 및 검증 절차

며, 세부내용은 다음과 같다[6]. ① collaborative(협력적): 데이터 축적을 통한 이해도 향상, ② lifelong(전주기적): 전 수명주기 동안의 의사결정 지원, ③ efficient(효율적): 가용자원의 최적 활용, ④ actionable(실행가능): 적시적 의사결정 지원, ⑤ realistic(현실적): 실제 운용환경 반영 등이다[9].

선진국들은 수치 중심의 기존 시험평가의 한계를 인정하고, 판단·상황 기반의 평가 필요성과 실험환경 구조화를 위해 단계별 시나리오 작성, 혼합시험환경 조성 등을 제시하였다. 또한 AI 기반 무기체계의 책임성을 내재화하기 위해 평가에 설명가능성과 책임성을 포함시켜야 됨을 강조하였다.

3.2 국내 동향

미래전 양상변화와 대·내외 안보환경 변화에 발맞추어 우리 군도 AI 기술이 적용된 무기체계를 지속적으로 도입하고 있다. 최근 육군의 다목적무인차량, 해군의 정찰용·전투용 무인수상정, 공군의 중요시설 경계시스템 등 도입에 따라 각 군은 시험평가 환경 조성 및 제도 개선을 추진 중이다.

먼저, 국방부와 합참차원에서는 AI 무기체계의 신뢰성, 요구성능 충족성, 효율성 향상을 위한 제반 요소를 반영토록 연구용역과 AI 워킹그룹회의 등을 통해 시험평가 관련 법령, 훈령, 규정에 반영을 추진하고 있다. 미국, NATO 등 AI 무기 선진국들 사례를 바탕으로 AI 무기체계에 대해 AI 모델 성능평가와 실 운용환경 시험평가를 단계적으로 실시할 수 있는 법적·제도적 근거를 마련에 집중하고 있다.

육군은 3군 중 가장 적극적으로 AI 무기체계 시험평가 발전을 추진하고 있다. 먼저, AI 기반 전력에 대해 단계별 시험평가 개념을 정립하여 수행업무를 구체화 하였다. 즉, 연구개발 단계에서 군 작전환경 데이터 적용 학습을 실시하고, AI 모델의 성능을 평가하여 인증한 후 실환경 시험평가를 현장데이터 성능평가, 체계성능 시험평가를 구분하여 수행하는 것이다. 특히, '22년 9월에서 '23년 12월까지 자율주행 무기체계 시험평가용 시뮬레이터 개발사업으로 통해 M&S 체계를 활용한 자율주행 무기체계 시험평가 가능성 및 효용성을 확인하였다. 이를 지속적으로 가상환경 모델링, 다양한 시나리오 적용을 통한 사후분석으로 가상시험환경을 고도화시키고 있다. 2025년 과학화경

계시스템 성능개량, AI 기반 유·무인복합 한국형공병 차량 등에 검토된 단계별 시험평가 절차를 적용하였으며, AI 성능평가용 HW(GPU H200*2)/SW 도입을 위한 예산반영과 AI 데이터의 품질평가·검증 및 보강 플랫폼 구축을 위해 '26년 국방실험사업 소요를 제기하였다. 또한 객체식별 AI 모델의 강건성 검증체계 구축을 추진하고 있으며, 국방 생성형 AI 성능평가를 위한 벤치마크 데이터셋 구축방안연구와 뉴런 커버리지 측정을 통한 화이트박스 도구를 확보하여 AI 모델 경량화(효율화) 및 XAI 구현을 추진 중이다[10].

해군은 국방부, 합참, 타군의 관련규정과 제도 개정과 연계하여 상위 훈령과 시험평가/시운전 업무편람 등에 AI 기반 무기체계 시험평가 관련 내용 반영을 추진 중이다. 그러나 ADD 국방 AI 기술연구원과 육군 시험평가단의 적극적인 시험평가 모델 개발 및 예산화 작업, 데이터셋 구축, 세부 기술에 대한 구체적인 연구활동과 비교하면 매우 미흡한 실정이다. 다수의 드론전력 도입과 무인수상정·잠수정 개발사업 등을 고려할 때 관련 산·학·연과의 공동연구 및 학술활동 실적(1건)도 매우 저조하며[11], 현장 근무자에 의한 연구도 시험평가의 방향성을 제시하는 담론 수준에 머물러 있다[12].

