Journal of The Korea Institute of Information Security & Cryptology ISSN 1598-3986(Print)
VOL.33, NO.3, Jun. 2023 ISSN 2288-2715(Online)
https://doi.org/10.13089/JKIISC.2023.33.3.577

AI 기반 국방정보시스템 개발 생명주기 단계별 보안 활동 수행 방안

박 규 도,[1†] 이 영 란[2‡]
[1,2]국군방첩사령부 국방보안연구소 (연구장교, 정보보호연구실장)

A Methodology for SDLC of AI-based Defense Information System

Gyu-do Park,[1†] Young-ran Lee[2‡]
[1,2]Defense Security Institute, Defense Counterintelligence Command
(Research Officer, Director of Information Security Research Lab.)

요 약

국방부는 국방혁신 4.0 계획에 기반한 첨단과학기술군 육성을 위해 AI를 향후 전력 증강의 핵심 기술로 활용할 계획이다. 그러나 AI의 특성에 따른 보안 위협은 AI 기반의 국방정보시스템에 실질적인 위협이 될 수 있다. 이를 해소하기 위해서는 최초 개발 단계에서부터 체계적인 보안 활동의 수행을 통한 보안 내재화가 필요하다. 이에 본 논문에서는 AI 기반 국방정보시스템 개발 시 단계별로 수행해야 하는 보안 활동 수행 방안을 제안한다. 이를 통해 향후 국방 분야에 AI 기술 적용에 따른 보안 위협을 예방하고 국방정보시스템의 안전성과 신뢰성을 확보하는 데 기여할 수 있을 것으로 기대한다.

ABSTRACT

Ministry of National Defense plans to harness AI as a key technology to bolster overall defense capability for cultivation of an advanced strong military based on science and technology based on Defense Innovation 4.0 Plan. However, security threats due to the characteristics of AI can be a real threat to AI-based defense information system. In order to solve them, systematic security activities must be carried out from the development stage. This paper proposes security activities and considerations that must be carried out at each stage of AI-based defense information system. Through this, It is expected to contribute to preventing security threats caused by the application of AI technology to the defense field and securing the safety and reliability of defense information system.

Keywords: Artificial Intelligence, AI Security, Defense Information System, Defense Acquisition, SDLC

I. 서 론

AI 기술은 4차 산업혁명 시대의 알파이자 오메가로 대한민국의 경제·사회·문화 全 영역에 영향을 끼치고 있으며, AI를 활용해 기존 기술 또는 제품을 고도화하고 부가가치를 창출하기 위한 재정적 투자가 활발히 이루어지고 있다. 맥킨지 연구소는 2030년까지 전 세계 기업의 70%가 AI를 활용할 것이며 이에 따른 글로벌 GDP 추가 성장액이 13조 달러에 이를 것으로 전망하였다[1]. 국방 분야 또한 AI와 기존 기술을 융합하려는 시도를 적극적으로 추진 중으로, 국방부는 적의 비대칭 위협 대응과 전투형 강군 육성을 위해 국방혁신 4.0 기본계획을 발표하고 향후 원격 통제와 자율 운용이 가능한 AI 기반의 유무인 복합 전투체계를 확대 구축할 계획을 밝힌 바 있다[2]. 이에 따라 각종 무기·전력지원체계의 핵심

Received(04. 03. 2023), Modified(05. 03. 2023), Accepted(05. 25. 2023)
† 주저자, brown4255@kakao.com
‡ 교신저자, yrleemath@gmail.com(Corresponding author)

또는 일부를 구성하는 국방정보시스템과 AI 기술의 결합이 국방 임무 전 분야에 다양한 형태로 활발하게 추진될 전망이다.

그러나 AI에는 편향성·불투명성 등 기술적 특성으로 인한 취약성이 존재하며 이로 인해 중독·적대적 공격과 모델 또는 학습 데이터의 무단 유출 문제 등 다양한 보안 위협이 발생할 수 있다. 더욱이 AI 보안 위협은 데이터 수집부터 모델 학습까지 AI 개발 과정 전체에서 발생할 수 있으며, 최초 잘못된 데이터 수집의 결과가 완성된 AI 시스템의 운용 과정에서 발견되는 등 문제 발생의 원인 분석과 예측이 어렵다는 특징이 있다. 따라서 AI 시스템에 대한 보안성 확보를 위해서는 최초 개발 과정에서부터 필요한 보안 활동을 식별하고 수행하며 그 결과를 검증하는 보안 내재화 개발이 필요하다.

이에 우리나라를 비롯한 주요 선진국 정부 및 공공기관은 '신뢰할 수 있는 AI' 또는 '윤리적 AI'를 목표로 보안적 측면에서의 다양한 고려 사항을 제시하고 이를 정책화하기 위한 노력을 지속하고 있으며[3][4][5], 미 국방부 또한 AI 윤리 원칙을 통해 책임있고 합법적인 군사적 AI의 개발 및 활용을 위한 다양한 보안 관점에서의 키워드를 제시한 바 있다[6]. 반면 우리 군의 국방정보시스템 도입 및 개발 절차를 살펴보면 대규모 SW 개발 시 시큐어 코딩 적용, 취약점 평가를 비롯한 보안대책 준수 여부 점검 등 소프트웨어 보편적인 보안 제도가 마련되어 있으나, AI만의 기술 특성을 고려한 체계적이고 통합된 보안 지침은 제도화되어 있지 않은 실정이다. 보안이 무엇보다도 중요한 가치를 지니는 국방정보시스템에서 AI 도입에 따른 보안 위협을 해소하지 못한다면 AI 기술에 대한 신뢰성 저하되고 기술 수용성이 약화되어 결국 국방 분야에 AI의 적극적인 도입 및 활용을 가로막는 장애물이 될 수 있다.

기존에 수행된 AI 보안에 관련된 연구는 데이터 및 모델의 알려진 개별 취약점에 대한 방어 기법[7][8][9]이 주를 이루며, SDLC 개념을 접목한 AI 개발방법론 연구[10][11]의 경우 국방 분야 및 우리 군의 특성을 반영하지 못하였다는 한계점이 있다. 반대로 AI의 국방 분야 활용에 따른 보안 연구[12][13]는 양질의 학습 데이터 확보를 위한 보안 규제 완화 등과 같이 주로 제도적 측면에서 접근하였으며 AI 보안을 고려한 군 특화 개발방법론의 필요성은 언급하였으나 구체적인 실행 방안을 제시하지는 못하였다.

따라서 본 논문에서는 먼저 국방정보화업무훈령 등 관련 훈령 및 규정과, 국제표준 및 주요 선진국 공공기관에서 제시한 AI 개발 생명주기를 바탕으로 국방 AI 개발공정을 정의하였다. 다음으로 주요 AI 보안 위협을 정리하고 이를 바탕으로 AI 기반 국방정보시스템에서 발생 가능한 보안 위협 시나리오를 제시하였다. 최종적으로 국방 AI 개발공정 및 보안 위협 시나리오에 기반하여 AI 기반 국방정보시스템의 보안 내재화를 위한 개발 생명주기 단계별 보안 활동 수행 사항을 도출하였다.

