



Received: 2022/07/07  
Revised: 2022/08/05  
Accepted: 2022/09/02  
Published: 2022/09/30

**\*Corresponding Author:**

Seunghoon Jung  
E-mail: hoonyjung9999@gmail.com

# 해상 모자이크전 환경에서 군사 목적 인공지능의 윤리적 접근 연구

## A Study on the Ethical Approach of Artificial Intelligence for Military Purposes in a Naval Mosaic Warfare

정승훈\*

해군소령/국방부 정보통신기반정책담당관실 한국형사이버보안제도 평가담당

Seunghoon Jung\*

LCDR, ROK Navy/K-RMF Evaluator, Information Infrastructure Division,  
Ministry of National Defense

### Abstract

AI와 자율시스템을 작전개념과 연결하여 최적의 방책을 제공하는 자동화된 전쟁의 개념을 모자이크전으로 명명하는 연구들이 진행되고 있다. 해상의 경우 함정 자체가 다양한 복합전 임무를 수행하는 점을 고려 시 모자이크전 환경에서 자율성을 가진 인공지능이 향후 큰 비중을 차지하게 될 것으로 예상된다. 이러한 상황에서 최근 여러 기관, 단체 등에서 발표한 윤리원칙을 살펴보고 해상 모자이크전 환경에서 군사 목적 AI의 윤리원칙 7개 항목(프라이버시, 책임성, 보안 및 신뢰성, 투명성 및 설명가능성, 공정성, 통제성, 적응성)을 제안하였다. 제안된 AI 윤리원칙은 향후 미래 해상작전 수행 시 준수되어야 관리될 필요성이 있다. 다소 제한적일 수 있겠지만 본 연구를 통해 해상 환경의 특수성을 이해하고 미래 모자이크전에서 향후 해상작전에 인공지능 운용 간 윤리적 문제점을 최소화하는데 효과적으로 활용될 수 있기를 기대한다.

Recently, studies that connects artificial intelligence (AI) and autonomous systems to operational concepts, approaches them as the concept of automated war that provides optimal measures, and calls them mosaic warfare. The point of performing various combined combat missions in battle ships should be considered. AI with autonomy is expected to occupy a large proportion in the future in the naval mosaic warfare. In this situation, This study considered the ethical principles recently announced by various institutions and organizations. In addition, seven items (privacy, responsible, security & reliable, transparency & explainability, equitable, governable, applicability) of military purpose AI were proposed in the naval mosaic warfare. The proposed AI ethical principles need to be followed and managed in future naval operations. Although it may be limited, it is expected that this study will understand the specificity of the maritime environment and be effectively used to minimize ethical problems between AI operations in future mosaic warfare.

### Keywords

인공지능(Artificial Intelligence),  
윤리적 접근(Ethical Approach),  
해상 모자이크전(Naval Mosaic Warfare),  
AI 취약성(AI Vulnerability)

### Acknowledgement

이 논문은 2022년도 해군사관학교 해양연구소 학술연구과제 연구비의 지원으로 수행된 연구임.

## 1. 서론

2022년 2월 24일 새벽에 러시아의 침공으로 시작된 우크라이나-러시아 전쟁은 많은 시사점을 보여주었다. 당시 러시아는 지상군 투입을 포함하여 여론전, 심리전, 사이버전 등 가용한 모든 전력을 총동원한 모습을 보였다. 개인적으로 관심을 가진 부분은 그 중 세 가지였다. 첫 번째는 러시아로부터 일방적으로 사이버공격을 받은 이후 국제 해커집단 어나니머스 및 우크라이나 IT부대를 통한 맞대응 차원의 사이버방어 및 공격이다[1]. 두 번째는 미군과 NATO군이 다수의 정찰기를 우크라이나 인근에 보내서 전장의 전방위적인 감시를 실시했다는 부분이다[2]. 세 번째는 ‘다른 나라에 살상 무기를 공급하지 않는다’는 독일의 오랜 원칙까지 반복하면서 제공한 대전차 무기와 지대공미사일 등이 적극 사용된 전쟁이란 점이다[3].

앞서의 세 가지 사안을 통해 최근의 전쟁 양상은 지/해/공/사이버/우주영역을 포함한 다영역 작전(multi-domain operations)[4]을 기본으로 이루어지며 다양한 출처를 통해 수집된 정보를 바탕으로 상대적으로 가격이 저렴한 대전차무기와 지대공미사일로 고가의 무기체계인 전차, 전투기를 격파하는 등 비대칭적인 요소가 복합적으로 작용하는 부분을 알 수 있었다. 미국에서는 예전부터 이러한 전쟁 양상을 ‘모자이크전’에 빗대어 표현하는 움직임들이 있었으며, 우리나라도 이와 관련된 연구들이 일부 진행된 바 있다. 하지만 이는 개념적인 연구 수준으로 진행되었으며, 군에 적극적으로 활용되기 시작한 인공지능(artificial intelligence, 이하 AI)과의 연계성까지 고려한 연구는 미비한 실정이다.

따라서 본 연구에서는 공군의 사례를 바탕으로 개념연구된 기존 모자이크전 연구를 해상 환경에서 바라보고, 이와 더불어 군사 목적의 AI가 접목되었을 경우 신뢰할 수 있는 AI의 윤리적 판단이 필요한 부분에 대해 연구하였다.

## 2. 관련연구

### 2.1 모자이크전

지난 수십 년 동안, 미국은 매우 강력하고 뛰어난 군사 플랫폼을 만들기 위해 첨단 기술을 활용해 왔으며, 이를 통해 경쟁국들의 위협에 성공적으로 대응해 왔다. 그러나 이러한 최신 기술이 민간분야에서 더욱 발전되고, 이를 상업시장에서 확보하기 용이해지면서 적대국들이 이용하기 쉬운 환경이 되었다. 국가가 복잡한 무기 시스템에 집중하는 것이 장점이 아니라 약점이 되는 상황도 발생하게 되었다. 2015년 미 국방고등연구계획국(DARPA, Defense Advanced Research Projects Agency)에서는 이러한 문제를 해결하기 위해 시스템 통합 기술 및 실험(SoSITE, system of systems integration technology and experimentation)이라고 하는 기술접근방식의 시작을 밝힌 바 있다[5]. 이 프로그램은 항공 플랫폼 중심으로 진행되었으며, 상호운용 가능한 다수의 유인플랫폼과 무인플랫폼에 항공전 능력을 분산시킬 수 있는 항공기, 무기, 센서 및 임무 시스템의 조합인 새로운 SoS 아키텍처를 통해 항공우위를 달성하는 것을 목표로 한다. Fig. 1에 보는 바와 같이 기존 플랫폼 중심의 종합적인 작전수행에서 기능 중심으로 분산하여 맞춤형 작전수행 형태로 발전하고자 하는 모습을 알 수 있다[5].

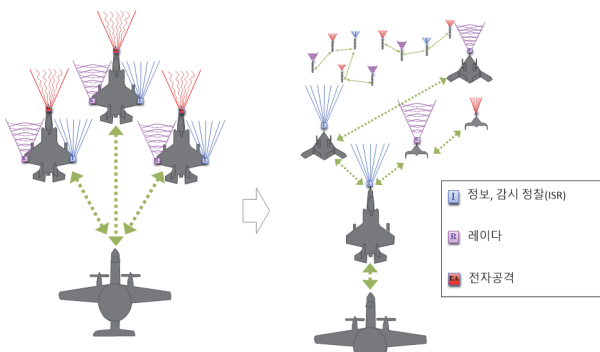


Fig. 1. 작전개념의 변화[5]

그로부터 몇 년 뒤 여기에서 더 나아가 새로운 개념이 언급되었다. 2018년 미 DARPA의 전략기술국장(director

of DARPA's strategic technology office) 티모시 그레이슨(Timothy Grayson)이 모자이크전에 대해 언급하면서 용어가 알려진 바 있다[6]. 그레이슨 국장은 “군은 경쟁국들보다 조금 더 빠르고, 조금 더 보호되고, 조금 더 위험한 무기를 개발하는 데 시간을 사용할 수 있다.”, “더 많은 비용을 사용하여 첨단 기술을 무기 시스템에 적용하여 앞서 나갈 수 있지만 이는 손해보는 방법이 될 수 있다.”, “개발하는 모든 단계는 이전보다 복잡하고, 도전적이며, 고비용에 시간도 많이 필요하다.”, “미군은 자신들이 보유하고 있는 것들을 혁신적인 방법으로 사용하여 적들을 압도하고, 적의 내면에서 지도자들이 의사결정을 수행하는 과정에 지장을 주어야 한다”고 언급하였다. 그리고, 이러한 내용을 DARPA 내부에서는 ‘모자이크전’이라고 부르고 있다고 하였다. 이 용어는 DARPA의 토마스 번스(Thomas J. Burns) 전(前) 전략기술실장(former director of the strategic technology office)과 댄 패트(Dan Patt) 전 부국장(former deputy director)이 만들었다고 한 바 있다. 이를 그레이슨 국장이 발전시켰다고 볼 수 있다.

