Journal of KIIT. Vol. 11, No. 12, pp. 125-134, Dec. 31, 2019. pISSN 1598-8619, eISSN 2093-7571 125
http://dx.doi.org/10.14801/jkiit.2019.17.12.125

AI 사이버보안 체계를 위한 블록체인 기반의 Data-Preserving AI 학습환경 모델

김인경*, 박남제**

Blockchain Based Data-Preserving AI Learning Environment Model for Cyber Security System

Inkyung Kim*, Namje Park**

이 논문은 2019년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임[2019-0-00203,선제적 위험대응을 위한 예측적 영상보안 핵심기술 개발]. 그리고, 2019년도 정부(교육부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행된 것임(과제번호:NRF-2019R1I1A3A01062789)

요 약

인공지능 기술은 작동과정에 대한 투명성이 보장되지 않는 수동적 인식 영역에 제한되는 한계점으로 인해, AI가 학습하는 데이터에 의존적인 취약점을 갖는다. 인공지능 학습을 위한 원시데이터는 AI 학습의 고도화를 위한 데이터 품질 확보를 위해 수작업으로 가공과 검수를 해야 하기에 인적 오류가 내재되어 있으며, 데이터의 훼손, 불완전함, 원시데이터와의 차이 등으로 인해 가공데이터를 통한 AI 학습 시 예상 치 못한 결과값을 도출할 수 있다. 이에 본 연구에서는 사이버 보안 관점에서의 접근을 통한 AI 학습데이터의 부정확한 사례 및 사이버보안 공격 방법 분석을 통해 기계학습 전 학습데이터 관리의 필요성을 살펴보고, 학습 데이터 무결성 검증을 위해 블록체인 기반의 학습데이터 환경 모델인 Data-preserving 인공지능 시스템 구축 방향을 제시한다. Data-preserving AI 학습환경 모델은 AI 학습데이터 제공 전 변조되지 않은 데이터로 학습됨을 보장 하여 데이터 가공 시 및 원시데이터 수집을 위한 오픈 네트워크에서의 데이터 제공 및 활용 시 있을 수 있는 사이버 공격, 데이터 변질 등의 위협을 사전에 방지할 수 있을 것으로 기대된다.

Abstract

As the limitations of the passive recognition domain, which is not guaranteed transparency of the operation process, AI technology has a vulnerability that depends on the data. Human error is inherent because raw data for artificial intelligence learning must be processed and inspected manually to secure data quality for the advancement of AI learning. In this study, we examine the necessity of learning data management before machine learning by analyzing inaccurate cases of AI learning data and cyber security attack method through the approach from cyber security perspective. In order to verify the learning data integrity, this paper presents the direction of data-preserving artificial intelligence system, a blockchain-based learning data environment model. The proposed method is expected to prevent the threats such as cyber attack and data corruption in providing and using data in the open network for data processing and raw data collection.

Keywords
data-preserving AI, block chain, cyber security, raw data

* 제주대학교 창의교육거점센터, 사이버보안 인재교육원 책임연구원
- ORCID: https://orcid.org/0000-0003-4055-6292
** 제주대학교 초등컴퓨터교육전공, 융합정보보안학과 교수(교신저자)
- ORCID: https://orcid.org/0000-0003-4434-8933

· Received: Sep. 20, 2019, Revised: Oct. 21, 2019, Accepted: Oct. 24, 2019
· Corresponding Author: Namje Park
Dept. of Computer Education, Teachers College, Jeju National University, 61 Iljudong-ro, Jeju-si, Jeju Special Self-Governing Province, 63294, Korea
Tel.: +82-64-754-4914, Email: namjepark@jejunu.ac.kr

Ⅰ. 서 론

인공지능(AI, Artificial Intelligence)은 머신러닝, 딥러닝의 기술과 함께 산업적으로 적용 가능한 수준으로 발전하면서 빠르게 상용화되어 의료, 금융, 로봇, 문화 등 다양한 분야에서 혁신을 이끌고 있다. 구글은 인공지능을 활용한 'AI 칼리코 프로젝트'를 통해 인간 평균 수명을 연장하는 것을 목표로 유전자 데이터와 가계도를 이용해 난치병을 치료하는 연구를 진행하고 있으며, 국내에서도 'AI 닥터'를 도입하여 질병을 관리하고 진단하는 등 기업을 넘어 국가 단위의 인공지능 기술이 현실화되고 있다.

국내 과학기술정보통신부에서는 해당 응용분야에 관한 질 좋은 데이터를 충분히 확보하는 것이 결정적임을 인공지능 알고리즘 개발의 초점으로 제시하고, AI 학습용 데이터 구축 과제 10가지를 통해 다방면에서의 데이터셋 구축 계획을 발표하였다. AI 위험물 식별 능력 증진, 질병 진단 AI 기술 역량 강화, 군집의 이상 행동 탐지 등을 위한 공문, 산업, 유통, 의료, 역사, 문화 등 다방면에서의 자료 수집을 목적으로 번역, 상황·동작·인지, 사물·위험요소 식별, 질병 진단 데이터 등 복합 인지 능력을 갖춘 AI 개발을 지원하기 위한 '멀티 모달' 영상 데이터 구축을 추진하고 있다.