본 논문의 구성은 다음과 같다. 1장은 서론, 2장은 관련 제도 및 표준으로 우리 군 훈령과 각종 AI 개발 생명주기 정의에 기반한 국방 AI 개발공정의 정의, 3장은 AI 특성에 따른 보안 위협과 AI 기반 국방정보시스템에서 발생 가능한 보안 위협 시나리오를 도출하였다. 이어 4장에서 AI 기반 국방정보시스템 개발 단계별로 수행해야 하는 보안 활동을 제시하였고, 5장에서는 결론을 맺는다.

II. 관련 제도 및 표준

2.1 AI 기반 국방정보시스템의 정의 및 국방 SW 표준 개발공정

국방정보화훈령에서는 국방정보시스템을 국방 무기·전력지원체계의 일부분을 구성하며 국방정보의 수집, 가공, 저장, 검색, 송·수신 및 그 활용과 관련되는 기기 등 응용소프트웨어와 기반운영환경의 조직화된 체계로 정의하고 있다[14]. 한편 지능정보화기본법에서는 AI를 전자적 방법으로 데이터를 수집·분석·가공하고 이를 바탕으로 학습·추론·판단 등을 구현하는 기술로 정의하였다[15]. 이를 바탕으로 본 논문에서는 AI 기반 국방정보시스템을 '국방 데이터를 전자적으로 수집·분석하고 이를 학습하여 추론과 판단 기능을 구현할 수 있는 국방정보시스템으로 무기·전력지원체계의 일부를 구성하는 것'으로 정의하였다.

국방정보시스템의 획득 절차는 무기체계로 분류된 경우 국방전력발전업무훈령 및 방위사업관리규정을, 전력지원체계로 분류된 경우 국방정보화업무훈령을 따른다[16]. 마찬가지로 AI 모델을 포함한 소프트웨어 부분에 대한 개발 절차는 무기체계 소프트웨어 개발 및 관리 매뉴얼 또는 국방정보화업무훈령을 따르며 세부 절차는 Table 1과 같다.

Table 1. Standard Development Process of Defense Software

Stage	Jobs
① Prepare	· Select Software Lifecycle Model · Plan for Development
② Analysis	· Define and Review System Software/DB Requirement · Create Specification
③ Design	· Design Inteface and DB · Define Unit/Integrated Test Requirements & Plan
④ Implement & Test	· Developmental Test · Certificate Software/System
⑤ Delivery	· System Installation · Operational Test

더불어 소프트웨어 개발 시에는 각 단계별로 지정된 산출물을 생산해야 하는데, 보안과 관련된 산출물로는 상세설계 단계에서의 시스템 보안설계서와 시험평가 결과에 따른 정보보호 적합성 평가 결과, 보안측정 및 조치 결과 등이 있다.

이외에 국방정보시스템 개발 및 운영에 따른 보안 관련 규정으로는 최초 사업계획 수립 시 SW에 대한 보안 요구사항을 식별하고 검토 결과를 포함하여야 하며 용역업체는 제안요청서 내의 기술적용 계획표를 통해 보안 분야에 대한 세부 기술 적용계획을 명시하여야 한다. 또한 국방보안업무훈령과 국방사이버안보훈령 등에 근거한 상호운용성 평가, 취약점 점검 등을 수행해야 하며 운용 단계에서는 국방정보시스템 보호요구사항을 충족해야 한다[17]. 그러나 상기 제도에도 불구 소프트웨어를 구성하는 기술적 요소에 대한 구체적인 보안 지침은 별도로 명시하고 있지 않다. 특히 AI는 기술적 특성과 이에 따른 보안 위협이 명확하므로 별도의 개발 단계를 정의하고 각 단계별 보안 기준과 지침을 수립하여야 하나 이와 관련된 국방 분야 훈령 및 규정은 미흡한 실정이다.

2.2 AI 개발 생명주기 관련 표준

AI 시스템의 개발 생명주기는 국제 표준(ISO), 美 NIST(National Institute of Standards and Technology), OECD(Organization of Economic Cooperation and Development) 등 다양한 기관에서 정의하고 있다.

먼저 국제 표준에서는 AI 개발 주기를 Figure 1

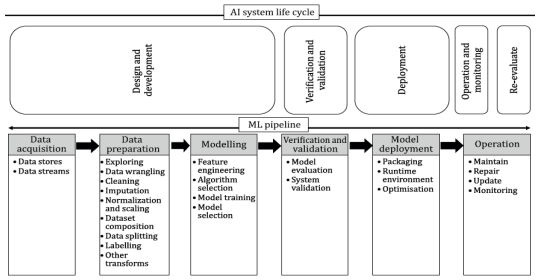

Fig. 1. ISO 23053, AI Systems Life Cycle

과 같이 ① 설계 및 개발(Design and Development), ② 검증 및 확인(Verification and Validation), ③ 배포(Deployment), ④ 운영 및 모니터링(Operation and Monitoring), ⑤ 재평가(Re-evaluation)의 5단계로 정의하였다[18]. 더불어 ML(Machine Learning) 기반의 AI를 구축하는 경우를 상정하여 설계 및 개발 단계 수행 절차를 ① 데이터 획득(Data Acquisition), ② 데이터 준비(Data Preparation), ③ 모델링(Modeling)으로 세분화하였는데, 데이터의 수집 및 학습 과정을 제시하였다는 점에서 ML 알고리즘을 사용하지 않는 AI 시스템의 개발 시에도 동일한 절차를 적용할 수 있다.

美 NIST의 AI RMF(Risk Management Framework)에서는 AI 개발 생명주기를 Figure 2와 같이 ① 사전설계(Pre-Design), ② 설계 및 개발(Design & Development), ③ 배포(Deployment), ④ 시험평가(Test & Evaluation)의 4단계로 정의한다[19].