패트 전 부국장은 “모자이크전의 장점은 킬체인을 더욱 탄력적으로 만든다.”, “모자이크전을 통해 수학적으로 모든 가능한 조합을 식별하고 수천 개의 연결고리를 통해 적에게 가능한 많은 공격 조합과 싸우도록 강요한다.”는 언급한 바 있다[7].

### 2.2 모자이크전 기대효과

2020년 패트 전 부국장은 미 전략예산평가센터(CSBA, Center for Strategic and Budgetary Assessments) 연구진과 함께 모자이크전에 대한 보고서를 발간했다[8]. 해당 보고서는 모자이크전이 실제 작전에 유용한지에 대해 설명하고 있으며, 위게임을 통해 그 결과를 증명하고자 하였다. 보고서가 2035년을 가정하여 설정한 3가지 위게임 상황은 다음과 같다.

첫 번째는 합동 특수임무부대가 콩고 민주 공화국 정부를 지원하고 적 해군 및 지상군에 대항하여 비전투원 철수를 수행하는 임무 상황이며, 두 번째는 합동 특수임무부대가 미국의 주변 동맹국들을 공격하는 강대국과 대치하는 상황에서 인질 구출, 석유 및 가스 시설 보호, 핵물질의 위치 확인 및 회수 등의 임무를 수행하는 상황이다. 마지막 세 번째는 합동 특수임무부대가 인도양 전역에서 적군에 대항하는 더 큰 전면전 양상으로 탄자니아와 케냐에 기반을 둔 적군의 탐지체제와 무기 복합체를

공격하는 임무를 가정한 상황으로 장거리 레이더, 중첩된 능동/수동의 공중/미사일, 첨단 전투기와 탄도/순항 미사일이 포함된 서인도양에 배치된 적군 함정과 잠수함에 대응을 가정한다.

각 시나리오별 두 개의 블루 팀(하나는 모자이크팀, 하나는 기존의 전통적인 작전팀)으로 나누고 컴퓨터 기반 시뮬레이션으로 수행하였으며, 실제 적의 역할을 하는 레드 팀은 동일한 상황을 부여하였다. 현실적으로 모자이크전에 필요한 기본 시스템과 기능이 존재하는 것은 아니므로 위게임은 향후 모자이크전을 원활히 구현하기 위한 네 가지 주요 기술적인 도전들(보급, 통신, AI 기반 기계 제어, 자율 시스템)을 극복했다는 가정하에 시행하였다. 물론 전통적인 작전팀도 지속 요구되는 사항(비행 연료, 전투단의 7일 연속작전 수행 등)도 충족하고 있다고 가정하였다(지난 우크라이나-러시아 전쟁의 경우 러시아 전차가 연료가 부족하여 방치한 부분도 있으나 이런 사항들은 가정 사항에서 제외). 주요 분석을 한 대상은 의사결정방식의 차이로 인한 결괏값이 다른가에 관한 부분이었는데 그 결과는 다음과 같았다.

첫 번째와 두 번째 위게임의 경우 모자이크팀과 전통팀 모두 작전임무를 달성하였으며, 세 번째 위게임에서는 모자이크팀만 작전임무를 달성할 수 있었다. 마지막 위게임 동안 두 팀 모두 인도양 전역에서 적군과 더 큰 분쟁의 일환으로 탄자니아와 케냐에 기반을 둔 적의 센서와 무기 복합체를 공격하는 임무를 수행하였다. 센서와 무기 복합체에는 장거리 레이더, 다량의 능동·수동 공중·미사일 방어체계, 첨단 전투기와 탄도·순항미사일 뿐만 아니라 서인도양에 배치된 적군 함정과 잠수함까지 포함된다. 이러한 적군의 위협에 대처하기 위해 적군의 결정을 압도하고 아군의 행동의 자유를 회복하고 적군의 지도자들에게 분쟁으로부터 벗어나도록 압력을 가하기 위해 서인도양에서 적군에 대한 일련의 빠르고 평행한 공격을 추구하기로 결정하였다. 그 결과 오직 모자이크팀만이 목표를 성공적으로 달성할 수 있었다.

첫 번째와 두 번째의 경우 임무는 달성되었으나 각 팀 별로 독립적인 작전의 다양성에 차이가 있었으며, 모자이크팀이 많은 수의 독립작전을 진행하였다. 세 번째 위게임에서 모자이크팀과 전통팀의 독립작전 수는 크게 차이나지 않았다. 전통팀의 경우에도 각각의 플랫폼별 고유한 임무가 부여되어 일정한 독립성을 보장하였기 때문으로 생각된다. 다만 6번째 턴에서 피격 등으로 전력의 손실이 발생하는 것을 고려한 작전의 재편성 등의 고려요소

가 증가하여 전통팀의 경우 정해진 시간 내에 의사결정을 시행하지 못하였다. 모자이크팀은 전력요소가 증가하여도 의사결정 판단시간에 큰 차이가 발생하지는 않았고 세 번째 위게임의 임무달성이 가능하였다.

모자이크팀은 적의 장거리 센서와 미사일을 파괴하기 위해 동시다발적인 작전팀을 운영하였고, 이로 인해 아군의 피해도 많이 발생하였다. 하지만 대부분의 손실은 소규모 모이거나 무인 부대에서 일어났다. 의도적으로 비대칭적 공격을 집중하고 더 적은 비용 대비 효과로 적군의 미사일 기지 파괴에 노력하여 성공확률은 낮았지만 작전적인 효과에서는 더 많은 긍정적인 결과가 도출되었다. 세 번째 위게임에서 모자이크팀이 손실한 플랫폼 전체 비용은 전통팀 손실의 3분의 1에도 미치지 못했다. 모자이크팀의 경우 지휘관이 더 많은 동시 행동을 취할 수 있게 함으로써 적군의 의사결정을 더욱 복잡하게 만들 수 있다고 보여진다.

상기 결과에서 유의할 점도 존재한다. 최초 가정한 주요 기술적인 도전들을 극복했다고 하더라도 실제 위 게임을 진행하면서 모자이크팀에서 제안하는 여러 제안사항을 비판 없이 수용하고, 모자이크팀의 경우 전통적인 팀에 비해 더 많은 자원(인력 등)이 포함되어 다양한 경우의 수를 적용할 수 있었다는 부분이 그에 해당한다.

또한, AI 기반 기계 제어의 활용부분을 의사결정 보조 장치로 접근한 부분이 주목할만하다. 소규모 유인 부대나 완전한 무인 시스템에 AI를 통합한 의사결정 보조 장치는 지휘관의 비교적 간단한 지시를 달성하기 위해 가용한 기계 장치들을 유기적으로 활용하는 방식을 취할 수 있다. 운전을 예로 들면 내비게이션이 맵을 기반으로 수치적인 알고리즘을 바탕으로 최단거리를 제시하는 경우와, 해당 지역에서 오랫동안 운전을 해온 택시운전기사가 내비게이션이 파악할 수 없는 주요 막히는 도로를 피해서 전체 거리는 증가하지만 실제 도착이 더 빠른 길을 즉흥적으로 제시하는 경우의 차이라고 볼 수 있다. 이를 군에 적용해 보면 기존의 작전교리를 바탕으로 위성사진 등을 기준으로 유용한 작전방향을 AI가 제시할 수 있는 반면, 다양한 경험을 가진 지휘관이 본인의 경험을 바탕으로 작전지시를 수행하여 비효율적으로 보이지만 결론적으로 더 나은 결과를 도출할 수도 있다. 이처럼 AI의 결과가 항상 최선이라고 할 수는 없으므로 AI의 활용을 보조 장치로 접근하였다고 볼 수 있다.

상기 위게임 결과에서 저자는 상대편, 즉 적군 지휘관과 작전에 대한 의사결정과 행동에 미치는 영향, 그리고 상대방도 모자이크전을 수행하는 것에 대한 영향은 추가

로 고민이 필요하다고 언급하고 있다. 유의미한 긍정적인 효과가 있지만 상대편과의 동등한 수준에서 대응은 다양한 경우의 수가 있으므로 고민이 필요하다는 의미로 생각된다.