지금까지의 인공지능은 인지, 의사결정, 예측 등의 정보를 제공할 때 결과에 대한 충분한 근거를 제공하지 못하며 수동적 인식 영역에 제한되는 인공지능 기술의 한계 극복을 위해 설명 가능한 인공지능(Explainable AI)이 요구되고 있다. EU에서는 GDPR(General Data Protection Regulation)을 통해 설명 가능한 알고리즘 개발 요구를 증가시키고, 2017년 DARPA에서는 XAI(Explainable AI) 프로젝트를 통해 설명 가능한 인고지능 알고리즘 개발을 추진하다. 딥러닝으로 대표되는 인공신경망의 블랙박스와 같은 기능으로 인해 작동과정에 대한 투명성이 보장되지 않음에 대한 신뢰성 문제 해결을 위한 정책적, 기술적 필요성이 야기되고 있다. 특히, 의료진단, 자율주행 등 생활 속 인공지능 도입을 위해서는 인공지능 행위 결과에 대한 판단의 불명확성에 대해 알고리즘 검증강화, 정확한 데이터 사용 등 인공지능 시스템 자체 내 기술 개발 및 오류를 최소화하고, 악의적인 공격을 방어할 수 있는 구조 도입이 필요하다.

이에 본 연구에서는 인공지능 신뢰성 향상을 위해 사이버 보안 관점에서의 접근을 통한 AI 학습데이터의 부정확한 사례 및 사이버보안 공격 방법 분석을 통해 기계학습 전 학습데이터 관리의 필요성을 살펴보고, 학습 데이터 무결성 검증을 위해 블록체인 기반의 학습데이터 환경 모델인 Data-preserving 인공지능 시스템 구축 방향을 제시하고자 한다.

Ⅱ. 관련 연구

2.1 AI 사이버위협

본 절에서는 AI 학습데이터의 위·변조 및 부정확성 사례를 통해 AI의 사이버보안 필요성을 제시한다. 먼저 AI 학습데이터로 인한 사고는 다음과 같다[1].

- AI 챗봇 악의적 학습 유도 : 2016년 마이크로소프트(MS)에서 인공지능 채팅로봇 Tay를 공개했으나 사용자의 의도적인 인종·성차별 메시지 학습으로 16시간만에 종료시켰다.

- 적대적 스티커 기반 공격 : 구글 리서치 그룹은 이미지 인식 인공지능 알고리즘을 오작동 시킬 수 있는 '적대적 스티커(Adversarial patch)'를 발표하였다. 적대적 스티커는 원형의 추상적 이미지를 담은 스티커로, 단순히 인쇄해서 사물 옆에만 붙여두면 이미지를 인식하는 인공지능 알고리즘의 오작동을 야기하였다.

- 미국 플로리다주 브로워드 카운티의 범죄 재발 가능성 예측 오류 : 약 18,000명의 범죄자를 대상으로 향후 2년동안 새로운 범죄를 일으킬 가능성에 대해 인공지능을 활용한 범죄자 예측 알고리즘을 통해 분석한 결과 흑인이 백인보다 범죄 재발 가능성이 약 54% 높은 것으로 파악되었으나, 동일기간 실제 데이터를 분석한 결과 오히려 백인의 재범 비율이 높게 나타났다.

- 딥러닝을 통한 의료 기록 위조 : 이스라엘 벤구리온대학교의 인공지능 연구진은 딥러닝 기술을 통해 만든 악성코드를 심어 환자의 3D 스캔 영상을

조작하는 실험을 진행하여 3명의 의사를 대상으로 전원을 속였으며, 실제 병원에서 사용되는 3D CT영상이나 엑스레이, MRI 등의 병원데이터가 보안 장치 없이 유통되어 적용되는 문제점이 제기되었다.

- AI를 활용한 딥페이크 : 2017년말 스와핑 알고리즘을 적용해 얼굴을 바꾸는 딥페이크를 통해 미국의 레딧(Reddit)에 가짜 동영상을 확산시켜 영상의 진위를 가리기 어려웠으며, AI 기술 플랫폼에서 AI 학습데이터의 유출에 대한 대응의 필요성이 제시되었다.

상기 사례를 통해 수많은 데이터를 통해 학습을 하는 인공지능은 부적절하거나 불안전한 데이터에 대한 취약성이 있으며, 악의적인 데이터 적용 시, 인공지능의 기능 및 성능과는 무관하게 잘못된 결과값을 도출할 수 있다. AI의 효율성 및 보급성을 기반으로, 공격자가 데이터를 통하여 AI를 부정적으로 사용하거나 AI 시스템을 교란시키는 공격을 수행하는 등의 보안 위협이 존재한다. 즉, 사이버공격으로 인한 데이터 위·변조 발생 시, AI의 학습효과, 학습방향 등을 훼손시킬 수 있기에 AI 학습데이터 제공 전 데이터 관리가 필요하다.

2.2 관련 연구 동향

인공지능의 학습데이터 중요성과 함께 학습데이터 수집 및 가공 서비스가 증가하고 있으며, 최근 블록체인과 AI의 융합을 위한 연구가 진행되고 있다. 본 절에서는 연구에서 제안하고자 하는 AI 학습데이터 환경모델 기반기술인 블록체인과 AI 데이터와의 연계를 위한 관련 연구 사례를 소개한다.