사전설계 단계에서는 데이터의 수집, 분류, 구조화, 선별 및 문제 정의, AI 체계와 관련된 이해관계

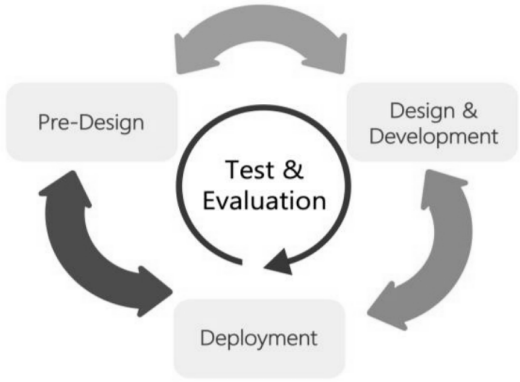

Fig. 2. NIST, AI System Lifecycle

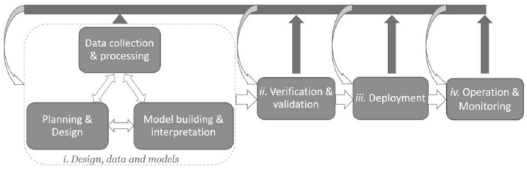

Fig. 3. OECD, AI System Lifecycle

자의 정의 등이 이루어진다. 설계 및 개발 단계에서는 데이터 분석 및 정제, 모델링 및 요구사항 분석 등을 수행하고, 배포 단계에서는 사용자 피드백 및 재정의, 운용 중인 AI 체계의 모니터링 및 폐기 등을 수행한다. 시험평가 단계는 상기 3개의 생명주기 단계 전반에 걸쳐 기술적 검증 및 확인(Verification & Validation)을 수행한다.

OECD에서는 AI 생명주기를 Figure 3과 같이 ①설계, 데이터 및 모델링(Design, Data & Modeling), ②검증 및 확인(Verification & Validation), ③배포(Deployment), ④운영 및 모니터링(Operation & Monitoring)의 4단계로 정의하였다[3].

각 단계별 세부 수행 내용을 살펴보면 첫 번째로 설계, 데이터 및 모델링 단계에서는 시스템의 컨셉과 목적, 전제 및 요구사항 등을 기반으로 프로토타입을 구축하는 시스템 계획 및 설계, 데이터를 수집 및 정제하여 품질과 완성도를 확인한 후 데이터셋의 특징을 문서화하는 데이터 수집 및 가공, 모델 및 알고리즘 선택, 세부 파라미터 수정, 결과 해석 등이 이루어지는 모델 구축 등의 작업을 순차적 또는 비순차적으로 수행한다. 검증 및 확인 단계에서는 구축된 모델을 실행하고 다양한 방식으로 검증 또는 수정하며, 배포 단계에서는 기존 시스템에 AI 모델 적용 시 호환성과 법(규정) 위반 여부 검증, AI 적용에 따른 조직의 업무 절차 및 사용자 경험 변경 사항 등을 확인한다. 운영 및 모니터링 단계는 AI 시스템이 원래의 목적에 맞게 지속해서 운영되고 있는지, 윤리적 문제가 발생하지 않는지 등을 확인하며, 문제 발생할 시 이전 단계로 돌아가 AI 체계를 수정하거나 AI 체계 자체를 폐기하는 방안을 검토한다.

2.3 국방 AI 개발공정

앞서 제시한 AI 개발 생명주기의 각 단계별 수행 활동을 종합한 후 국방 표준 SW 개발공정의 단계별 수행 내용과 유사성을 비교 분석하여 Table 2와 같이 국방 AI 개발공정을 정립하였다.

국방 AI 개발공정은 국방 분야에서 개발 및 활용되는 AI 모델이 국방정보시스템 응용소프트웨어의 일부로 포함되는 점을 감안, 국방 표준 SW 개발공정과 유사한 절차를 따르도록 구성하였다. 이를 통해 AI 개발이 별도의 독립된 절차가 아닌 국방 응용소프트웨어 개발의 일부로서 함께 개발되어 개발 효율성 향상을 도모할 수 있을 것으로 생각된다. 또한 4장에 서술할 AI 기반 국방정보시스템 개발 단계별 보안 활동 수립 시에도 본 개발공정을 기반으로 함으로써 보안이 내재화된 AI 개발 절차를 수립하는데 기여할 수 있을 것이다. 단, 국방 표준 SW 개발공정의 ⓐ준비 단계는 응용소프트웨어의 수명주기 모델을 선택하고 사업수행계획서를 작성하는 등 시스템 전반의 개발을 준비하는 단계로 직접적인 AI 개발의 절차에는 포함되지 않는다.

국방 AI 개발공정의 ①설계 단계는 국방 표준 SW 개발공정의 ⓑ분석 및 ⓒ설계 단계에 해당되며 AI 모델의 기능 요구사항 및 성능 목표치 설정, 학습 데이터 수집을 위한 데이터 세부 구성(메타데이터) 항목 선정, 모델 개발을 위한 SW 및 알고리즘 선택, 성능 평가 지표 선정 및 수행 계획 작성 등이 이루어진다. ②데이터 전처리와 ③모델링 및 평가는 국방 표준 SW 개발공정의 ⓓ구현 단계에서 수행한다. 이 단계에서는 수집된 데이터의 전처리, AI 모델링, 결과 해석 및 세부 파라미터 수정 등 데이터 가공 및 모델 학습을 수행하며 그 결과를 시스템 통합시험 결과보고서 등 개발 산출물에 포함해야 한다. 마지막 ④운영 및 모니터링 단계는 국방 표준 SW 개발공정의 ⓔ인도 단계에 대응되며 AI 모델의 정상적인 동작 여부를 지속 모니터링하고 사전 설정한

Table 2. Comparison of Development Process of Defense Software & AI

Dev. Process of Defense Software	Dev. Process of Defense AI
ⓐ Prepare	-
ⓑ Analysis	① Design
ⓒ Design	
ⓓ Implement & Test	② Data Preparation
	③ Modeling & Evaluation
ⓔ Delivery	④ Operation & Monitoring

조건에 따라 유지보수 또는 재학습을 수행해야 한다.

III. AI 보안 위협

3.1 주요 AI 보안 위협

3.1.1 중독 공격(Poisoning Attack)

국제표준(ISO)에 따르면 AI를 대상으로 한 특징적 보안 위협은 크게 중독 공격, 적대적 공격, 프라이버시 위협 등 세 가지로 분류할 수 있다[20]. 우선 중독 공격은 훈련 데이터를 고의적으로 조작하여 잘못된 모델링을 유도하는 공격 방식으로, 성공할 경우 AI 시스템이 오동작이나 잘못된 데이터 입력 등을 감지할 수 없게 된다. 중독 공격에 사용되는 조작된 데이터를 Poison Data라고 하며, 최소한의 데이터 변조를 통해 최대한의 오동작을 일으키는 것을 공격 목표로 한다[7]. 특히 인터넷 연결 환경에서 지속적으로 재학습이 이루어지는 형태의 모델이 중독 공격에 가장 취약하며, 독립망 등의 분리된 네트워크 환경이라도 본 공격에 노출 시 모델 재학습이 필요하다. 단, 전제 조건으로 공격자가 공격 대상 AI 체계의 학습 알고리즘과 데이터를 알고 있어야 하며, 공격 시뮬레이션이 가능해야 한다[21]. 이외에도 중독 공격은 백도어 공격과 같은 후속 응용 공격의 준비 단계로도 악용할 수 있다. 백도어 공격이란 오작동을 유발할 수 있는 데이터를 미리 학습한 후, 완성된 AI에 해당 데이터를 입력하여 실제로 오작동을 유발하는 공격 방식이다[22].