### 2.3 해상 모자이크전

우리 해군은 미래혁신연구단과 같은 조직을 창설하여 미래를 대비한 해군력 향상을 위해 다양한 사업과 연구 등을 추진하고 있다. 모자이크전의 경우에도 앞서 미군의 개념을 연구하고 한국군 발전방향을 제언하는 연구가 진행된 바 있다[9]. 해당 연구는 모자이크전을 대비한 미국의 미래 전력건설 방향을 소개하고 있는데, 향후 대형함 위주의 전력에서 대형함을 줄이는 대신 중·소형함의 비중 확대와 USV의 신규 편성이 언급되었다.

미래의 해상 모자이크전을 개념적으로 생각해보면 Fig. 2와 같이 임무별, 또는 해역별 작전수행을 위해 함정, 항공기를 포함한 다양한 전력이 해상에서 전투를 수행하는 모습을 그려볼 수 있을 것이다[10]. 앞으로의 함정 역할은 기존의 센서와 무장을 단일 플랫폼에서 모두 수행하던 것에서 Fig. 3와 같이 센서나 무장을 무인함정(USV, un-



Fig. 2. 해상 모자이크전 상상도[10]



Fig. 3. 해상 플랫폼을 활용한 무인항공기 운용[11]

manned surface vehicle), 무인항공기(UAV, unmanned aerial vehicle), 무인잠수정(UUV, unmanned underwater vehicle) 등에 분산하고 역할을 나눠서 수행하는 방향으로 발전할 가능성이 높다[11].

이와 더불어 지상에서 출격하는 항공기나 UAV, 센서의 정보를 공유하고, 해당 전력이 지휘함의 작전구역으로 접근하는 경우 통제권을 받아 작전을 수행하는 모습 등이 복합적으로 적용되는 모습이 해상에서 모자이크전 수행이라 할 수 있을 것이다. 여기서의 핵심은 모자이크전에 참여하는 모든 센서나 무장 등과 같은 전력을 운영하는데 있어 복잡한 작전수행이 적기에 원활히 진행될 수 있도록 하는 AI 참모의 역할이 중요하다는 것이다.

### 2.4 군사목적의 AI

미래전 양상은 기존의 무기체계들을 복합적으로 운용하여 F-32와 같이 다수의 기능을 통합한 소수의 정교한 무기체계보다는 한두 개의 기능으로 분리된 전력 요소들을 신속하게 구성 또는 재구성할 수 있는 군사력을 바탕으로 아군에겐 적응성과 유연성을 확보하고 적군에겐 복잡성과 불확실성을 부여하게 될 것이다.

예를 들어 함정에서 대형 3D 프린터로 소모성 소형 무인기를 제작(소형 엔진 등은 별도 사전 탑재하고, 외형을 제작하여 조립하는 형태 등)하고, 목적에 따라 정찰용, 공격용, 지원용 등으로 유기적으로 활용할 수 있을 것이다. 1대의 무인기에 관련 기능을 전부 포함할 경우 배터리 용량, 주요 기능 입력을 위한 HW 성능이 높게 필요하지만, 일부 단순한 기능만 부여하는 소모성 무기로 운용할 경우 HW 사양이 높을 필요가 없으며 유기적으로 활용할 수 있을 것이다.

또한, 최근 AI와 자율시스템을 작전개념과 연결하여 적보다 빠르고 효과적으로 의사결정을 도출하도록 하는 연구가 적극적으로 진행되고 있으며, 지휘관과 기계의 상호작용을 통해 가용전력을 분석하고 최적의 방책을 제공하는 자동화된 절차개발이 다양한 분야에서 추진되고 있는 상황이다.

이와 관련된 기술들은 국내에서 개발중인 ‘국방 AI참모’(한화시스템은 AI를 활용한 국방 지휘통제 지능화 서비스 개발 선도형 핵심기술 과제인 ‘지능형 전장 인식 서비스 및 플랫폼·서비스 통합 기술’ 과제를 추진중에 있음), AI 자율운항 무인수상정 등의 가시적인 성과들을 통해 구체화되고 있는 추세이다[12].

이러한 분야 이외에도 군사 목적의 AI는 광범위하게 적용될 수 있다. 한국국방연구원에서 미국 사례를 통해 군사 목적의 AI 적용 대표적인 분야를 종합하였으며 그 내용을 정리하면 Table 1과 같다[13].

**Table 1.** 군사적 목적의 AI 활용: 미군 사례[13]

구분	분야
감시정찰 (ISR)	무인비행체(UAV) 촬영 영상 분석, 표적 자동 식별, 다중언어 음성인식 (speech recognition) 및 번역, 2D 영상을 융합하여 3D 모델 생성 등
군수 (logistics)	공군 항공기 예방정비 (F-35 자동군수정보체계), 육군 장갑차 맞춤형 정비 (IBM Watson과 계약 등)
사이버작전 (cyberspace operations)	DARPA의 AI기반 사이버도구 (Cyber Grand Challenge, 2016), 기존 악성코드 변조 신속 대응, 공격/방어를 동시 시행 등
정보작전 (information operations)	군 요원, 정보요원 등 관찰이 필요한 사람 대상으로 포괄적인 행동 프로파일 (구매 이력, 신용보고서, 전문경력, 가입현황 등)을 만들고 관찰대상 이상행동 포착, 대응 등
영상합성/조작 (deep fakes)	AI의 영상합성/조작을 활용한 위조사진, 모조 오디오/비디오 등으로 가짜뉴스를 만들어 대중의 신뢰를 무너뜨리거나 외교관 협박, 이에 대응하는 AI 모조품 식별 도구 등
지휘통제 (command & control)	AI 기반 공통작전상황도 (common operating picture, COP)가 모든 정보를 결합, 정보 불일치 자동 보정, 작·아 전체 상황을 하나의 화면으로 제시, 실시간 전장상황 분석하여 지휘관에게 행동방책 제시 등 ※ 모자이크전에서 AI 적용의 주된 분야
반자율/자율운송체 (semiautonomous & autonomous vehicles)	유인전투기(F-35 or F-22)와 무인전투기(F-16)를 한 조로 편성하여 작전수행 (Loyal Wingman Project), 전장에서 전투원을 뒤따르는 무인차량, 잠수함 추적용 무인함정, 군집활동 무인체 등
자율살상무기 (lethal autonomous weapon system)	사람의 직접 통제 없이 다양한 센서와 컴퓨터 알고리즘을 이용하여 독자적으로 표적 식별, 타격하는 시스템, 통신이 불가능하거나 열악한 환경에서 활용 등

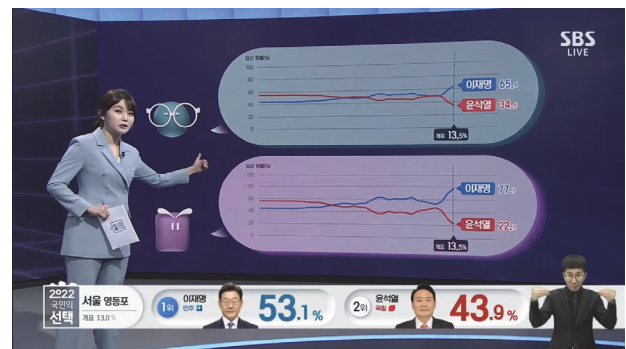
앞서 표에서 모자이크전에서 사용되는 군사 목적의 AI가 활용되는 가장 주된 분야는 지휘통제분야라고 볼 수 있다. 그리고 지휘통제분야는 자율살상무기를 통제하는 기본 상위개념에서 운용될 수 있는데, 자율살상무기의 경우 AI의 판단에 대한 의존도가 매우 높다고 볼 수 있다. 따라서 이에 대한 정확한 판단 및 검증이 필요할 것이다. 이

점을 고려하여 미 국방부에서는 훈령에 관련 내용을 포함하고 있다. 미 국방부의 ‘타격체계의 자율성(autonomy in weapon system)’ 관련 훈령[14]에서는 자율성을 가진 무기체계를 개발하는 경우 체계 사용 전반에 걸쳐서 지휘관이나 운영자가 적절한 수준의 판단을 행사하도록 설계되어야 하고, 개발 및 전력화 단계에서는 국방부와 합참의 검토 및 승인을 받도록 언급되어 있다.