Kim(2019)는 블록체인 구조에서 블록들이 직접 병렬적으로 데이터를 수집하게 하고 각 블록들이 수집한 데이터를 타 블록의 데이터와 비교하여 양질의 데이터만을 선별하는 방안을 제안하고, 비교를 통해 분석된 데이터를 통한 학습용 데이터셋을 구성하였다[2].

Aum(2019)은 블록체인을 이용한 레이블 타입 데이터 관리 기반의 AI 학습 데이터 생산성 향상 시스템 및 방법을 통해 해킹에 대비하여 안정적 데이터 저장 체계를 확립할 수 있도록 하였고[3], 블록체인 노드의 스토리지 자원이 보다 효율적으로 활용 될 수 있는 블록체인 기반 데이터 관리 방법 및 그 장치를 제시하였다[4].

Asaph(2016)의 연구사례는 의료 연구를 목적으로 블록체인을 기반으로 하는 데이터 공유 메커니즘을 제안하였다[5].

III. 제안 방안

3.1 AI 학습데이터 특성

딥러닝을 포함한 다양한 종류의 인공지능 기술은 분석 및 학습 알고리즘, 컴퓨팅 시스템, 데이터로 구성되며 데이터 학습 여부가 알고리즘 고도화에 직결된다. 인공지능 학습을 위해서는 데이터 예시가 필요하며, 딥러닝을 활용하여 특정 기능의 AI 모델을 구현하기 위해서는 목적에 맞는 데이터셋을 구성하여 학습을 해야 한다. AI 머신러닝 학습 흐름은 아래 그림과 같으며, AI 모델에 적용하기 위한 데이터는 일반적으로 다음과 같은 프로세스를 따른다.

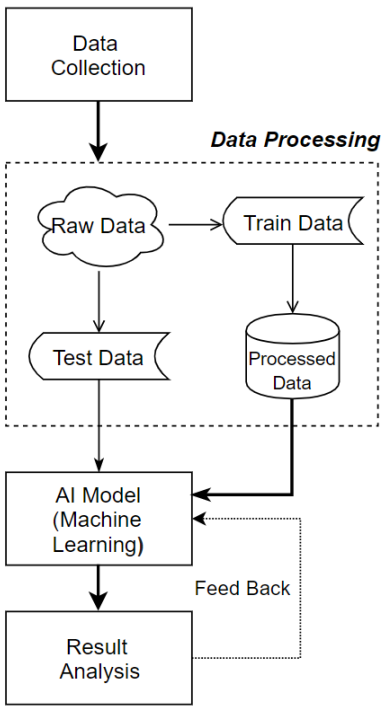

그림 1. AI 머신러닝 학습 흐름
Fig. 1. AI machine learning flow

- 데이터 수집 : 비정형 데이터의 수집단계로 이미지, 텍스트, 음성 등의 데이터 중 개발하려는 인공지능의 목적 및 기능에 맞는 데이터를 추출하는 1차 전처리 수행

- 데이터 전처리 : 수집된 데이터를 머신러닝 모델의 입력에 적합한 형태로 바꿔주는 단계로, 데이터에 빠진 부분(결측값)을 채우거나 삭제, 중복 또는 필요한 데이터 속성을 선택 혹은 삭제 또는 데이터의 기존 속성을 조합하거나 모델의 필요에 따라 원시데이터를 지정된 형식으로 변환하는 과정을 포함

- 데이터 분석 : 정형화된 데이터의 패턴을 탐색, 데이터 매핑, 탐색 및 추론에 기반한 데이터 추출, 일부 데이터를 통한 데이터의 학습 등 인공지능에 적용하기 위한 데이터 분석 과정

데이터 수집 과정에서의 원시 데이터(Raw data)는 인공지능 알고리즘에 적용하기에는 노이즈가 많고 일관성이 없으며 중복된 경우가 많다. 인공지능 모델의 품질이나 신뢰성, 정확성, 성능을 위해서는 고품질의 데이터를 위한 단계가 필수적이며, 본 단계에서 데이터 오류 수정, 중복제거, 불일치 데이터 제거, 충돌 데이터 조정 등 사람의 전문성 및 통찰력을 기반으로 하는 데이터 분석과 조직화가 수행된다. 이러한 데이터의 전처리 작업에 평균 80% 이상 노력이 필요하다고 알려져 있다[6]-[9]. 질 좋은 데이터를 충분히 확보하는 것이 결정적인 인공지능 개발의 초점이 되고, 인공지능이 학습하기 위한 데이터에 대한 품질 보장이 요구된다.

3.2 AI 학습데이터 요건

현재 인공지능 기술은 결과에 대한 충분한 근거를 제공하지 못하며 작동과정에 대한 투명성이 보장되지 않는 수동적 인식 영역에 제한되는 한계점으로 인해, AI가 학습하는 데이터에 의존적인 취약점을 갖는다. 또한 인공지능의 특성 중 유익한 용도 또는 유해한 용도 모두 사용이 가능한 이중성의 특성과 블랙박스 구조적 특성으로 인한 취약성으로 인해 AI 보안 위협 등장이 우려된다.