3.1.2 적대적 공격(Adversarial Attack)

중독 공격이 AI 모델 구축 前 단계의 학습 데이터를 대상으로 하는 반면 적대적 공격은 이미 완성되어 운영 중인 AI 모델(체계)에 악의적인 데이터를 입력하여 AI의 오동작을 유도하는 공격 방식을 지칭하며, 회피(Evasion)· 기만(Deception)·입력(Input) 공격으로도 불린다. 사람의 눈에 보이지 않는 극소량의 데이터 변조만으로도 AI가 잘못된 판단을 일으킬 수 있으며[23], 이미지 이외에도 음성, 텍스트, 의료 등 다양한 AI 응용 분야에서 적대적 공격이 가능하다[24]. 또한 이미 개발이 완료되어 운용 중인 AI 체계를 대상으로 공격이 가능하므로 중독 공격에 비해 보다 넓은 공격표면을 갖는 특징이 있다. 적대적 공격의 세부 유형은 공격자의 목표가 특정 결과값으로의 오분류 혹은 임의의 결과로 오분류인지에 따라 표적 공격 또는 무표적 공격으로 나누거나, 공격 대상 모델의 정보량 등에 따라 화이트박스, 그레이박스, 블랙박스 공격 등으로 나눌 수 있다[8][25].

3.1.3 프라이버시 위협(Privacy Threat)

프라이버시 위협 또는 탐색(Exploratory) 공격은 AI 체계에 포함된 민감 데이터를 유출할 수 있는 공격이다. 중독·적대적 공격이 AI 체계에 특정한 값을 '입력'함으로써 오동작을 유도하였다면, 프라이버시 위협은 반대로 AI 체계가 제공하는 '출력' 값을 악용하여 AI 모델이나 학습 데이터에 포함된 민감 정보를 획득한다. 프라이버시 위협의 유형은 학습에 사용된 실제 데이터를 추출하는 모델 도치(Model Inversion), 특정 데이터가 학습 데이터 셋에 포함되었는지 여부를 확인하는 회원정보 추론(Membership Inference), 학습 모델의 파라미터 값 또는 모델 자체를 획득하는 모델 추출(Model Extraction) 공격 등이 있으며[9], 데이터 수집 단계, 학습 단계, 검증 단계 등 AI 생명주기 전체에서 보안 위협이 될 수 있다.

3.2 AI 기반 국방정보시스템 보안 위협 시나리오

2장에서 제시한 다양한 유형의 AI 보안 위협은 AI를 적용하는 국방정보시스템에도 실질적인 위협이 될 수 있다. AI 기반 국방정보시스템의 특징은 크게 ① 외주 용역을 통한 시스템 개발 및 운영, ② 이미지・영상판독 분야의 AI 활용 다수, ③ 군사(비밀)자료를 활용한 AI 학습 등 세 가지로 정의할 수 있다. 이를 바탕으로 3.1에서 제시한 주요 AI 보안 위협과 결합하여 AI 기반 국방정보시스템에서 발생 가능한 보안 위협 시나리오를 다음 세 가지로 제시하였다.

3.2.1 AI 개발 공급망 공격을 통한 개발용 SW 및 학습 데이터 오염

현재의 국방정보시스템은 대부분 군내 인력이 아닌 외주 용역업체를 통해 개발되므로 이 과정에서 필연적으로 공급망 위험에 노출될 가능성이 높다. AI 개발 시에는 데이터 저장 및 전처리, 모델링, 테스팅

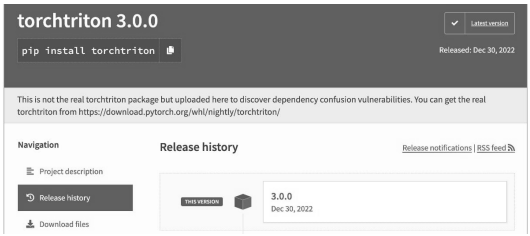

Fig. 4. Malicious PyTorch Dependency 'torchtriton' on PyPI Repository[27]

및 배포 등 AI 개발 주기 전 과정에서 다양한 오픈소스 SW가 활용되며 이 과정에서 해당 오픈소스에 취약점이 내재될 경우 이를 활용해 개발된 AI 시스템 전체가 보안 위협에 노출될 수 있다.

이와 관련하여 2022년 12월 머신 러닝 개발 프레임워크인 PyTorch의 저장소가 의존성 혼동(Dependency Confusion) 공격을 받아 주요 시스템 정보를 탈취하는 악성 패키지 torchtriton이 Figure 4와 같이 업로드된 사례가 있다[26].

마찬가지로 외주 개발업체가 개발 기간 단축 등을 이유로 공개된 데이터셋이나 사전 학습된 모델을 활용할 수 있다. 이때 공격자가 국방 AI에서 많이 활용되는 이미지 또는 영상 데이터셋에 악성 데이터를 포함하여 공개된 저장소에 업로드하고, 외주 개발업체가 이를 활용할 경우 해당 AI 시스템이 중독 공격에 노출될 가능성이 있다.

3.2.2 운용 중인 AI 기반 국방정보시스템 대상 적대적 공격 수행

두 번째로 이미지·영상 판독 기반의 AI 시스템을 대상으로 하는 적대적 공격 시나리오를 생각해 볼 수 있다. 예를 들어 기지경계용 CCTV 시스템이 관제 중인 구역에 적대적 공격이 가능한 형태의 스티커를 부착하거나 가짜 이미지를 투사하여 시스템의 오동작을 일으킬 수 있다. 마찬가지로 공격용 드론에 부착된 카메라를 공격하여 아군기나 민간인에 대한 오인 사격을 일으키는 시나리오도 가능하다.

정찬호 등[28]은 Figure 5와 같이 Shadow Adversarial Example 기법을 통해 원본 전투기 사진에 대한 적대적 이미지를 생성하고 이를 VGG19 모델을 통해 분류한 결과 까치로 오인식되는 실험 결과를 얻었으며, 이를 통해 국방 분야에서 자주 사용하는 이미지 데이터에 대해 적대적 공격이 가능함을 증명한 바 있다. 이를 통해 상기한 CCTV

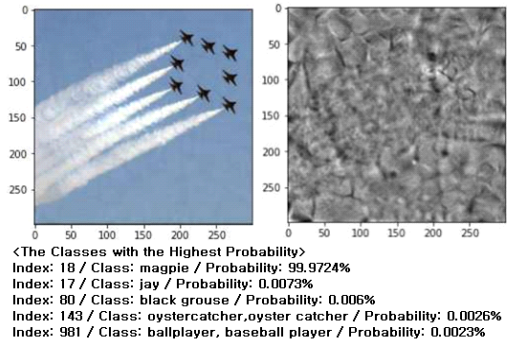

Fig. 5. Classification Result of Fighter Image Created by Adversarial Sample[28]

및 드론봇 등 실세계를 영상으로 관찰하는 형태의 AI 체계를 다수 운용하게 될 우리 군에 적대적 공격이 실질적인 위협이 될 수 있음을 확인할 수 있다.