## 2.5 AI 활용의 유의점

2021년 8월, 아프가니스탄 전쟁 종료가 선언되면서 미군이 철군하기 하루 전날 테러단체 이슬람국가 호라산(IS-K)을 겨냥해 단행한 드론 공격에 2살 어린이를 포함한 일가족 10명이 사망한 사건이 발생하였다(그 중 18세 미만이 8명)[15]. AI참모가 이러한 실수를 하지 않도록 사전에 편향되지 않은 정보의 제공(정보 종합/취합 단계의 검증), 건전한 판단(판단이 적절한지에 대한 검증)을 통한 지휘부 참모 건의가 필요한 부분이다.

2022년 대한민국 제20대 대통령 투표 당시 SBS에서 투표결과를 AI 예측분석으로 제시한 사례도 있다[16]. 당시 ‘AI 유확당’이라는 이름으로 사전 정보(과거 유권자 투표 성향, 읍면동 단위 초정밀 데이터, SBS 선거방송 노하우) + 실시간 정보(선관위 개표 데이터) + 신중도를 종합적으로 판단하여 특정 후보가 1등할 확률을 계산한 다음 이를 Fig. 4와 같이 표현하였다.



**Fig. 4.** 대통령 선거 당시 AI 분석 사례[16]


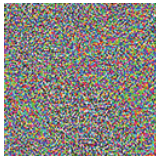


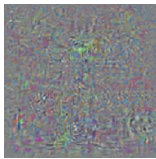


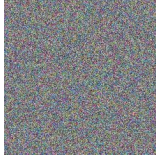




개표가 진행되면서 여당의 후보가 대통령으로 당선될 확률이 매우 높다고 판단하고 있었으며, 시간이 지날수록 그 격차는 더욱 커지고 있었다. 하지만 실제 개표가 진행되면서 AI의 판단을 신뢰하기가 어려워져서 개표방송 중간에 이러한 분석내용은 사라지게 되었다. 물론 최종 투표결과도 AI의 예측내용과는 다르게 진행되었다.

이러한 일부의 사례에서도 알 수 있듯이 AI가 제시하는 각종 정보를 막연히 신뢰하기에는 어려움이 있으며 그 정확성과 근거에 대한 재판단이 필요하다고 볼 수 있다.

2.6 AI 취약요소

AI의 판단에 유의할 부분으로 보안 및 신뢰성도 고민해볼 요소이다. AI 취약요소로 언급되고 있는 적대적 공격(adversarial attack)이 그 대표적인 사례라 할 수 있다. 평소에 정상적으로 인식하고 사용되는 정보가 사람들은 인식할 수 없지만 기계만 인식하는 적대적 변화사항 또는 노이즈로 인해서 잘못된 판단을 내릴 수 있다. 공개된 정보를 종합하면 Table 2와 같다.

Table 2. AI 적대적 공격 사례[17]-[20]

No.	최초 인식	적대적 변화사항 또는 잡음	합성 후 인식
1	 판다		 긴팔원숭이
2	 멈춤		 양보
3	 고릴라		 독일 사냥개
4	 안녕하세요		 문 열어라

표에서 보는 바와 같이 사람이라면 하지 않을 판단이지만 AI의 경우 데이터 기반으로 결과를 도출하는 만큼 그 데이터가 오염되고 변경된다든지, 또는 개별의 데이터는 관련이 없지만 합성되는 과정에서 최종 판단결과가 변한다든지의 취약점이 존재하므로 AI에만 판단을 전적으로

맡기는 부분은 신중히 생각해볼 필요가 있다. 물론 이에 대한 대응책들도 많이 연구되고 있으나, 특히 국방영역에서의 적용을 고민할 때는 이러한 취약점이 매우 중요한 사안으로 취급되어야 할 것이다.

2.7 AI의 윤리적 접근

과거 프로젝트 메이븐(Project Maven, 2019년 미 국방부가 구글과 AI 관련 계약을 맺으면서 당시 구글 직원들이 반발하여 약 12명의 직원이 회사를 떠났고, 결국 구글은 2019년 국방부와 맺었던 계약을 갱신하지 않았음)을 둘러싼 논란으로 신뢰에 타격을 입은 미 국방부가 신뢰 회복을 위해 AI 윤리 가이드라인을 마련하였다. 미 국방부와 기업 간의 계약 체결 책임을 맡은 국방혁신단(DIU, Defense Innovation Unit)은 투명성 제고 차원에서 이른바 ‘책임 있는 AI’ 가이드라인을 공개했다. 해당 가이드라인에는 “국방부는 AI가 가진 능력에 대해 의도하지 않게 갖게 된 편견을 최소화하기 위해 신중한 조치를 취할 것” 등의 내용이 포함되어 있다[21].

앞서 관련 연구에서 확인된 내용들에서 알 수 있듯이, 미 국방부와 마찬가지로 우리나라에서도 전면적으로 대두되고 활용범위를 넓히고 있는 AI를 국방분야, 그리고 해상에서 복합적으로 운용되는 모자이크전 환경에 적용함에 있어서 윤리적 문제를 고민해봐야 할 것이다.

3. 군사목적 AI의 윤리적 접근 방향

3.1 국제기구 윤리 규범

경제개발협력기구(이하 OECD) 및 국제연합교육과학문화기구(이하 UNESCO)에서 AI에 대한 윤리 규범을 각각 발표한 바 있다.

OECD의 경우 AI 전문가 그룹(한국 과학기술정보통신부에서 주도)에서 제시한 AI 이사회 권고안을 회원국 모두 만장일치로 2019년 5월 공식채택하였다. 해당 권고안에는 AI 시스템은 정보를 투명하게 제시하고 결과에 책임을 져야하며 AI 결과에 대한 인간이 통제권을 행사할 수 있어야 한다는 내용 등이 포함되어 있다. OECD 권고안의 경우 법적 구속력은 없지만 국제 표준을 설정하는데 영향을 발휘(OECD가 2013년 결의한 개인정보보호지침은 미국과 유럽을 포함한 전 세계 개인정보보호법의 기초로 활용)하고 있다[22].

UNESCO의 경우 2018년 AI 윤리 관련 UNESCO의 역할 마련을 위한 연구 준비를 시작으로 2020년 AI 윤리 권고안 초안을 만들어 회원국으로부터 의견을 수렴하였다. 이후 2011년 11월 제41회 총회에서 193개 회원국 만장일치로 AI 윤리 권고안을 채택한 바 있다. 해당 권고안은 4대 가치와 10대 원칙으로 정리되어 있으며 OECD 권고안과 마찬가지로 AI의 책임, 의무, 인간의 감독 및 결정 등에 대한 내용이 포함되어 있다[23].

이들 국제기구별 권고안은 군사 목적의 AI 윤리적 접근 방향을 명확하게 기술하고 있지는 않지만 AI 행위자 및 회원국은 국내·외 법에 따른 윤리적·법적 책임을 준수해야 하는 부분을 통해 간접적으로 그 개념에 대해 언급하고 있다고 볼 수 있다.

### 3.2 미군의 AI 윤리원칙

미 국방부는 2020년 2월 AI 사용을 위한 일련의 윤리적 원칙을 공식 채택하고 이를 공개한 바 있다. 이 권고안은 15개월 동안 정부부처, 산업계, 학계, 민간 전문가 등 주요 AI 전문가들과 협의한 끝에 도출된 내용으로 군사 목적의 AI 시스템 합법적 활용을 목표하고 있는 미 국방부의 전략 목표를 따르고 있다. 미군의 AI 윤리원칙은 미국 헌법 10호(Title 10 of the U.S. Code, Law of War)에서 생각하고 있는 기존 국제조약과 가치관에 기반을 두고 있다. 군에서 AI를 사용하는 경우 기존에 없는 새로운 윤리적 모호성과 위험을 가져올 수 있다. 미군은 이러한 문제를 해결하고 AI의 책임 있는 사용을 보장하는데 AI 윤리원칙이 도움이 될 것으로 보고 있다[24].

미군은 AI를 사용하면서 법적, 윤리적, 정책적으로 전투 및 비전투 기능에 모두 적용하기 위한 5가지 원칙을 제시하였다. AI 개발, 배치 및 사용에 대한 책임을 유지하면서 적절한 수준의 판단과 관리를 수행하기 위한 책임성(responsible), AI의 의도하지 않은 편견을 최소화하기 위한 공평성(equitable), AI 데이터 출처, 설계 절차, 문서화, 기술, 개발 프로세스 및 운영 방법 등에 대한 이해가 가능한 추적성(traceable), AI 기능을 명확하게 정의하여 안전, 보안, 효과가 전체 수명주기에 정의된 용도 내에서 시험 및 보장이 가능한 신뢰성(reliable), AI 기능이 의도한 기능을 구현하도록 하고, 의도되지 않은 결과를 탐지/회피할 수 있는 능력과 의도하지 않은 행위를 보이는 시스템을 분리 또는 비활성 할 수 있는 통제성(governable)이 미군의 5대 AI 윤리원칙에 해당한다[25].