따라서 조만간 전력화 예정인 AI 객체인식 기술과 자동 항해 및 타격·교전 등 복합업무를 수행해야 하는 무인수상정·잠수정 개발 로드맵에 따른 구체적인 시험평가 방안 정립이 시급한 실정이다.

4. AI 기반 무기체계 신뢰성 강화를 위한 시험평가 발전 방안

4.1 AI 기반 무기체계 시험평가 프레임워크 정립

AI 기반 무기체계 특성을 고려한 시험평가 체계 구축이 선행되어야 할 것이다. 무기체계 수명주기 전반에 걸친 시험평가 체계를 설계하여, 각 단계에 맞는 테스트 프로토콜을 정의하고, 데이터 수집, 모델학습, 배포 후 성능평가에 이르기까지 일관된 기준과 방법론을 적용해야 한다.

구체적으로 육군 사례를 벤치마킹하여 AI 모델, 주체계(SW/HW)에 대한 단계적 시험평가를 수행하는 것이다. 1단계는 개발업체와 기술품질원 주관으로 AI 모델의 요구성능 충족여부를 시험하여 성능평가

와 인증을 실시하고, 2단계는 시험평가처 주관하에 AI 기반 무기체계(AI 모델+주체계) 통합시험을 실시하여 ‘전투용 적합’ 여부를 판정한다. 3단계는 시험평가 후 운용부대 주관으로 성능평가 및 환류를 통해 성능을 향상시키고 고도화하는 작업을 반복하는 체계를 정립하는 것이다.

시험평가 방법은 다음과 같다. 먼저 AI 모델은 데이터 기반 성능평가를 수행한다. 첫째, 군에서 사전 구축한 시험평가용 데이터를 이용하여 정상 상황, 이상 상황, 극한 상황 등을 포함한 시험 데이터셋을 구성한다. 둘째, 반복 시나리오보다는 다양한 상황 데이터에 대한 AI 반응을 수집·분석한다. 셋째, 운용치, 로그, 센서 정보 등 실제 수집된 대규모 데이터 기반 분석을 통해 AI 알고리즘의 오탐율, 미탐율 등을 정량적으로 분석한다. 다음 주체계는 M&S를 통해 다양한 전장환경을 가상으로 구성하고, 핵심기능은 실선시험으로 직접 검증한다. 첫째, M&S 단계는 기상, 지형, 적 상황 등 가상 전장환경 및 위협 시나리오를 디지털 트윈 또는 HIL⁴⁾방식으로 평가한다. 둘째, 실선시험 단계는 실제 플랫폼(무인수상정, 무인항공기 등)을 대상으로 실제 환경에서 작동성을 검증한다. 기타 표적식별, 자율 결정 등 판단 알고리즘 시험은 M&S 시험 후 실선시험을 실시한다. GPS 재밍 등 극한상황 반응시험은 시뮬레이션 기반 반복시험 후 실선시험을 수행하고, 항법, 조작성 등 기본기능 시험은 실물시험을 수행한다.

구체적인 시험평가 절차는 다음과 같다. 1단계는 AI 모델에 대한 성능평가 및 인증을 군 작전환경 및 객체 특성을 고려한 무기체계용 AI 모델에 적용할 군사데이터를 이용하여 기품원 등 전문기관이 수행하며, AI 모델 성능평가 시 학습용 데이터와 시험평가용 데이터를 분리하여 활용한다. 2단계는 1단계에서 인증된 AI 모델을 추가학습 및 성능평가를 통해 학습능력을 강화하고, 강화된 AI 모델이 탑재된 주 체계와의 통합된 성능을 체계분석처 주관으로 M&S 후 검증된 AI 기반 무기체계를 최종적으로 시험평가한다. 이때 AI 모델 학습 시 군에서 구축한 데이터셋 유출 방지를 위한 보안대책을 강구해야 한다. 3단계는 시험평가 및 운용유지 간 구축된 데이터셋을 지속적으로 축적하여 시험평가 완료된 AI 기반 무기체계 고도화 및 유

사 무기체계 환류를 위해 AI 기반 무기체계 운용 간 지속적으로 데이터를 수집하고 추가 학습을 통해 고도화를 유지한다.