3.2.3 탐색 공격 등을 통한 군사(비밀)자료 유출

AI 기반 국방정보시스템의 학습 데이터로 군사(비밀)자료를 사용하는 경우를 상정하였을 때, 접근통제, 데이터 무결성 관리 등 보안 규정을 준수한다면 중독 공격에 직접 노출될 가능성은 비교적 적은 반면 탐색 공격 등을 통해 모델이나 학습 데이터가 공격자에게 유출되는 가능성은 상존한다. 특히 공격자가 모델의 기초 정보를 모르는 상태에서 입력에 대한 출력값 획득만으로 학습 데이터 유출이 가능한 블랙박스 공격에 대한 다양한 실증 사례가[29][30] 존재하므로 이 역시 우리 군에 현실적인 위협이 될 수 있다. 다만 시스템의 출력 결과를 군 내부의 인가된 인원만이 얻을 수 있는 국방정보시스템의 특성상 공격 조건이 앞서의 두 가지 시나리오에 비해서는 까다로울 수 있다. 그럼에도 불구하고 군 내부자에 의한 위협이나 AI 시스템의 유지보수 등을 위해 모델 데이터를 군 외부로 반출하게 될 경우 등에 대해서는 고려가 필요하다.

IV. AI 기반 국방정보시스템 개발 생명주기 단계별 보안 활동 수행 절차

본 장에서는 AI 기반 국방정보시스템 개발 시 수행해야 하는 보안 활동을 Table 3과 같이 제안한다. 앞서 제시한 국방 AI 개발공정을 기반으로 국내외 AI 관련 보안제도 및 지침과 AI SDLC 관련 선

Table 3. Security Activity in Development Process of AI-based Defense Information System

Dev. Process of Defense Software	Dev. Process of Defense AI	Security Threats to Consider	Security Activity	ref.	
ⓐ Prepare	-		-	-	
ⓑ Analysis	① Design	Supply Chain Attack	· Cost/Benefit Analysis of applying AI to System · Set AI Stakeholders · Set Access Control Policy for each Development Stage · Scan Vulnerability of SW for AI Dev.	[3] [4] [5] [6] [14] [19]	[34]
ⓒ Design					
ⓓ Development & Test	② Data Preparation	Poisoning Attack	· Analysis of Data Quality/Quantity · Data Preprocessing/Auditing		[31] [33]
	③ Modeling & Evaluation	Adversarial Attack	· Scan & Fix Model Security Threat · Apply AI Explainability · Test & Evaluation		[34] [37] [39]
ⓔ Operation	④ Operation & Monitoring	Privacy Threat	· Monitor & Review of Sys. Operation · Set Procedure of Model Retraining · Manage Data History · Documentation & Sharing to AI Stakeholders		[31]

행연구 등을 기반으로 우리 군의 실정에 부합하는 단계별 세부 보안 활동을 선정하였다. 또한 각 보안 활동들을 국방 SW 표준 개발공정 각 단계별로 분류함으로써 AI 기반 국방정보시스템의 보안 내재화 개발을 가능케 하고 시스템이 전력화되기 전 보안 위협을 사전 제거함으로써 AI 기술 적용에 따른 보안 위협을 예방하고 국방정보시스템의 안전성과 신뢰성을 확보하는 데 기여할 수 있을 것으로 기대한다. 이하에서는 각 단계별 세부 보안 활동과 우리 군 특성을 반영한 이의 적용 방안을 함께 제시한다.

4.1 설계 단계

설계 단계에서 수행해야 할 보안 활동은 ①AI 기술 적용 가능성 검토, ②AI 체계 이해관계자 선정, ③ 개발 과정에 필요한 접근 통제 절차 수립, ④AI 개발용 SW 취약점 점검의 4가지로 정의하였다.

먼저 AI 기술의 적용 가능성 검토란 AI 기술의 국방정보시스템 적용이 해당 체계의 운용 목적에 부합하고 궁극적으로 군의 임무 수행 능력 향상에 기여할 수 있을지 판단하는 절차이다. 이를 위해 국방 CIO 협의회 등 국방정보체계 소요 심의 및 의사결정 기구가 국방정보시스템 소요 심의 시 AI 기술 도입의 적절성을 함께 검토하는 방안을 생각해볼 수 있다. 이때 검토 기준은 훈령 등에 근거하여 경제성(AI 도입 비용 대비 편익), 개발 가능성(양질의 학습 데이터 확보 가능성 등), 보안성(AI 관련 보안 위협 예방 및 대응 계획), 상호운용성(기반 시스템-AI 모델 또는 AI 시스템-타 시스템 간의 입·출력 데이터 연계 방안 등) 다양한 측면을 고려해야 한다. 만약 AI 도입이 본래의 목적을 달성하지 못하거나 효과성이 떨어짐에도 불구하고 무리하게 AI 기술을 도입하는 경우 국방정보시스템에 공격 표면이 추가되어 보안 관점에서 부작용을 초래할 수 있으므로, 초기 시스템 도입 검토 단계에서부터 상기한 내용을 포함한 다각적인 고려가 필요하다.

다음으로 AI 시스템의 이해관계자를 선정하여야 한다. 이해관계자란 시스템 운영 및 설계자, 데이터 과학자, 모델 엔지니어 등 AI 개발 과정에서 특정한 역할을 수행하는 인원을 의미한다. 특히 보안 관련 업무를 수행하는 이해관계자는 AI의 데이터, 모델 및 시스템 전반의 보안 취약요소를 지속적으로 점검하며 문제 발생 시 수정 또는 모델 재학습 등의 판단을 내릴 수 있어야 한다. 더불어 이러한 판단이 의사결정 과정에서 반영될 수 있도록 제도와 조직을 함께 구축해야 한다. 현재 우리 군은 AI 이해관계자에 대한 규정화된 개념과 직책·직무별 역할이 정의되어 있지 않은 실정이나, 해당 AI 시스템의 사용 주무부서·유지보수 기관·장비 운영기관[14] 등과 더불어 시스템이 운영되는 분야의 전문가 및 데이터 관리자,

정보통신 보안 및 시스템 시험평가 담당자 등을 국방 AI 이해관계자에 포함하는 방안을 생각해볼 수 있다.

세 번째로 개발 단계별 접근통제 정책을 국방 보안 관련 훈령 등에 근거하여 수립 및 시행해야 한다. 민감한 군사정보를 취급하는 국방정보시스템의 특성상 데이터 취급자, 개발자, 사용자 등 각 주체별로 접근 가능한 범위가 다를 수 있으며 데이터의 결합 전과 결합 후, 모델 생성 전과 후의 보안 등급이 변화할 수 있다. 이에 시스템 보안책임자를 중심으로 개발 단계별로 정보에 접근 가능한 주체와 경로를 설정하여 접근통제 정책을 수립하고, 모든 개발 단계에서 이의 준수 여부를 지속 관리·감독해야 한다.