미군은 해당 윤리원칙을 참고하여 AI 개발을 시행하고 있으며, 이 점을 참고하여 우리 실정에 맞는 해상에서의 윤리원칙을 고민해 볼 필요성이 있다.

### 3.3 한국의 AI 윤리기준

2020년 12월 대통령 직속의 4차산업혁명위원회는 과학기술정보통신부와 정보통신정책연구원에서 제안한 AI 윤리기준을 심의 의결한 바 있다. 해당 기준은 윤리학, 법학 등을 포함한 산학연과 주요 전문가 논의, 자문, 공청회, 시민 의견수렴 등을 거쳐 정리되었다. ‘사람 중심의 AI’라는 핵심 키워드를 바탕으로 마련된 기준은 3대 기본원칙과 10대 핵심요건으로 구성되어 있다. 3대 기본원칙은 인간의 존엄성 원칙, 사회의 공공선 원칙, 합목적성 원칙이며, 10대 핵심요건은 인권 보장, 프라이버시 보호, 다양성 존중, 침해금지, 공공성, 연대성, 데이터 관리, 책임성, 안정성, 투명성이다[26].

해당 기준은 법적 구속력이 있는 지침은 아니지만 도덕적으로 자율성을 가지는 규범의 성격을 지니며, 이번 제정으로 영구히 결정되는 것이 아닌 지속 발전하는 형태로 접근하고 있다. 우리 국방 영역에서도 이러한 정부의 방향을 이해하고 AI 윤리적 방향을 검토해야 할 것이다.

### 3.4 군사목적 AI의 윤리적 접근방향

앞서 살펴본 국제기구, 미군, 우리나라의 AI 윤리원칙 또는 기준을 그룹화하여 정리해보면 유사한 부분을 확인할 수 있다. 그 예로 미 하버드대학교 버크만 클라인 센터(Berkman Klein Center)에서는 2016~2019년 발표된 AI 윤리에 대한 문서를 분석하여 프라이버시, 책임성, 안전 및 보안, 투명성 및 설명가능성, 공정성 및 차별금지, 기술에 대한 인간의 통제, 전문적 책임, 인간 가치 증진 등 총 8개의 테마를 도출하였다[27].

해당 8개 테마를 중심으로 앞서 관련 연구에서 언급한 각 원칙들을 종합[28]하면 Table 3와 같다. Table 3에 나타난 바와 같이 각 테마에 모두 해당하는 AI 윤리원칙은 UNESCO가 유일하다. UNESCO는 8대 테마 이외에도 비례성 및 무해성, 지속가능성 등 다양한 내용의 AI 윤리원칙을 다루고 있다. OECD에서는 프라이버시, 기술에 대한 인간의 통제, 전문적 책임 등을 고려하지 않고 있다. 미군의 경우에도 프라이버시, 전문적 책임, 인간가치 등을 미 포함하고 있어 다른 윤리원칙에 비해서 다소 적게 느껴

**Table 3.** 하버드대학교 버크만 클라인 센터의 8개 AI 테마를 기준으로 정리한 윤리원칙

AI 8대 테마	OECD	UNESCO	미군	한국
프라이버시 (privacy)	-	프라이버시권 및 데이터보호	-	프라이버시 보호, 데이터 관리
책임성 (accountability)	책임성	책임 및 의무	책임성	침해금지, 책임성
안전 및 보안 (safety and security)	견고성, 보안 및 안전	안전 및 보안	신뢰성	안정성
투명성 및 설명가능성 (transparency and explainability)	투명성 및 설명가능성	투명성 및 설명가능성	추적성	투명성
공정성 및 차별금지 (fairness and non-discrimination)	공정성	다양성 및 포용성 보장 공정성 및 비차별성	공평성	다양성 존중
기술에 대한 인간의 통제 (human control of technology)	-	인간의 감독 및 결정	통제성	침해금지
전문적 책임 (professional responsibility)	-	다자적이고 조정가능한 거버넌스 및 협력	-	-
인간가치증진 (promotion of human values)	인간중심가치, 포용적 성장, 지속가능한 발전	인권, 기본적 자유, 인간 존엄성의 존중·보호·증진	-	인간의 존엄성, 사회의 공공선, 인권보장, 공공성
기타	웰빙(well-being)	비례성 및 무해성, 지속가능성, 의식 및 리터러시, 환경 및 생태계 변형 등	-	기술의 함묵적성 원칙 연대성

질 수 있으나 책무성 등 5개 분야는 잘 고려하고 있어 실제적으로 주요한 핵심적인 내용은 모두 언급하고 있다고 생각된다.

해당 표를 중심으로 군사목적 AI 윤리적 접근방향을 고민해 볼 수 있는데, 미군에서 언급하고 있는 5가지 내용은 필수적인 사항으로 포함하고, 해당 원칙에서 고려하지 않은 다른 테마에서 추가적인 요소를 도출하는 것이 유용한 방향으로 생각된다. 특히, 이를 해상이라는 특수한 환경과 향후 미래 전장이라 생각되는 모자이크전을 고려해보면 AI 윤리원칙을 추가로 적용할 필요성이 있다.

## 4. 해상 모자이크전 환경에서 군사 목적 AI의 윤리적 접근

### 4.1 주요 고려요소

2019년 1월, 미 해군 연구소(US Naval Institute)에서는 상선을 미사일 탑재함으로 개조하자는 의견을 개진하였다. 이에 따르면 미 해군의 함정 건조 비용을 절감을 위해 기존에 있는 컨테이너 운반선이나 이중선체 유조선을 미사일로 무장시키라고 제안하고 있다. Fig. 5는 민간 선박을 활용한 대표적인 사례인 미 해군 수송사령부 소속 중형 고속수송함이다[29].



**Fig. 5.** 미 중형 고속수송함(T-KAR 313)[30]

영국도 2019년 민간여객선과 화물선 각 1척을 사들여 이를 군함으로 개조할 것이라고 발표하였다. 영국은 이미 지난 2013년에도 유빙 감시선을 임대해서 군함으로 개조한 바 있다[31].

중국의 경우 2021년 9월 자국 영해를 진입하는 ‘위험 선박’에 대해 사전 신고를 의무화하였다. 중국 ‘해상교통 안전법’에 따르면 위험 물질을 실은 선박의 경우 사전에 중국 정부에 신고하게 되어 있는데 이는 군사적 목적에 활용되는 민간 선박에 주로 적용될 것이라고 예상되고 있다[32]. 이를 통해 민간 선박이라 할지라도 군사 목적으로 활용하는 사례가 많고, 이 상황을 통제하고자 하는 중국의 방향성을 알 수 있다.

그 외에도 최근 한국전자통신연구원에서는 ‘딥러닝 기반 해상감시영상 내 선박 검출 식별 기술’을 개발 완료하



고 민간에 기술 이전을 추진하고 있으며, 이러한 기술들은 국내에서 개발 중인 국방 AI참모 등과 연계하여 함정에서 활용할 수 있을 것이다. 2018년 평양공동선언 당시 한강 하구에 공동이용수역을 설정하고 남북의 민간선박이 자유롭게 항행할 수 있도록 하는 군사합의가 추진된 바가 있었는데[33], 향후 민간선박이 자유롭게 움직이는 상황에서 군사적 목적의 활용이 식별되었을 경우 AI 활용에 있어 중요한 부분은 AI가 건전한 판단을 할 수 있는지, 지휘관은 이러한 AI의 판단을 신뢰할 수 있는지의 여부라 할 수 있다.

이상의 내용들을 고려해볼 때 군함과 민간 선박이 군사적 목적으로 혼재되어 있을 때 AI참모는 어떠한 판단으로 이를 명확히 구분할 것이며, 군사적 위협이 증가되는 상황에서 민간선박이지만 AI가 군사 목적으로 식별해서 위험하다고 판단하는 경우 미사일 타격을 건의할 수 있는지 등에 대한 고민이 필요하다. 특히, 해상에서 AI의 판단을 신뢰하기 위해서는 해당 AI의 수준을 고려할 필요가 있다. 단순 데이터 자동분석 등의 낮은 수준에서부터 작전수행 및 표적타격 추천, 자동 시행 등의 높은 수준에 이르기까지 AI를 구분하여야 하며 각 AI 수준별 윤리기준에 맞는 개발, 관리 및 운용이 이루어져야 할 것이다.