4.2 시험평가환경 조성 및 여건 보장

먼저 AI 기반 무기체계 시험평가에 필요한 물리적·가상적 인프라를 구축해야 한다. 가상시험 환경(virtual testbed) 구현을 위해 AI 시스템의 다양한 운용 시나리오를 검증하기 위한 시뮬레이션 환경을 설계하여 구축하고, 군사작전 시나리오의 성능과 안정성을 모의 환경에서 시험해야 한다. 물리적 시험 시설은 AI 무기체계, 드론, 센서, 네트워크 등을 평가할 수 있는 전문 시험장에서 실질적인 시험을 수행하기 위한 필수 시설이다. 또한 다양한 지리적 조건에서 시험을 병행할 수 있는 분산형 T&E 플랫폼을 구축하고 데이터 통합체계인 분산형 시험평가 네트워크 구축이 필요하다. 이를 위해 ‘국방 데이터 가시화체계 고도화’ 사업에 ‘해양 객체/표적 이미지 데이터 구축’을 추진하고, 해수부, 국과연 주관 무인체계 전용시험장 구축 사업 진행 간에 해군의 의견이 반영될 수 있도록 해야 한다. 끝으로 국방부 주관 자율주행 무기체계 시험평가 시뮬레이터 구축 고도화 사업을 참고하여 해군 시험평가용 M&S 체계 확보를 추진해야 한다.

다음 AI 기반 무기체계 시험평가 수행 주체인 인력에 대한 전문성을 향상시켜야 한다. 먼저 해군 자체적으로 AI 분야가 포함된 「무기체계 평가관 관리과정」을 개설하고, AI 모델과 AI 기반 무기체계 관련 교육을 교육사령부 예하의 전투병과학교, 정보통신학교 및 기술행정학교 등에서 실시해야 한다. ADD 국방기능인공기술연구원 등 전문기관을 활용하여 AI 무기체계 최신 연구·동향 학습을 위한 초빙강의를 적극적으로 추진해야 한다. 또한 민간, 학계 및 전문가들과의 협력 등 인적 네트워크를 강화해야 한다. 민간 기업과 학술 기관이 보유한 최신 AI 시험평가 기술을 활용하고, 데이터 과학자, 시스템 엔지니어 등 다양한 분야 전문가와의 협력을 통해 포괄적인 평가체계를 확립해야 할 것이다.

끝으로 AI 기반 무기체계의 윤리적·법적 위험완화와 책임성 확보를 고려해야 한다. 설명가능성 측면에서 운용자가 AI 판단 근거를 얼마나 명확히 이해할 수 있는지, 그 설명이 실제 의사결정에 기여할 수 있는지

4) HIL: Hardware in the loop : 실제 하드웨어 포함 시험

검증할 수 있어야 한다. 통제 가능성 측면에서는 비상 정지, 수동전환, 임무 중단 등의 조치가 실제 작전시나리오 속에서 안정적으로 작동하는지를 평가해야 한다[13]. AI의 윤리성 시험평가는 편향, 차별적 결과, 비윤리적 결정 등을 탐지할 수 있는 검증 프로세스를 설계해야 하며, 투명성, 해석가능성, 공정성 기준을 명시적으로 적용해야 한다. 또한 국제인권법, 무력충돌법 등 법적규제 준수여부를 평가하기 위해 각국의 법적 요구사항과 국제조약, 데이터 프라이버시 및 안전성 관련 법률 등을 준수하는지 확인할 수 있어야 할 것이다.

5. 결론

오늘도 계속되고 있는 러-우전쟁, 이스라엘-미국-이란전쟁에서 AI를 활용한 다양한 시스템들과 다수의 무인체계는 전쟁의 양상을 바꾸고 있다. 특히, Gospel, Lavender & Where's Daddy 등 AI 기반 무기체계를 활용하여 전장에서 신속한 의사결정과 행동으로 군사작전의 효율성과 치명성이 증대되고 있다. 한반도 전장에서도 미래에는 누가 더 우수한 AI 기반 무기체계를 보유하느냐가 전쟁의 승패를 좌우하는 결정적 수단이 될 것이다.

또한 AI 기반 무기체계의 능력은 최초의 시험평가 점수보다 배치 이후, 전장 환경에 얼마나 빠르게 학습하고 적응하느냐에 의해 좌우된다. 따라서 완벽한 전력화를 전제로 한 일괄 배치 방식에서 벗어나, 제한된 구역·임무·위협 수준 내에서 실제 부대와 함께 반복 운용하며 지능을 고도화하는 ‘운용 중 학습 루프(operational learning loop)’를 제도적으로 내재화해야 할 필요가 있다. 이를 위해 operational sandbox를 단순한 시험장이 아닌, 전장의 유사성, 위험 통제성(risk controllability), 데이터 수집 구조(data instrumentation), 즉각적 피드백 체계(immediate feedback architecture)가 결합된 반(半)실전형 통합 실증 환경으로 재정의할 필요가 있다[12].