마지막으로 AI 데이터 처리 및 모델 학습에 사용되는 SW 및 라이브러리의 상당수가 오픈소스인 점을 감안, 공급망 공격에 대비할 필요가 있다. 이를 위해 체계 설계 단계부터 사용이 예상되는 개발 관련 프로그램에 대한 목록을 관리하고 이에 대한 취약점 점검 및 형상 관리가 동반되어야 한다. 따라서 AI 개발 SW에 대한 보안(취약점) 점검 업무 세부 수행 지침·절차와 수행조직 등을 국방보안업무훈령 및 국방사이버안보훈령 등에 반영해야 한다.

4.2 데이터 처리 단계

데이터 처리 단계에서 수행해야 할 보안 활동은 ① 충분한 데이터 수집 여부 분석, ② 데이터 전처리 및 감사의 2가지로 정의하였다. 이를 통해 중독 공격을 예방하기 위한 데이터 편향성과 악의적 데이터를 제거하는데 그 의의가 있다.

우선 AI 성능 목표를 달성할 수 있는 양질의 학습 데이터가 수집되었는지를 확인할 필요가 있다. AI 학습을 위한 데이터 수집 과정에서는 AI 체계의 운용 목표와 특성에 가장 부합하는 대푯값을 찾는 것이 힘들고, 특히 데이터의 출처가 다양할 경우 정규화 및 가중치 부여 문제 등이 발생할 수 있다[3]. 이에 데이터가 과부족하거나 편향성을 띄는 경우 과소적합(Underfitting), 과다적합(Overfitting) 등의 문제가 발생할 수 있으며, 이는 곧 중독·적대적 공격 등 다양한 보안 위협에 취약한 결과를 낳을 수 있다[7]. 또한 데이터의 결함은 일단 모델 개발이 시작된 이후에는 발견하기 힘들다는 제한사항이 존재하므로 학습 전에 양질의 데이터 수집 여부를 면밀히 점검하고 데이터 정제를 수행할 필요가 있다.

다음으로 데이터 전처리 및 감사 과정에서는 데이터의 특성을 나타낼 수 있는 메타데이터를 확보하고 AI 시스템 활용 분야를 고려한 데이터 정제 기준을 수립해야 한다[31]. 예를 들어 경계감시용 CCTV의 경우 이미지 데이터의 학습 및 처리를 위해 촬영일시·위치, 조리개 값 등의 메타데이터 항목을 선정하고 이미지 크기·비율·화질 등을 기준으로 데이터를 정제하여야 한다. 더불어 편향성 방지를 위한 보안활동으로 데이터 분포 검증, AI 모델의 판단 결과에 영향을 미치지 않아야 하는 변수(보호 변수)의 설정, 인적·물적 요인에 따른 편향성 제거 등을 수행해야 한다.

현재 우리 군은 보안 규정에 따른 군사(비밀)자료의 AI 활용 제한, 부대별 데이터 분산 저장에 따른 융합·공유 및 데이터 품질 관리 미흡 등 원활한 학습 데이터 확보에 다소 어려움을 겪고 있는 실정이다[32]. 이에 국방부는 AI 개발 및 활용의 핵심이 데이터의 활용 및 관리에 있음을 인식[33]함에 따라 국방 데이터 관리 및 활용 활성화 훈령을 제정하는 한편 국방데이터관리위원회 개최를 통해 군사 자료의 체계적인 활용 및 보안 관리를 도모하고 있다. 이에 덧붙여 향후 본격적인 AI 기반 국방정보시스템의 도입 및 운용을 위해서는 군사(비밀)자료를 포함한 국방 데이터의 품질 관리 및 평가 방안을 체계적으로 정립할 필요가 있다.

4.3 모델링 및 평가 단계

모델링 시에는 앞서 3장에서 제시한 내용을 기반으로 모델에 대한 적대적 공격을 예방하기 위해 ① 다양한 모델 보안 위협 발생 가능성의 점검 및 조치, ② 국방 AI 모델의 설명 가능성 확보, ③ 국방 AI 특성을 반영한 시험평가 등의 보안 활동을 수행하여야 한다.

모델 보안 위협에 대한 방어 기법으로는 반복적 질의를 통한 모델 공격을 방어하기 위한 질의 횟수 제한[34], 모델 구축 시 적대적 이미지를 추가 학습시켜 AI 모델에 입력되는 적대적 이미지에 대한 분류 능력을 갖게 하는 적대적 학습 기법[35], 예측 결과가 결정 경계에 가까운 경우 정확도를 임의로 낮추거나 노이즈 데이터를 삽입하여 모델 세부 속성 추출을 방해하는 예측 결과 난독화[36] 등이 있다. 모델 보안 위협 대응을 위해서는 우선 AI 체계 도입 시점을 기준으로 예방 및 대응 가능한 모델 보안 위

협을 점검 및 조치하고, 이후 운영 및 모니터링 단계에서 최신화된 모델 공격 및 방어 기법들을 분석하여 체계 보완 또는 모델 재학습을 수행해야 한다.

다음으로 설명가능한 인공지능(eXplainable Artificial Intelligence, XAI)기술의 국방정보시스템 적용이 필요하다. AI의 설명가능성이란 AI의 블랙박스(Black-box) 특성을 극복하고 연산 과정 및 판단 결과에 대해 사람이 그 근거와 과정을 이해할 수 있는 성질을 의미하는 것으로[37], 기술적 관점에서는 AI 체계의 오류와 보안 취약점 수정에 활용할 수 있으며 더불어 정책적 관점에서는 국방 분야에서 AI 사용에 대한 불안감을 해소하고 기술 수용성을 강화하여 AI 도입 활성화에 기여할 수 있다. 예를 들어 AI 기반 유무인 복합체계 도입이 확산 시, 각종 전장 데이터를 종합 및 분석하여 지휘관의 지휘결심을 지원하는 AI가 군에 도입될 수 있다. 이 때 만약 AI가 내린 판단의 이유를 제시하지 못한다면 해당 시스템의 전력화는 제한될 것이다[38].

마지막으로 완성된 AI 모델에 대한 시험평가 시에는 데이터 편향성이 심한 국방 AI의 특성을 반영할 필요가 있다[39]. 예를 들어 기지경계용 CCTV 구축에 필요한 사진·영상 데이터의 경우 실제로 침입자가 등장한 화면에 비해 그렇지 않은 평시의 화면 비중이 압도적으로 높다. 마찬가지로 사이버보안 분야에서 AI 기반의 사이버방호체계 구축을 위한 학습 데이터 수집 시에도 실제 침해시도 데이터에 비해 정상 또는 오탐 데이터가 대부분일 가능성이 높다. 이 경우 특정 경우의 수가 지나치게 높은 경우를 배제하고 성능평가가 가능한 F_{β}-Score 등의 성능평가 지표를 활용할 필요가 있다.