#### 4.2 해상 모자이크전 환경에서 군사목적 AI의 7대 윤리원칙 제안

앞서의 관련 연구 및 고려 요소 등을 종합적으로 판단하여 해상 모자이크전 환경에서 군사 목적의 AI 활용에 대한 7대 윤리원칙을 제안하고자 한다. 본 논문에서 제안하는 7대 윤리원칙은 다음과 같다.

##### 4.2.1 프라이버시(privacy)

미군의 AI 윤리원칙에는 포함되어 있지 않지만 향후 프라이버시에 대한 부분은 매우 중요하다고 생각된다. 더욱이 총 인구수의 감소로 기존에 인력이 하던 많은 부분이 AI가 포함된 무인체 등으로 대체될 가능성이 높다. 그러면 중요한 위치에서 작전임무를 수행하는 개인에 대한 정보보호가 필요하며 AI가 수집해서 분석하는 정보에 대한 프라이버시 관리가 필요할 것이다. 사적인 영역과 작전과 관련된 공적인 영역의 구분이 될 필요성이 있다.

##### 4.2.2 책임성(responsible)

책임을 표현하는 영어 용어가 AI 윤리원칙별 다른 부

분이 있어 미군에서 사용하는 영어 용어로 책임성을 따르는 것이 군의 입장에서 좀 더 명확하다고 생각된다. AI가 개발될 때 판단 기준과 근거는 결국 개발자의 사고와 판단에 많은 영향을 미치게 된다. AI 개발에 참여하는 인원들은 군사 목적의 활용임을 인식하고 책임을 지고 개발에 참여해야 할 것이다. 향후 AI가 내린 판단이 올바른지 아닌지에 대해 여러 상황에 추가적인 논의가 있을 수 있겠지만 그럴 경우에도 개발에 참여한 책임 있는 인원들을 통해 지속적 개선, 보완, 관리가 필요할 것이다.

##### 4.2.3 보안 및 신뢰성(security and reliable)

AI에 대한 적대적 공격이 가해지거나 또는 예상치 못한 예러 데이터가 AI 판단에 영향을 미칠 경우, AI가 추천 또는 시행하는 여러 결과에 전장의 승패를 좌우할 만큼 큰 실수로 이어질 수 있으며, 또는 민간을 대상으로 공격을 실시하여 전쟁에 대한 여러 비난 요소 및 악영향을 유발할 수 있을 것이다. 이를 위해 AI는 보안성을 가지고 외부 영향 요소를 최소화한 가운데 신뢰성을 바탕으로 동작되어야 한다. 보안 및 신뢰성은 AI를 운영하는 데 있어 매우 기본이 되는 요소라 할 수 있다.

##### 4.2.4 투명성 및 설명가능성(transparency and explainability)

미 국방부에서는 추적성(traceable) 항목을 투명하고 감사가 가능한 방법론을 적용하고 데이터 소스, 설계 절차 및 문서화를 포함하여 AI 개발 담당자가 AI 기술, 개발, 운영, 프로세스에 대한 이해를 가진 상태에서 AI 기능을 개발하고 배포되어야 하는 내용으로 언급하고 있다. 해당 항목을 AI 8대 윤리원칙 중 투명성 및 설명가능성에 정확히 일치한다고 보기는 어려우나 투명성이라는 부분에 유사성이 일부 있다고 볼 수 있다.

실제적으로 본 항목에서 언급하고자 하는 투명성은 다음과 같은 의미이다. AI 알고리즘에 따라서 도출된 결과는 그 결과에 이르기까지에 대한 충분한 정보가 제공되어야 하고, 그 일련의 과정들은 투명하게 관리되고 AI 행위자가 그 정보를 요청할 수 있는 기회가 주어져야 한다.

설명가능성은 투명성에 따라 제공되는 정보들, 즉 AI 알고리즘의 입력값, 결괏값, 해당 결괏값을 도출하게 된 이유 등을 사람이 이해할 수 있도록 설명하는 부분을 말한다. 지휘관이 작전임무를 수행하면서 AI가 추천한 각종 결과물에 대해서 본인의 군 작전 경험과 비교하여 쉽

게 납득이 가지 않는 경우 이러한 설명 결과를 바탕으로 본인의 판단을 수정/보완할 수 있을 것이다.

#### 4.2.5 공평성(equitable)

미군에서 언급하고 있는 공평성은 AI의 의도하지 않은 편견을 최소화하기 위해 신중한 조치를 취해야 하는 부분이다. AI 8대 윤리원칙에서 언급하고 있는 공정성 및 차별금지 AI 행위자들이 차별 없이 모두에게 AI의 혜택을 보장하고 수명주기 전반에서 차별적이고 편향된 관행을 최소화하며, 피해에 대한 구제 방안을 갖추는 부분으로 보고 있다.

본 논문에서 표현하고자 하는 공평성은 군의 특성을 포함한 내용이다. 해상 환경에서 작전은 함정 단위로 이루어지게 되며 단일 함정이라 할지라도 당시의 지휘권을 어떻게 가지느냐에 따라 작전임무 수행에 차이가 있을 수 있다. 예를 들어 함장의 임무수행이 제한되는 경우 부장이 그 임무를 대리하며 부장도 임무수행이 제한되는 경우 작전관이 수행하는 상황, 24시간 작전수행을 위해 당직 순번에 따라 긴급한 업무에 대한 임무를 수행하는 상황 등이 그러한 경우에 해당한다고 볼 수 있다. 이처럼 AI 시스템의 지원 대상이 특정 인원인지, 그 직책을 수행하는 당시의 인원인지에 따라 알고리즘 접근 방향이 달라질 수 있는데, 해상 모자이크전 환경에서는 특히 직책/임무 관점에서 공평성을 가지고 특정 대상이 아니라 당시의 해당 직책 및 임무를 수행하는 인원을 대상으로 지원하는 개념에서의 AI 활용 및 접근이 가능해야 할 것이다.

#### 4.2.6 통제성(governable)

통제성의 의미는 두 가지 관점에서 생각해 볼 수 있다. AI의 활동을 인간이 통제하는 부분, 그리고 AI 스스로 자정적 역할에서 통제하는 부분이 그것이다. 전자의 경우 UNESCO에서 언급하고 있다. UNESCO에서는 인간의 감독 및 결정 항목에서 인간이 AI 시스템에 의지할 수 있지만 최종적으로 인간의 몫인 책임 및 의무를 대신할 수 없으므로 AI 시스템을 관리하는 인원 또는 조직에 윤리적·법적 책임을 물을 수 있도록 해야 한다고 접근하고 있다. 후자의 경우 미 국방부 통제성 항목에서 언급하고 있다. AI를 개발할 때부터 의도치 않은 결과를 탐지·회피하는 능력과 의도치 않은 행위가 탐지되면 해당 시스템을 분리 또는 비활성화하는 능력을 보유한다는 내용이 이에 해당한다.

본 논문에서 제안하는 통제성은 이 모두를 포함하고 있다. AI의 활동을 인간이 통제할 수 있도록 최종적 결정 권한을 인간이 보유해야 하며 AI 스스로 자체적인 이상 탐지 기능을 보유하여 통제가 가능토록 구현되어야 할 필요성이 있다.

#### 4.2.7 적용성(applicability)

적용성과 관련된 윤리원칙이 유사하게 표현되거나 정의된 내용은 없으며, 본 윤리원칙은 해상 모자이크전 환경에서 군사 목적의 AI가 가져야 할 다소 특화된 윤리원칙이라 할 수 있다. 해상 모자이크전 환경에서 함정의 역할은 매우 다양하게 이루어진다. 예를 들어 최초 작전임무를 부여받았을 때 해상전투라는 기본 임무로 시작을 하였으나 예상외로 적의 규모가 방대하여 후방으로 이동하여 호송임무를 하다가 다시 격침된 아군 함정의 인원을 구조하는 임무를 하다가 인근 잠수함이 식별되어 탐지업무에 투입되다가 기뢰가 탐지되어 기뢰탐지 업무를 병행하는 등 하나의 작전만 수행하는 것이 아닌 당시의 임무와 역할에 따라 동시다발적 복합임무를 수행하는 경우 등이 이에 해당한다. 이처럼 복합임무를 수행하는 환경에서 AI가 그 모든 경우의 수를 예측하고 거기에 따른 AI 참모로서 조언을 제공하거나, 또는 시스템 자동화 관점에서 스스로 조정·통제·관리가 이루어지려면 함정에 상당한 규모의 AI 시스템이 탑재되어야 하고, 해당 AI 시스템이 정상적으로 동작하기 위해 수천 종의 센서 정보가 소통될 수 있는 통신 여건이 마련되어야 할 것이다.