이러한 문제의식을 바탕으로 본 연구는 미래 다영역전장 환경에서 AI 무기체계 운용 선진국인 미국의 AI T&E와 RAI Toolkit, 영국의 FATE 등의 사례를 참고하여 시험평가 기준과 안정적인 평가절차 등 AI 기반 무기체계 신뢰성 강화를 위한 시험평가 발전방안을 제시하였다. 먼저 AI 기반 무기체계 특성을 고려한

AI 모델, 주체계에 대한 단계적 시험평가 수행 등 프레임워크를 확립해야 한다. 또한 시험평가에 필요한 물리적·가상적 인프라를 구축, 시험평가 인력의 전문성 강화, AI 기반 무기체계의 윤리적·법적 위험완화 조치와 책임성 확보 등을 제안하였다.

실무적 차원에서는 개발 중인 정찰용 무인수상정의 시험평가 수행을 위해 TEMP⁵⁾, DT 및 OT 계획서 확정과 평가관 구성·교육이 선행되어야 할 것이다. 또한 국방부, 합참의 업무추진 로드맵에 따라 해군 전력발전업무규정과 시험평가/시운전업무 편람 등에 AI 기반 무기체계의 단계별 AI 모델 및 체계성능 시험평가 수행절차를 반영해야 한다.

참고문헌

- [1] U.S DoD, 2026 National Defense Strategy, 23 January 2026.
- [2] 대한민국 국방부, ‘안규백 국방부 장관, ‘50만 드론전사 양성’ 선언 [보도자료],’ 2025.09.04.
- [3] 합동참모본부, 무기체계 시험평가 업무 규정, 합참규정-543-01, 2025.05.13. 제정.
- [4] 김백중, ‘인공지능(AI) 기반 무기체계 시험 평가 개선방안 연구(운용시험평가를 중심으로),’ 광운대학교 대학원 박사학위논문, 2023.
- [5] 국방과학연구소 국방인공지능기술연구원, 국방인공지능 자율무기체계 시험평가 발전방안(비공개), 2025.05.22.
- [6] US DoD, Developmental Test and Evaluation of Artificial Intelligence-Enabled Systems Guidebook, Office of the Director, Developmental Test, Evaluation, and Assessments Office of the Under Secretary of Defense for Research and Engineering, February 2025.
- [7] 대한민국 국방부, 안전하고 신뢰할 수 있는 AI 활용 국방가이드라인(비공개) [연구보고서], 2024.12.
- [8] Bayliss Daniel, De Lucia Angelo, Martin-Blanco Alvaro, Reding Dale F., Regan Laura A., & Wells Bryan, Science & Technology Trends: 2023-2043 Vol. 1, 2023.
- [9] UK Ministry of Defence, Guidance Defence Test and Evaluation: Future Advantage Through Evaluation (FATE), 9 September 2024.
- [10] 대한민국 육군(육군시험평가단), ‘AI 기반 자율 무기체계 시험평가 발전방안,’ 2025 육군 시험평가 및 감항인증세미나 발표집, 2025.07.29., pp. 24-39.
- [11] Shin-Bae Park, Won-Jae Kim, & Kurnchul Lee, ‘A Study of the Development Test and Evaluation and Verification

⁵⁾ TEMP(Test and Evaluation Master Plan, 시험평가 기본계획서): 연구개발 무기체계의 시험평가계획을 종합적으로 명시한 문서로 개발시험평가계획서와 운용 시험평가계획서를 작성하기 위한 기준을 제공한다.

Procedure of a Multi-Mission USV, M-Searcher,' Journal of Ocean Engineering and Technology, VOL. 32, NO. 5, 2018, pp. 402-409.

[12] Jinwoo Kim & Jungkyu Yang, 'A Study on Mission Engineering-Based Development Methodology for AI Unmanned Systems: Focusing on Explainability and

Controllability,' Journal of KNST, VOL. 8, NO. 3, 2025, pp. 345-353.

[13] 김진우, '국방 AX 시대, AI 기반 무인체계 획득 방법의 진화(進化) - 나선형 개발과 진화적 획득을 통한 지능형 무인체계 구현,' 군사저널, 4월호, 2025, pp. 37-42.