우리 군은 상기한 AI 모델 취약점 점검 및 시험평가를 국방 소프트웨어 취약점·보안성 점검의 일환으로 간주하고 해당 절차를 국방사이버안보훈령 및 무기체계 소프트웨어 개발 및 관리 매뉴얼 등에 반영할 필요가 있다. 또한 시스템의 최초 발주 시 체계요구규격서 등에도 AI 모델의 특성을 고려하여 적절한 성능평가 방안을 명시해야 한다.

4.4 운영 및 모니터링 단계

AI 개발 생명주기의 마지막 운영 및 모니터링 단계에서는 ① 특히 탐색 공격 등을 노린 악의적 데이터 입력을 탐지하기 위한 지속적인 동작 모니터링 및 분석, ② 모델의 재학습에 대비한 절차 수립 및 이력관리, ③ 우발상황 대비를 포함한 운영 절차 수립 및 문서화 등의 보안 활동을 수행해야 한다.

구체적으로는 우선 AI 모델의 동작(입력·출력)과 체계 운영 과정 및 결과에 대한 로그를 수집하고 이를 통해 탐색 공격 등 모델 및 학습 데이터 유출을 목적으로 의도적인 입력이 지속 발생하는지 여부를 점검해야 한다. 이를 위해 훈령에 따른 시스템 운영지침서 수립 시 AI 시스템 입력 횟수 및 주기의 임계치를 설정해야 한다. 또한 AI 시스템이 유지보수 등을 이유로 군 외부로 반출되는 상황을 대비하여 형상 관리를 시행하고 앞서 설계 단계에서 수립한 접근통제 정책을 적용한다.

다음으로 AI 성능 평가, 오작동 여부 확인 등을 지속 수행하고[31] 이 과정에서 모델 성능이 기준치 이하로 하락할 경우 모델 재학습을 수행해야 하며 이에 필요한 성능 기준치, 재학습 시기, 수행 절차 등을 규정화하여야 한다. 추가적으로 시스템 운영지침서에 AI의 운용 목적과 범위, 제한사항과 더불어 동작 구조, 절차, 취급 방법 등의 세부사항을 가능한 상세히 포함해야 한다[15]. 특히 AI 시스템이 의도치 않은 오작동 또는 기능 정지로 인해 타 체계 또는 임무 수행에 영향을 끼칠 경우 즉시 체계를 정지시킬 수 있어야 하므로[6], 이에 대한 비상계획 또한 운영지침서에 포함해야 한다.

마지막으로 모델의 개발 과정 및 결과를 규격화된 양식으로 문서화 하여 AI 이해관계자 또는 보안 활동 수행 인원에게 투명하게 공유할 수 있는 제도 및 체계를 구축해야 한다[19]. AI 기반 국방정보시스템의 보안 활동은 어느 시점에만 국한된 것이 아니라 최초 설계부터 모니터링까지 全 수명주기에 걸쳐 수행되므로, 시스템 정보 및 보안 진단 결과에 대한 문서화와 공유가 상시 가능함을 항상 전제해야 한다.

V. 결 론

국방부는 국방혁신 4.0 계획을 바탕으로 국방 全 분야에 AI를 적용하기 위한 노력을 지속하고 있으며, 향후 더욱 다양한 AI 기반 정보시스템을 전력화할 예정이다. 그러나 AI의 기술적 특징으로 인해 다양한 취약점과 보안 위협이 발생할 수 있어 이를 해소하기 위한 체계적이고 심도 깊은 정책적·기술적 노력이 필요하다. 우리 軍이 체계적인 AI 보안 정책을 마련하고 이를 통해 신뢰성과 기술 수용성을 확보해야, AI를 통한 전력 증강이라

는 목표를 달성할 수 있을 것이다.

이에 본 논문에서는 AI 기반 국방정보시스템의 보안 내재화 개발을 위해 국방 AI 개발공정 및 우리 군에 실질적인 영향을 끼칠 수 있는 보안 위협 시나리오를 도출하였으며 이를 바탕으로 AI 개발 단계별 고려 및 수행해야 하는 보안 활동을 제시하였다. 또한 각 보안 활동을 국방 SW 표준 개발 공정 단계별로 분류함으로서 자연스러운 보안 내재화 개발을 도모하고 관련 제도의 보완 및 발전에도 활용할 수 있도록 하였다. 이를 통해 AI 특징적인 보안 위협 및 취약성을 제거함으로써 안전하고 신뢰할 수 있는 국방 AI의 개발, 구현 및 활용이 가능할 것으로 기대된다.

다만 제시한 보안 활동을 구체적으로 국방 업무 절차에 적용하기 위해서는 데이터 및 모델에 대한 정성·정량적 보안 평가지표와 방법의 개발, 국방 사이버보안 위험관리제도[40]등 변화하는 국방 보안업무에 기반한 AI 시스템 맞춤형 위험관리 전략과 거버넌스 구축 방안 마련 등의 추가적인 연구가 필요하다.

또한 AI 보안 위협 및 공격 기술이 지속 발전하고 다양화되는 반면 이를 완벽하게 방어하기 위한 보안 기술은 다소 제한되는 실정이다[9]. 이에 향후에도 軍이 민간·학계 및 선진국 등에서 정립된 AI 보안 정책 및 기술에 대해 지속 관심을 경주하고 이를 국방 분야에 맞게 체계화하려는 정책적 노력이 필요하다.

References

[1] McKinsey Global Institute, "Notes from the AI frontier: Modeling the impact of AI on the world economy," https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-applications-and-value-of-deep-learning, Accessed, Feb. 2023

[2] Ministry of National Defense, "The Basic Plan of Defense Innovation 4.0," Mar. 2023

[3] OECD, "The OECD Artificial Intelligence(AI) Principles," May. 2019

[4] Europian Commision High-Level Expert Group on Artificial Intelligence, "Ethical Guidelines for Trustworthy AI," Apr. 2019

[5] Ministry of Science and Technology, "Strategy to Realize Artificial Intelligence Trustworthy for Everyone," May. 2019

[6] U.S. Department of Defense, "DOD Adopts Ethical Principles for Artificial Intelligence" Feb. 2020

[7] Sohee Park, Daeseon Choi, "Artificial Intelligence Security Issue," Journal of the Korea Institute of Information Security & Cryptology, 27(3), pp. 27-32, Jun. 2017

[8] Seulgi Lee, et al, "Research Trend in Attacks of Machine Learning Model," Journal of the Korea Institute of Information Security & Cryptology, 29(6), pp. 67-74, Dec. 2019

[9] Gwonsang Ryu, Daeseon Choi, "Reseach Trend in Artificial Intelligence Security Attack and Defense," Journal of the Korea Institute of Information Security & Cryptology, 30(5), pp. 93-99, Oct. 2020

[10] Samuli Laato, et al, "AI Governance in the System Development Life Cycle: Insights on Responsible Machine Learning Engineering," IEEE/ACM 1st International Conference on AI Engineering - Software Engineering for AI(CAIN), pp. 113-123, Jun. 2022

[11] Erick Galinkin, "Towards a Responsible AI Development Lifecycle: Lessons from Information Security", arXiv preprint arXiv:2203.02958, Mar. 2022

[12] Yoon Junghyun, "Issues and Prospects of AI Utilization in the Defense Field," STEPI Insight, pp.