표 1. AI에 적용 가능한 공격기법
Table 1. Attack techniques applicable to AI

Attack method	Description
Poisoning attack	A type of causal attack that injects a hostile sample into a training data set, disrupting the availability and integrity of a machine learning model.
Evasion attack	Attacks that significantly reduce the overall security of the target system by generating some hostile samples that an attacker can evade detection
Impersonate attack	Generate specific hostile samples so that existing machine-learning-based systems use a different label than the impersonated sample to misclassify the original sample.
Inversion attack	It is an attack that collects some basic information about the target system model by using API provided by the machine learning system, and reversely analyzes the basic information to steal sensitive information.

표 1은 악의적인 데이터에 취약성을 이용한 AI에서 적용 가능한 기술적 공격기법으로, AI가 학습하기 위한 데이터의 무결성이 요구된다[10]-[15].

특히, 원시데이터는 AI 학습의 고도화를 위한 데이터 품질 확보를 위해 수작업으로 가공과 검수를 해야 하기에 인적 오류가 내재되어 있다. 데이터의 훼손, 불완전함, 원시데이터와의 차이 등으로 인해 가공데이터를 통한 AI 학습 시 예상 치 못한 결과값을 도출할 수 있다. 즉, AI 모델 구축 과정에서 AI의 상이한 결과값에 대한 검증을 위해 가공된 학습데이터 추적이 필요하며, 이 때 원시데이터에 대한 무결성이 보장되어야 한다[16]-[19].

또한 AI 기술 적용을 위해 다양한 데이터 수집 및 가공 서비스가 제공되고 있으나 정보의 신뢰성 이슈가 부각되고 있다. 악의적인 목적을 위해 신뢰할 수 없는 다량의 데이터 업로드에 대한 출처를 분석하고 추적 가능한 학습데이터 수집 환경이 필요하며, 일반적으로 오픈 데이터를 관리해주는 중앙에서의 서버가 존재한다. 이에 중개 서버가 중단되거나 변조가 가해지면 정보의 신뢰성, 가용성이 침해되는 등 여러 보안 위협을 내재하고 있다. 또한 AI가 학습하기 위한 데이터는 수집 과정에서 빅데이터 형태를 가지기에 빅데이터의 서버에 대한 안

전성이 요구되어 진다[20]-[24].

이에 본 연구에서는 데이터 위·변조 방지 및 무결성 보장을 위해 블록체인 기반의 학습데이터 관리 방안을 제안한다.

3.3 AI 학습환경 모델

인공지능 학습모델은 원시데이터를 그대로 사용하여 학습과정을 수행하거나, 원시데이터를 학습에 적합한 형태로 변형하여 생성한 가공 데이터를 이용한 학습 데이터 수행 시 제 3자의 악의적인 공격으로 학습과정에서 사용되는 데이터의 위·변조를 방지해야 한다. 특히, 오픈 네트워크를 통해 복수의 데이터 제공자로부터 다양한 원시데이터를 수집하여 학습데이터로 이용 시, 원시데이터 암호화를 통한 개인정보에 대한 프라이버스 보호가 필요하다[25].

본 연구에서 제안하는 블록체인 구조화를 통한 Data-preserving AI를 위한 학습환경 모델은 블록체인 구조의 특성을 통해 AI가 학습하기 위한 데이터에서 원본 및 가공된 데이터의 무결성의 요건을 만족시키며, AI 학습데이터 제공 전 변조되지 않은 데이터로 학습됨을 보장 할 수 있다[26][27].

그림 2는 본 연구에서 제안하는 구조로, 적어도 하나 이상의 데이터 제공자로부터 수신한 원시데이터 및 상기 원시데이터의 해시코드를 블록체인에 저장하고, 인공지능 학습모델에 블록체인에 저장된 원시데이터를 제공, 인공지능 학습모델에서 사용한 데이터의 해시코드와 블록체인에 저장된 원시데이터의 해시코드 비교를 통한 데이터의 무결성 검증의 구조를 포함한다. 인공지능 학습을 위한 원시데이터는 적어도 하나 이상의 데이터 제공자로부터 수신이 가능한 오픈 네트워크 환경에서 원시데이터의 위변조를 원천적으로 방지하고, 데이터 제공자를 정확하게 추적할 수 있다.

원시데이터를 저장하는 블록체인은 블록헤더와 블록데이터 구성된 블록에서 데이터 제공자로부터 수신한 원시데이터의 블록데이터에 해시코드를 포함하여 암호화 저장하거나, 인공지능 모델의 특성에 따라 원시데이터의 가공 시의 편리성 추구를 위해 원시데이터의 해시코드를 저장하는 블록데이터를 따로 구성할 수 있다.