함정의 공간을 효율적으로 활용하기 위해 체계통합 관점의 접근[34]이 추진되고 있는 상황에서 이러한 큰 규모의 시스템이 탑재될 수 있는 별도의 공간은 제한된다고 볼 수 있다. 만약 적정 규모의 AI 시스템을 함정에 탑재한다고 하더라도 최상급의 AI 수준과 그 역할을 기대하기는 어려울 것이다. 또한, 필요에 따라서 변화하는 작전환경에 특화하여 AI 자원을 해당 분야로 집중하여 활용하는 노력 등이 필요하다. 물리적 자원의 한계성을 고려해서 AI의 분석정보, 수준 등을 우선순위 지정 등을 통해 효과적으로 운용 및 관리해야 할 필요성이 있다. 이상의 내용을 고려하여 적용성이라는 원칙을 제안하고자 한다.

#### 4.3 접근방안

그렇다면 앞서 제안한 7대 윤리원칙을 실제 어떻게 적용할 것인가 하는 부분에 고민이 생긴다. AI 수준별 구분

을 통해 적용하는 것도 한 방법으로 검토될 수 있을 것이다. AI 수준별 구분을 통한 윤리기준은 Table 4와 같이 적용할 수 있다고 생각된다. 예를 들어 단순분석수준의 보고는 기본 윤리 수준으로 하는 등의 방식으로 접근할 수 있을 것이다. AI는 사람의 통제를 기본으로 하기 때문에 적용되는 윤리기준은 매우 기초적인 수준일 것이며, 관리자의 인식수준이 더 중요한 요소이기 때문이다. AI참모수준을 구분하는 데 있어 특히 표적정보를 활용하는 경우 윤리 수준은 매우 정확하고 높은 수준이 되어야 할 것이다. 현재 수치적으로 AI 정확도를 판단하는데 어려움이 있고, 기계적인 판단 이외에 국내외 정치적 상황, 작전적/전략적 수준의 내용들이 보완되어야 하므로 지휘관의 최종 판단을 지원하는 수준의 지원이 타당하다. 전략적 제대와 작전적/전술제대로 구분하여 적용하되 AI 수준별 구분[35]이 필요하다.

**Table 4.** 해군에서의 윤리적 AI 접근방안

AI 수준	적용규모	주요작전	윤리적용
레벨 1 (simple control program)	무인체 포함 (레벨 2 이상 제외)	단독작전	적용 수준 자체 판단, AI 개발 수준 고려 일부 생략 가능
레벨 2 (classical AI)	전단급/기함	합동/단독 작전	
레벨 3 (AI accepts mechanical learning)	작전사/함대사	연합/합동/ 단독작전	전 단계 적용 및 검증, 지속 관리
레벨 4 (AI embracing deep learning)	전략제대	연합/합동 작전	

AI참모 등과 같이 전반적인 예측의 영역을 포함하여 작전임무수행을 권고하는 입장의 경우 잘못된 AI 판단을 비판 없이 수용하였을 때 발생하는 문제가 매우 커질 위험이 있다. 해군의 경우 함정 기(既)탑재되는 AI 수준이 육상 시스템보다 낮은 수준을 가질 수밖에 없다. 하지만 최신의 전장환경 정보를 획득하여 즉각적으로 대응하기에는 함정이 유리한 측면이 있다. 함정이 식별한 최신의 정보교환이 위성 또는 무선으로 이루어지는 과정 속에 대용량의 분석 자료들을 끊임없이 보내기에는 한계가 있을 수 있다(날씨 등으로 무선환경이 나빠진 경우 등). 그럴 경우 제한된 정보로 분석(육상) 또는 제한된 수준의 장비/시스템/체계(해상)로 AI 분석이 이루어질텐데 최상의 상황을 가정한 AI 분석 수준을 100 % 신뢰하는 것은 관련

연구에서 본 바와 같이 예상치 못한 부작용(민간 공격 등)을 가지게 될 수 있을 것이다.

## 5. 결론

향후 미래전은 기존의 무기체계들이 매우 복잡한 형태로 필요에 따라 다양한 방식으로 운용되는 양상으로 진행될 가능성이 높다. 소수의 정교한 고가치의 무기체계 보다는 한두 개의 기능만 보유한 분리된 저가치의 전력 요소를 신속하게 구성 또는 재구성하여 전쟁이 진행될 것이다. 기존의 네트워크 중심의 정적인 형태의 킬체인(kill-chain)은 이미 적이 작전개념을 인식하고 무력화 수단도 갖추는 노력을 하고 있어 공격의 위험도가 높다. 이러한 위험도를 낮추기 위해 하나의 노드가 무력화되어도 작전운용이 가능한 동적인 킬웹(kill-web) 형태의 작전 능력이 필요하다. 최근 AI와 자율시스템을 작전개념과 연결하여 적보다 빠르고 효과적으로 의사결정이 되도록 하는 연구가 적극적으로 진행되고 있으며, 지휘관과 기계의 상호작용을 통해 가용전력을 분석하고 최적의 방책을 제공하는 자동화된 절차의 개발이 다양한 분야로 추진되는 상황이다.

이러한 부분을 종합하여 모자이크전으로 명명하는 연구들이 진행되고 있다. 해상의 경우 함정 자체를 유틸리티(utility)로 접근하고 다양한 복합전 임무를 수행하는 점을 고려하여 모자이크전 환경에서 단순화된 임무를 유연한 작전상황과 접목하여야 하며, 자율성을 가진 AI가 그 역할에서 큰 비중을 차지하게 될 것으로 예상된다. 미래의 해상 모자이크전을 개념적으로 생각해보면 임무별, 또는 해역별 작전수행을 위해 함정, 항공기를 포함한 다양한 전력이 해상에서 전투를 수행하는 모습을 그려볼 수 있을 것이다. 앞으로의 함정 역할은 기존의 센서와 무장을 단일 플랫폼에서 모두 수행하던 것에서 다양한 센서나 무장을 USV, UAV 및 UUV 등에 분산하고 역할을 나눠서 수행하는 방향으로 발전할 가능성이 높다.

AI의 경우 기본적으로 사전 입력된 규칙에 따라 관련 정보를 종합/분석하고 의견을 제시하는 점, 민간 주축으로 개발되어 군에 특화된 작전적 판단에 대한 의사결정 지원 신뢰성 문제 등이 우려되는 상황이다. 따라서, 해상 모자이크전 환경에서 군사목적의 AI가 효과적으로 활용될 수 있도록 윤리적 접근에 대한 고민이 필요한 실정이다.

이러한 상황에서 최근 여러 기관과 단체에서 발표한 윤리원칙을 살펴보고 해상 모자이크전 환경에서 군사목적

AI의 윤리원칙 7개 항목을 제안하였다. 제안된 AI 윤리 원칙은 향후 미래 해상작전 수행 시 준수되어 관리될 필요성이 있다. 물론 해당 원칙이 절대적으로 고정된 형태는 아니고 상황에 따라 가감이 가능한 부분이므로 발전적으로 적용할 수 있을 것이다. 다만 본 연구에서 말하고자 하는 부분은 앞으로 더욱 AI의 윤리적 문제에 대해 고민할 필요성이다. 다소 제한적일 수 있겠지만 본 연구가 해상 환경의 특수성을 이해하고 미래 모자이크전에서 향후 해상작전에 AI 운용 간 윤리적 문제점을 최소화하는데 효과적으로 활용될 수 있기를 기대한다.