1-55, Aug. 2021
〔13〕 Keun Ha Choi, et al, "The Implications to ROK Armed Forces from the Artificial Intelligence Strategy of U.S. Department of Defense and Army," Journal of the Korea Association of Defense Industry Studies, 27(1), pp. 41-52, Jun. 2020
〔14〕 Ministry of National Defense, "Instruction on Defense Informationization," May. 2022
〔15〕 Ministry of Science and Technology, "Framework Act on Intelligent Informatization," Jun. 2020
〔16〕 Kwang-Chun Go, "Improvement of the Acquisition Procedure for Defense Information System," KIDA Defense Issues & Analyses, 1916(22-37), Oct. 2022
〔17〕 Sung Rim Cho, et al, "Development of the Information Security Methodology for Defense Organization," Korea Society of IT Services, 12(4), pp. 77-90, Dec. 2013
〔18〕 ISO, "ISO/IEC 23053: Framework for Artificial Intelligence(AI) Systems Using Machine Learning(ML)," Jun. 2022
〔19〕 NIST, "AI Risk Management Framework," Jan. 2023
〔20〕 ISO, "ISO/IEC TR 24028: Overview of Trustworthiness in Artificial Intelligence," May. 2020
〔21〕 "Poison attacks against machine learning, Security and spam-detection programs could be affected," The Kurzweil Accelerating Intelligence, July. 2012
〔22〕 Y. Yao, et al, "Latent Backdoor Attacks on Deep Neural Networks," Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security, pp. 2041-2055, Nov. 2019
〔23〕 D. Song, "AI and Security: Lessons, Challenges, and Future Directions," IEEE Symposium on Security and Privacy, 1st Deep Learning and Security Workshop, May. 2018
〔24〕 Hyun Kwon, Youngchul Kim, "Trend in Technology of Adversarial Examples in Deep Learning Model," Journal of the Korea Institute of Information Security & Cryptology, 31(2), pp. 5-12, Apr. 2021
〔25〕 Hyun Kwon, et al, "Research Trend in Evasion Attacks of AI Security Problems," Communications of the Korean Institute of Information Scientists and Engineers, 36(2), pp. 32-36, Feb. 2018
〔26〕 The Pytorch Team, "Compromised PyTorch-nightly dependency chain between December 25th and December 30th, 2022", https://pytorch.org/blog/compromised-nightly-dependency/#how-to-check-if-your-python-environment-is-affected, Accessed, Feb. 2023
〔27〕 Ax Sharma, "PyTorch discloses malicious dependency chain compromise over holidays," https://www.bleepingcomputer.com/news/security/pytorch-discloses-malicious-dependency-chain-compromise-over-holidays/, Accessed, Jan. 2023
〔28〕 Chan Ho Jeong, Hyun Kwon, "A Study on Adversarial Examples for Image-Based Deep Learning Model in the Military," Proceedings of Symposium of the Korean Institute of communications and Information Sciences, pp. 1452-1453, Jun. 2021
〔29〕 W. Brendel, et al, "Decision-Based Adversarial Attacks: Reliable Attacks Against Black-Box Machine Learning Models," International Conference on

Learning Representations, Feb. 2018
〔30〕 Yoonsoo An, Daeseon Choi, "Model Type Inference Attack Using Output of Black-box AI Model," Journal of the Korea Institute of Information Security & Cryptology, 32(5), pp. 817-826, Oct. 2022
〔31〕 Ministry of Science and Technology, Telecommunications Technology Association, "Guideline for Trustworthy AI Development," Jan. 2022
〔32〕 Seyong Kim, et al, "A Study on the Strategic Application of National Defense Data for the Construction of Smart Forces in the 4th IR," Convergence Security Journal, 20(4), pp. 113-123, Oct. 2020
〔33〕 Ministry of National Defense, "Second Session of the Defense Data Management Committee Held by Ministry of National Defense," Press Release, Apr. 2023
〔34〕 Financial Security Institute, "Guideline for Financial Sector AI Security," Apr. 2023
〔35〕 Tramèr, Florian, et al. "Ensemble adversarial training: Attacks and Defenses." arXiv preprint arXiv:1705.07204, Apr. 2020
〔36〕 J. Jia, et al, "MemGuard: Defending against Black-Box Membership Inference Attacks via Adversarial Examples," Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security, pp. 259-274, Dec. 2019
〔37〕 DARPA, "Explainable Artificial Intelligence," Aug. 2016
〔38〕 Donghan Oh, "Utilization of Artificial Intelligence Technology in the Military and Suggestion of XAI Technology Application Direction," Journal of Digital Contents Society, 23(5), pp. 943-951, May. 2022
〔39〕 Yongbok Lee, et al, "Development of Artificial Intelligence Weapon System Test and Evaluation Methods: Performance Evaluation of the Classification Model," Journal of Applied Reliability, 22(1), pp. 1-9, Mar. 2022
〔40〕 Woo-sung Yang, et al, "Korean Security Risk Management Framework for the Application of Defense Acquisition System," Journal of the Korea Institute of Information Security & Cryptology, 32(6), pp. 1183-1192, Dec. 2022

〈저자소개〉

박 규 도 (Gyu-do Park) 정회원
2022년 2월: 성균관대학교 정보통신대학원 석사
2017년 8월~2019년 12월: 공군 군수전산소 정보보호실
2019년 12월~2022년 12월: 국군방첩사령부 국방보안연구소 정보보호실
2023년 1월~현재: 국군방첩사령부 국방보안연구소 RMF TAG
〈관심분야〉 정보보호, 사이버보안, 인공지능

이 영 란 (Young-ran Lee) 정회원
2005년 2월: 이화여자대학교 수학과 이학박사
2009년~2014년: 국가수리과학연구소 선임연구원
2015년 1월~2019년 11월: 국군방첩사령부 국방보안연구소 책임연구원
2019년 12월~2020년 12월: 국군방첩사령부 한국형 사이버보안제도개발 TF
2021년 1월~2022년 4월: 국군방첩사령부 국방보안연구소 책임연구원
2022년 5월~현재: 국군방첩사령부 국방보안연구소 정보보호연구실장
〈관심분야〉 사이버보안, 공개키암호, RMF, C-SCRM