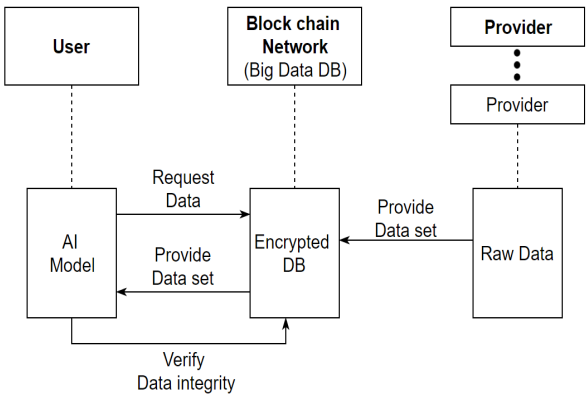

그림 2. 블록체인 기반 AI 학습데이터 오픈 네트워크 구조

Fig. 2. Blockchain-based AI learning data open network

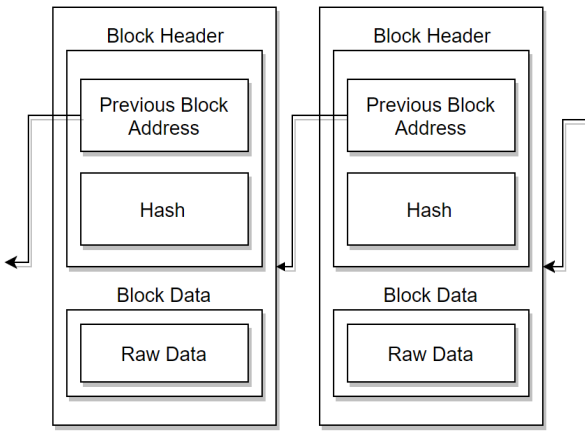

그림 3. 제안한 블록 구조

Fig. 3. Proposed block configuration

본 연구에서의 블록체인 적용방안은 불가변성의 특성을 통해 AI가 학습하기 위한 데이터는 위·변조되지 않았음을 통해 무결성을 보장할 수 있다. 또한 악의적 공격, DDoS 등 특정서버의 무력화로부터 안전성을 제공하며, 내부자의 조작을 방지할 수 있다. 또한 학습데이터의 블록 암호화로 인해 데이터와 관련한 유출로부터 자유로우며, AI 결과값에 대한 데이터 추적이 가능하다.

3.4 AI 검증환경 모델

AI 학습환경에서는 원시데이터 및 가공데이터의 추적을 통한 검증 가능한 데이터 감사가 필요하다. 그림 3은 인공지능 학습데이터의 무결성을 검증하기 위한 구조로, 데이터 제공 노드에서 검증노드를 통

한 AI 서버와의 유무선 통신망을 통해 연결된 시스템에서 검증노드는 블록체인 서버와의 연결을 통해 데이터의 검증을 수행한다. 검증노드는 데이터 제공 노드로부터 데이터를 수신하면 이를 블록체인에 저장하고, AI 서버로부터 데이터 요청 시 블록체인에 저장된 데이터를 AI 서버 서버로 제공하는 과정을 통해 AI 학습에서의 검증환경을 구축할 수 있다.

검증환경 모델에서의 데이터 제공자 층은 원시데이터 및 원시데이터의 해시코드를 같이 데이터 층으로 전송한다. 데이터 층은 데이터 관리 및 수집을 위한 층으로, 제공자 층에서 전송받은 데이터 및 해시코드를 암호화하고, 블록체인에 저장한다. AI 머신러닝을 위한 층에서는 데이터 층으로부터 데이터 및 암호화된 해시코드를 제공받아 AI를 학습시키고, 학습된 패턴을 통한 향후 결과를 도출하여 모델을 관리한다.

검증노드에서는 데이터 제공 노드에서 수신한 데이터를 블록체인에 암호화키 및 복호화키를 포함하여 암호화하여 저장한다. 또는 검증노드 내에서 암호화키 및 복호화키 생성 후, 데이터 제공 노드로 수신하여 데이터 제공자가 데이터 제공 시 부여받은 암호화키를 이용하여 데이터 암호화 후, 데이터 제공이 가능하다. 이는 오픈 네트워크에서 다수의 데이터 제공자에게 동일한 암호화키를 이용하여 데이터를 암호화하므로, 하나의 복호화키만을 관리하는 실효성을 갖는다.

표 2. 검증노드 구조
Table 2. Construct of verification

Scope	Description
Storage	Save data hashcode with data storage. Store hashcodes together in one blockchain or save each one separately
Provider	When requesting data from AI server, it transmits to data server stored in blockchain. Providing after decrypting encrypted data in data encryption mode
Verification	Integrity verification using hashcode of data stored in blockchain. Compare hashcode of data received from AI server to hashcode of data stored in blockchain