## 참고문헌

- [1] 박효재, (2022. 2. 27). 우크라이나 전쟁 사이버 전선으로 확대 ... 우크라이나 'IT부대' 창설, 전세계 해커에 도움 요청, 경향신문, <https://www.khan.co.kr/world/world-general/article/202202271310001> (검색일: 2022. 3. 12)
- [2] 김진명, (2022. 3. 20). 美, 유럽에 최첨단 감시·정찰기 띄워 우크라이나 정보 제공, 조선일보, <https://www.chosun.com/international/us/2022/03/20/MN3JFMFYHJHDZFTDHR2BJGP2QM/> (검색일: 2022. 3. 21)
- [3] 장영은, (2022. 2. 27). 독일, '무기 수출 자제' 원칙 깨고 우크라이나 미사일 지원, 이데일리, <https://www.edaily.co.kr/news/read?newsId=01312006632234784> (검색일: 2022. 4. 2)
- [4] 허광환, (2019). "미국의 다영역 작전(Multi-Domain Operations)에 대한 비판과 수용." 군사연구, 147, pp. 125-153.
- [5] Defense Advanced Research Projects Agency (2015, March 30). Operating in Contested Environments, DARPA, <https://www.darpa.mil/news-events/2015-03-30> (검색일: 2022. 4. 12)
- [6] Stew Magnuson. (2018. 11. 16). DARPA Pushes 'Mosaic Warfare' Concept, National Defense, <https://www.nationaldefensemagazine.org/articles/2018/11/16/darpa-pushes-mosaic-warfare-concept> (검색일: 2022. 5. 4)
- [7] Defense Advanced Research Projects Agency (n.d). DARPA Tiles Together a Vision of Mosaic Warfare, <https://www.darpa.mil/work-with-us/darpa-tiles-together-a-vision-of-mosaic-warfare> (검색일: 2022. 5. 7)
- [8] Bryan Clark, Dan Patt, and Harrison Schramm, (2020). "Mosaic Warfare," Center for Strategic and Budgetary Assessments.
- [9] 장진오, 정재영, (2020). "미래전을 대비한 한국군 발전방향 제언: 미국의 모자이크전 수행개념 고찰을 통하여," 해양안보(Maritime Security), Vol. 1, No. 1, pp. 215-240.
- [10] Raymond McConoly, (2021. 2. 13). The Future of Naval Warfare to Define the Winner of the Big Power Struggle, Naval Post, <https://navalpost.com/the-future-of-naval-warfare-to-define-the-winner-of-the-big-power-struggle/> (검색일: 2022. 3. 7)
- [11] Dan Parsons (2021. 8. 5). Navy Considering Drone Delivery for Essential Parts At Sea, USNI News, <https://news.usni.org/2021/08/05/navy-considering-drone-delivery-for-essential-parts-at-sea> (검색일: 2022. 3. 7)
- [12] 배성수, (2020. 2. 20). "AI참모가 지휘 돕는 스마트 국방 머지않아," 한국경제, <https://www.hankyung.com/it/article/2022022055361> (검색일: 2022. 5. 21)
- [13] 김의순, (2020). "국방에서의 AI 적용, 미국 사례와 알아야 할 몇 가지," 한국국방연구원.
- [14] Ashton B. Carter, (2021. 11. 12). Autonomy in Weapon Systems (Department of Defense Directive No. 3000.09), US Department of Defense, <https://www.esd.whs.mil/portals/54/documents/dd/issuances/dodd/300009p.pdf> (검색일: 2022. 5. 21)
- [15] 김다솔, (2021. 8. 31). 美 드론 공격에 카불 일가족 10명 사망, 이데일리, <https://www.edaily.co.kr/news/read?newsId=02082806629152896> (검색일: 2022. 5. 21)
- [16] 김혜민, (2022. 3. 9). AI 유희당...꿈꿈이·화끈이의 분석은?, SBS, [https://news.sbs.co.kr/news/endPage.do?news\\_id=N1006669844](https://news.sbs.co.kr/news/endPage.do?news_id=N1006669844) (검색일: 2022. 3. 9)
- [17] OpenAI (2017. 2. 24). Attacking Machine Learning with Adversarial Examples, <https://openai.com/blog/adversarial-example-research/> (검색일: 2022. 4. 17)
- [18] Kashif Ahmad, Majdi Maabreh, Mohamed Ghaly, Khalil Khan, Junaid Qadir, and Ala Al-Fuqaha (2020) Developing Future Human-Centered Smart Cities: Critical Analysis of Smart City Security, Interpretability, and Ethical Challenges, arXiv (DOI: 10.48550/arXiv.2012.09110, 검색일: 2022. 4. 17)
- [19] Arunava Chakraborty, (2019. 10. 16). Introduction to Adversarial Machine Learning, FloydHub, <https://blog.floydhub.com/introduction-to-adversarial-machine-learning/> (검색일: 2022. 5. 13)
- [20] Sai Prabhu Gonigintala, (2020. 5. 21). Adversarial Attack Machine Learning, What After College, <https://whataftercollege.com/machine-learning/adversarial-attack-machine-learning/> (검색일: 2022. 5. 13)
- [21] Will Douglas Heaven, (2021. 11. 19). The Department of Defense is issuing AI ethics guidelines for tech contractors (美 국방부가 기술기업을 위해 AI 윤리 가이드라인을 마련한 이유), MIT Technology Review, <https://www.technologyreview.kr/departments-of-defense-government-ai-ethics-military-project-maven/> (검색일: 2022. 5. 13)
- [22] 김용래, (2019. 5. 22). "인공지능, 인권·민주주의 존중해야" 국제사회 AI원칙 첫 채택, 연합뉴스, <https://www.yna.co.kr/view/AKR20190522171100081> (검색일: 2022. 5. 24)
- [23] 방은주, (2021. 11. 24). 유네스코, 193개 회원국 만장일치로 AI윤리 권고 채택, ZDNet Korea, <https://zdnet.co.kr/view/?no=20211124105459> (검색일: 2022. 5. 27)
- [24] JAIC Public Affairs, (2021. 2. 26). AI Ethical Principles - Highlighting the Progress and Future of Responsible AI in the DoD, Joint Artificial Intelligence Center, [https://www.ai.mil/blog\\_02\\_26\\_21-ai\\_ethics\\_principles-highlighting\\_the\\_progress\\_and\\_future\\_of\\_responsible\\_ai.html](https://www.ai.mil/blog_02_26_21-ai_ethics_principles-highlighting_the_progress_and_future_of_responsible_ai.html) (검색일: 2022. 5. 27)

- [25] U.S. Department of Defense, (2020. 2. 24). Immediate Release: DOD Adopts Ethical Principles for Artificial Intelligence, <https://www.defense.gov/News/Releases/Release/Article/2091996/dod-adopts-ethical-principles-for-artificial-intelligence/> (검색일: 2022. 5. 27)
- [26] 박수형, (2020. 12. 23). '한국판 AI 윤리기준' 10대 원칙 담았다, ZDNet Korea, <https://zdnet.co.kr/view/?no=20201223105913> (검색일: 2022. 5. 27)
- [27] Jessica Fjeld, Nele Achten, Hannah Hilligoss, Adam Chirsofer Nagy, and Madhulika Srikumar, (2020). "Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI," Berkman Klein Center.
- [28] 오연주, 김소미, 권태균, (2021. 12). "유네스코 AI 윤리 권고 주요 내용 및 시사점," IT & FUTURE Strategy 보고서 제10호, 한국지능정보사회진흥원(NIA).
- [29] Army Times (2015. 8. 13), 7 hurt in Black Hawk 'hard landing' released from hospital, <https://www.armytimes.com/news/your-army/2015/08/12/7-hurt-in-black-hawk-hard-landing-released-from-hospital/> (검색일: 2022. 5. 29)
- [30] Unofficial UN Navy Site (n.d.) USNS Red Cloud (T-AKR 313), <https://www.navysite.de/akr/akr313.htm> (검색일: 2022. 5. 29)
- [31] 유영준, (2019. 2. 12). 예산 절감 영 해군, 민간선박 군함으로 개조, 연합뉴스, <https://www.yna.co.kr/view/AKR20190212055500009> (검색일: 2022. 5. 29)
- [32] 박수찬, (2021. 9. 21). 中, 자국 영해 진입 '위험 선박'에 사전 신고 의무화... 외교적 충돌 가능성, 조선일보, <https://www.chosun.com/international/china/2021/09/01/US6GEGP4PRB3BDAFKBKDO FVGWI> (검색일: 2022. 6. 1)
- [33] 백나리, (2018. 9. 19). [평양공동선언] 한강 하구 공동이용수역 설정... 민간선박 이용 보장, 연합뉴스, <https://www.yna.co.kr/view/AKR20180919099300014> (검색일: 2022. 6. 1)
- [34] 정승훈, 지해근, 최삼욱, 정남식, 임진국, (2020). "함정 체계통합 발전방향," Bulletin of the Society of Naval Architects of Korea, 57(1), pp. 15-20.
- [35] 정승훈, (2020). "해상 전자전 수행을 위한 인공지능(AI) 활용방안 연구," Journal of the KNST, Vol. 3, No. 2, pp. 118-124.