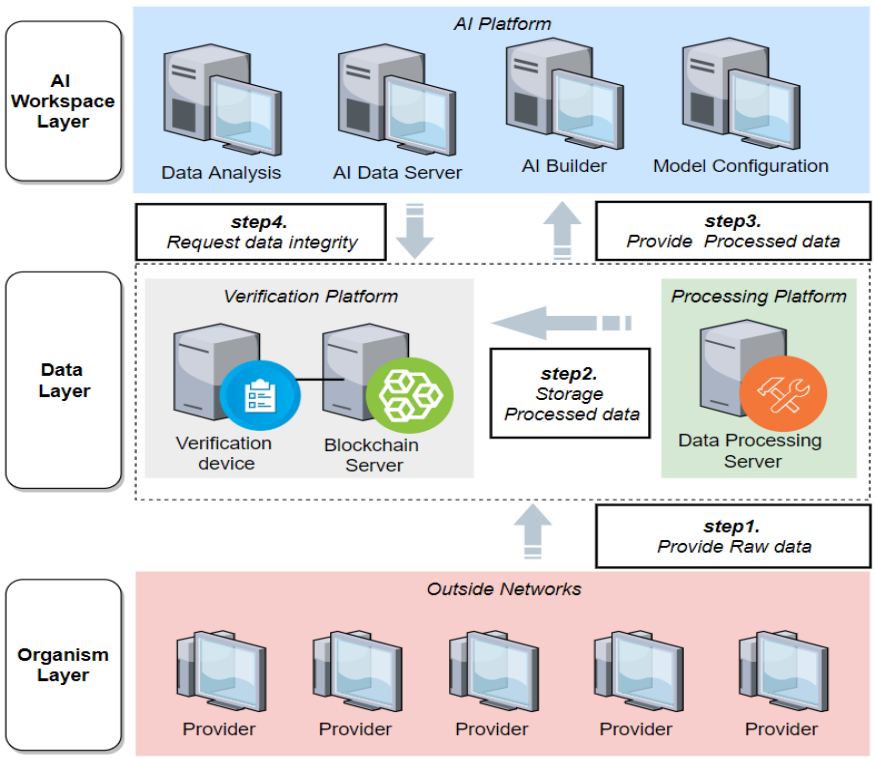

그림 4. AI 검증환경 구조
Fig. 4. AI environment structure for verification

데이터 무결성 검증을 위해서는 AI 서버로부터 받은 데이터 및 데이터의 해시코드를 포함하는 무결성 검증 요청 시, 데이터 저장 모듈에서의 블록체인 원장으로부터 해당되는 데이터의 암호화된 해시코드를 제공모듈로 전송한다. 제공모듈에서는 암호화된 해시값을 복호화하여 검증모듈로 전송하고, 검증 모듈에서는 복호화된 해시코드와 AI 서버로부터 수신된 데이터의 해시코드를 비교하여 위·변조 여부를 검증한다.

3.5 기존 연구와의 비교분석

인공지능 학습데이터를 위한 연구로는 인공지능 모델 개발에 필요한 데이터를 수집하고, 가공해주는 서비스를 위한 양질의 학습용 데이터셋 제공, 가공·검수 프로세스와 인공지능 기반 자동화 도구 개발에 초점이 맞추어져있다. 또한 AI 모델용으로 쓸 개인 데이터를 공개하도록 유도하는 탈중앙화된 데이터 교환 프로토콜 및 네트워크로, 데이터를 올린 사람은 암호화폐로 보상하기 위한 시스템 및 대화형 인터페이스 개발자와 데이터를 소유한 사람이 네트워크에 가치를 제공하고 보상을 받을 수 있는 프레임워크 등 AI 학습데이터 수집을 통한 보상 방안을 위한 연구가 활발하다.

본 연구에서의 AI 학습환경 모델은 AI 모델 기계학습 시 기반이 되는 학습데이터를 블록체인 구조를 통한 수집, 저장을 통해 데이터 무결성, 암호화를 통한 기밀성을 보장할 수 있다. 이를 통해 AI 학습데이터에 대한 추적 및 검증이 가능하여 AI 모델의 신뢰성을 확보할 수 있으며, 데이터 가공 및 원시데이터 수집을 위한 오픈 네트워크에서의 데이터 제공 및 활용 시 사이버 공격, 데이터 변질 등의 위협을 사전에 방지할 수 있는 장점을 가진다. 즉, 본 연구에서는 AI 환경에서 보안의 필요성을 강조하고, AI 학습의 고도화를 위한 신뢰성 확보에 초점을 두어 학습데이터의 무결성을 보장하기 위한 환경 모델을 제안한 것으로 기존 연구와의 목적에서 차별성을 갖는다.

표 3. 기존 연구와의 비교분석
Table 3. Comparative analysis with existing research

	Conventional AI learning environment model	Proposed AI learning environment model
Research element	Raw data for AI learning	
Purpose	- Collect and process data for AI models - Compensation through decentralized data collection	- Ensure AI learning data integrity - Preparing for AI environmental infringement accidents from cyber attacks
Characteristic	- Providing and processing high quality datasets - Interface environment with data providers and AI model developers	- Verifiable data audit through AI training data tracking - Traceable when unintended AI results due to processed data

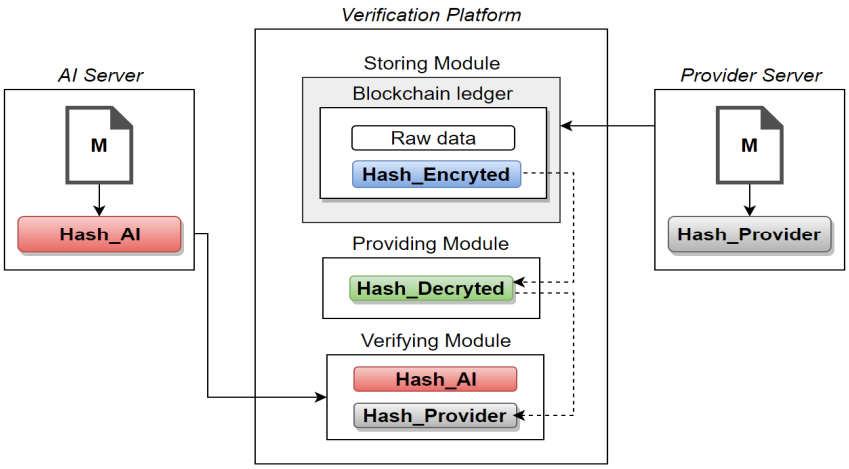

그림 5. AI 데이터 검증 절차
Fig. 5. Process of verification for AI data

Ⅳ. 결 론

인공지능은 행위 결과에 대한 판단의 불명확성에 대해 알고리즘 검증강화, 정확한 데이터 사용 등 인공지능 시스템 자체 내 기술 개발 및 오류를 최소화하고, 악의적인 공격을 방어할 수 있는 구조 도입이 필요하다. 이에 본 연구에서는 인공지능 신뢰성 향상을 위해 사이버 보안 관점에서의 접근을 통한 AI 학습데이터의 부정확한 사례 및 사이버보안 공격 방법 분석을 통해 기계학습 전 학습데이터 관리의 필요성을 살펴보고, 학습 데이터 무결성 검증을 위해 블록체인 기반의 학습데이터 환경 구축 모델인 Data-preserving 인공지능 시스템 구조를 제시하였다. 본 연구에서의 제안방안은 AI 모델의 기계학습 시 기반이 되는 학습데이터의 블록체인 구조를 통합 수집, 저장을 통해 데이터 무결성, 암호화를 통한 데이터 기밀성을 보장할 수 있다. 또한 AI 학습데이터에 대한 추적 및 검증이 가능하여 AI 모델의 신뢰성을 확보할 수 있다. 데이터 가공 시 및 원시데이터 수집을 위한 오픈 네트워크에서의 데이터 제공 및 활용 시 있을 수 있는 사이버 공격, 데이터 변질 등의 위협을 사전에 방지할 수 있을 것으로 기대된다.

References

[1] Yongsik Moon, "The Malicious Use of Artificial Intelligence Forecasting, Prevention, and Mitigation", National Information Society Agency(NIA), NIA Special Report, 2018-12, Aug. 2018

[2] Youngrang Kim, Junghoon Woo, Jaehwan Lee, and Ji Sun Shin, "High-quality data collection for machine learning using block chain", Journal of the Korea Institute of Information and Communication Engineering, Vol. 23, No. 1, pp. 13-19, Jan. 2019.

[3] Sungmin Aum, "Artificial Intelligence Learning Data Productivity Improvement System based on Label Type Data Management Using Block Chain, and Method thereof", KR Patent 1020180153330, filed Nov 3, 2018, issued May 10, 2019.

[4] Sungmin Aum, "Automatic inspection system for label type data based on Artificial Intelligence Learning to improve data productivity, and method thereof", KR Patent 1020180153327, filed Nov 3, 2018, issued Apr 5, 2019.

[5] A. Azaria, A. Ekblaw, T. Vieira, and A. Lippman, "MedRec: Using Blockchain for Medical Data Access and Permission Management", Conference of Open and Big Data, Vienna, Austria, pp. 25-30, Aug. 2016.

[6] Yeongchun Woo, Seongteob Lee, Wan Choi, Chanwon Ahn, and Okki Baek, "Trend of Utilization of Machine Learning Technology for Digital Healthcare Data Analysis", Electronics and Telecommunications Trends, Vol. 34, No. 1, pp. 98-110, Feb. 2019.

[7] Namje Park, Byung-Gyu Kim, and Jinsu Kim, "A Mechanism of Masking Identification Information regarding Moving Objects Recorded on Visual Surveillance Systems by Differentially Implementing Access Permission", ELECTRONICS, Vol. 8, No. 7, pp. 735, Jul. 2019.

[8] Jinsu Kim, Namje Park, Geonwoo Kim, and Seunghun Jin, "CCTV Video Processing Metadata Security Scheme Using Character Order Preserving -Transformation in the Emerging Multimedia", ELECTRONICS, Vol. 8, No. 4, pp. 412, Apr. 2019.

[9] Namje Park, Younghoon Sung, Youngsik Jeong, Soo-Bum Shin, and Chul Kim, "The Analysis of the Appropriateness of Information Education Curriculum Standard Model for Elementary School in Korea", International Conference on Computer and Information Science, Springer, pp. 1-15, Jun. 2018.

[10] Donghyeok Lee, Namje Park, Geonwoo Kim, and Seunghun Jin, “De-identification of metering data for smart grid personal security in intelligent CCTV-based P2P cloud computing environment", Journal of Peer-to-Peer Networking and Applications, Vol. 11, No. 6, pp. 1299-1308, Nov. 2018.

[11] Donghyeok Lee and Namje Park, "Electronic

identity information hiding methods using a secret sharing scheme in multimedia-centric internet of things environment", Journal of Personal And Ubiquitous Computing, Vol. 22, No. 1, pp. 3-10, Feb. 2018.

[12] Donghyeok Lee and Namje Park, "Geocasting-based synchronization of Almanac on the maritime cloud for distributed smart surveillance", Supercomputing, Vol. 73, No. 3, pp. 1103-1118, Mar. 2017.

[13] Namje Park and Hyochan Bang, "Mobile middleware platform for secure vessel traffic system in IoT service environment", Journal of Security And Communication Networks, pp. 500-512, Nov. 2014.

[14] Namje Park and Namhi Kang, "Mutual Authentication Scheme in Secure Internet of Things Technology for Comfortable Lifestyle", Journal of Sensors (Basel), Vol. 16, No. 1, pp. 1-16, Dec. 2015.

[15] Namje Park, Jin Kwak, Seungjoo Kim, Dongho Won, and Howon Kim, "WIPI Mobile Platform with Secure Service for Mobile RFID Network Environment", Journal of AWNTA, pp. 741-748, Jan. 2006.

[16] Namje Park, Hongxin Hu, and Qun Jin, "Security and Privacy Mechanisms for Sensor Middleware and Application in Internet of Things (IoT)", Journal of Distributed Sensor Networks, Vol. 2016, Article ID 2965438, 3pages, Jan. 2016. https://doi.org/10.1155/2016/2965438

[17] Jaehyun Se, "Business Value of Blockchain and Applications of Artificial Intelligence", Journal of AJMAHS, Vol. 8, No. 7, pp. 779-789, Jul. 2018.

[18] Jin-Hee Ku, "A Study on Adaptive Learning Model for Performance Improvement of Stream Analytics", Journal of Convergence for Information Technology, Vol. 8, No. 1, pp. 201-206, 2018.

[19] JungYul Choi, "A study on the standardization strategy for building of learning data set for machine learning applications", Journal of Digital Convergence, Vol. 16, No. 10, pp. 205-212, Oct. 2018.

[20] R. Frost, D. Paul, and F. Li, "AI pro: Data Processing Framework for AI Models", IEEE 35th International Conference on Data Engineering (ICDE), Macau SAR, China, pp. 1980-1983, Apr. 2019

[21] A. Aoaddah, A.A. Elkalam, and A.A. Ouahman, "FairAccess: a new Blockchain-based access control framework for the Internet of Things", Journal of Security and Communication Networks, Vol. 9, No. 18, pp. 5943-5964, Feb. 2017.

[22] Jiseop Lee, Sooyoung Kang, and Seungjoo Kim, "Study on the AI Speaker Security Evaluations and Countermeasure", Journal of the Korea Institute of Information Security and Cryptology, Vol. 28, No. 6, pp. 1523-1537, Dec. 2018.

[23] Jinsu Kim, Sangchoon Kim, and Namje Park, "Face Information Conversion Mechanism to Prevent Privacy Infringement", Journal of KIIT, Vol. 17, No. 6, pp. 115-112, Jun. 2019.

[24] Jinsu Kim and Namje Park, "Inteligent Video Surveilance Incubating Security Mechanism in Open Cloud Environments", Journal of KIIT, Vol. 17, No. 5, pp. 105-116, May 2019.

[25] Namje Park and Marie Kim, "Implementation of load management application system using smart grid privacy policy in energy management service environment", Cluster Computing, Vol. 17, No. 3, pp. 653-664, Sep. 2014.

[26] Donghyeok Lee and Namje Park, "A Proposal of SH-Tree Based Data Synchronization Method for Secure Maritime Cloud", Journal of the Korea Institute of Information Security & Cryptology, Vol. 26, No. 4, pp. 929-940, Aug. 2016.

[27] Donghyeok Lee and Namje Park, "A Secure Almanac Synchronization Method for Open IoT Maritime Cloud Environment", Journal of Korean Institute of Information Technology Vol. 15, No. 2, pp. 79-90, Feb. 2017.

저자소개

김 인 경 (Inkyung Kim)

2015년 2월: 고려대학교 수학과 석사
2017년 5월 ~ 2019년 4월 : 한국원자력통제기술원 사이버보안실 전문연구원
2019년 5월 ~ 현재 : 제주대학교 창의교육거점센터, 사이버보안인재교육원 책임연구원
관심분야 : 사이버보안 평가, 기반시설 보안, 융합기술 보안 등

박 남 제 (Namje Park)

2008년 2월 : 성균관대학교 컴퓨터공학과 박사
2003년 4월 ~ 2008년 12월 : 한국전자통신연구원 정보보호연구단 선임연구원
2009년 1월 ~ 2009년 12월 : 미국 UCLA대학교 공과대학 Post-Doc, WINMEC 연구센터 Staff Researcher
2010년 1월 ~ 2010년 8월 : 미국 아리조나 주립대학교 컴퓨터공학과 연구원
2010년 9월 ~ 현재 : 제주대학교 초등컴퓨터교육전공, 대학원 융합정보보안학과 교수
2011년 9월 ~ 현재 : 창의교육거점센터장, 과학기술사회(STS)연구센터 부센터장, 정보영재 주임교수, 사이버보안인재교육원장
관심분야 : 융합기술보안, 컴퓨터교육, 스마트그리드, IoT, 해사클라우드 등