ITU-T SG17 AI 보안 국제표준화 동향

오흥룡
한국정보통신기술협회
hroh@tta.or.kr

International Standardization Trends for AI Security in ITU-T SG17

Oh Heung Ryong
Telecommunications Technology Association(TTA)

요 약

인공지능(AI)은 현대 사회에서 가장 유망받고 있는 기술이며, 기존의 여러 기술 및 서비스들과도 접목되어 점차 발전하고 있는 기술이다. 특히, AI 기반의 LLM(Large Language Model) 발전은 인간의 학습 능력을 모방하고, 부족한 정보량을 강화하거나 지식의 역량을 확장하는 시스템으로 역동적이고 급속히 발전하고 있다. 그러나 AI 기술의 발전에 따른 역작용으로 사용자들의 정보가 무분별하게 활용되거나 잘못된 정보가 확산되는 사례가 있고, 국가 및 조직을 위협하는 새로운 보안 위협들이 확대되는 문제점도 있다. 따라서, AI 기술 발전과 함께 이에 대한 문제점들을 완화하거나 해결하기 위한 AI 보안기술이 함께 논의되고 개발되어야 한다. 본 논문에서는 UN 산하에서 정보통신 관점에서 국제표준을 개발하고 있는 ITU-T SG17 Q7(안전한 응용서비스 보안)에서 논의되고 있는 AI 보안 국제표준화 활동 현황에 대해 살펴보고자 한다.

Ⅰ. 서론

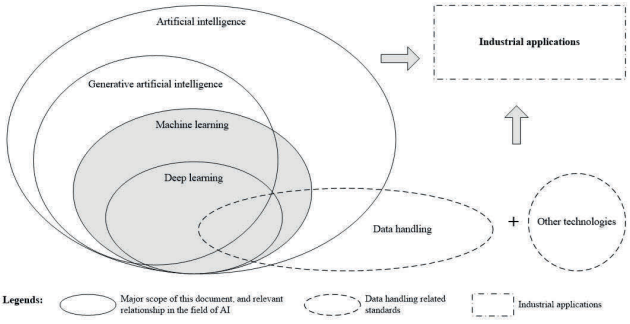
AI 기술의 발전은 인간의 인지 능력을 강화하고, 산업 현장에서의 효율성 및 생산성 강화와 사회적 위험에서 사용자들을 안전하게 보호할 수 있는 기술로 자리매김하고 있다. 또한, 생성형 인공지능은 AI 기술 기반으로 텍스트, 이미지, 오디오, 비디오 등 다양한 형식의 콘텐츠를 분석 및 재생성하는 기술로 활용되고 있다. 그러나 AI 시스템의 역량이 확장됨에 따라 보안상의 도전 과제와 잠재적 취약성도 점점 더 복잡해지고 있으며, 생성형 AI 기술이 사이버범죄에 활용되는 사례 증가와 사용자들의 프라이버시를 위협하고 있어, AI 기술의 발전과 함께 법, 제도, 규제, 대응책들이 함께 고려되어야 한다. AI 기술은 국가 간, 기업 간, 서비스 도메인 간에 다양한 형태로 구현 및 활용될 수 있어, 이해당사자들 간에 이해 격차를 줄이고 안전하게 구현될 수 있도록 국제표준들이 반드시 활용되어야 한다. 본 논문에서는 ITU-T SG17 Q7에서 논의되고 있는 AI 보안 기술의 국제표준화 활동 현황들을 살펴보고, 향후 국내외 표준화 활동에 관심 있는 전문가들에게 동향 정보를 제공하고자 한다.

Ⅱ. AI 보안 국제표준화 현황

ITU-T SG17에서는 AI 보안 표준 개발을 위한 사전 조사를 위해 서신그룹(Correspondence Group)을 신설하였으며, 2024.9-2025.3월까지 아래와 같은 3가지 목표로 활동을 하였다.

- AI 시스템에서의 보안리스크, 보안위협, 보안도전 관련 상황 분석
- AI 보안에서의 차세대 기술 식별 및 현재 기술 수준 분석
- AI 보안 구현을 위한 상위 수준의 기술 지침 개발

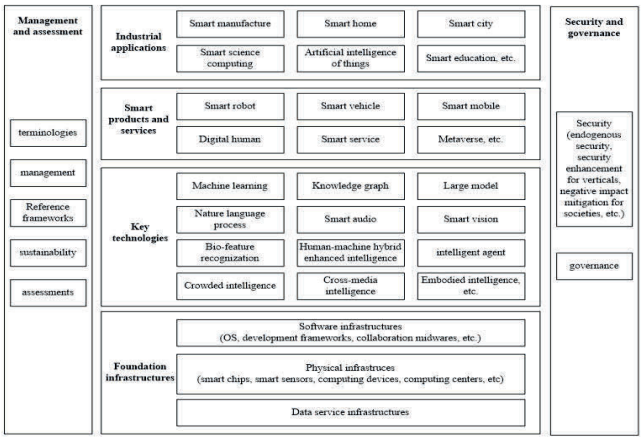
AI 보안 서신그룹은 활동 결과물로 “AI 보안 전략 및 표준화 환경” 보고서를 개발하였으며, 이 보고서는 앞서 언급된 3가지 목표와 다른 국제표준화기구(ITU-T SGs, ISO/IEC, IEEE, ETSI SAI)에서의 AI 보안 활동 현황을 포함하고 있다[2].



[그림 1] AI 기술의 관계도

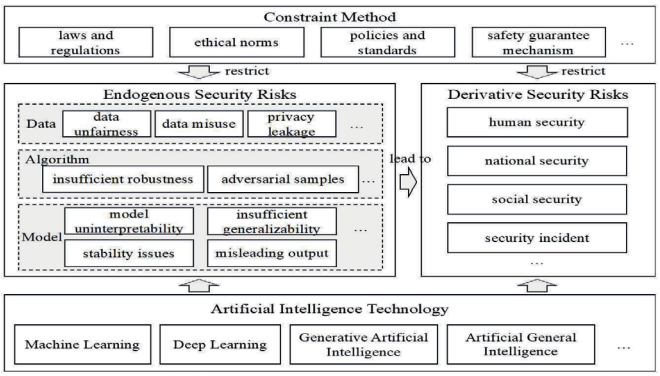
[그림 1]은 AI 기술의 관계도와 데이터가 활용되어 산업 환경에 적용되는 개념을 설명하고 있다. 딥러닝은 대용량 데이터에서 특정 학습(예: 개나 고양이 이미지)을 통해 정답을 찾는 개념이며, 머신러닝은 대용량 데이터에서 주어진 패턴을 학습하고 예측 및 판단의 개념이 포함되어 있다. 이러한 개념들을 포함하고 있는 인공지능은 사람처럼 추론하고, 학습하고, 스스로 진화하여 최적화된 판단을 결정할 수 있는 개념이다. 결과적으로 이러한 학습 모델은 학습데이터가 중요하게 작용을 미치고, 데이터 생명주기 관점에서 각 단계별 보안 위협 분석과 정보보호 기술들이 함께 고려되어야 한다.

ITU-T SG17에서 AI 보안 표준화는 2021.4월, 한국 주도로 빅데이터 환경에서 프라이버시 및 민감한 데이터를 모니터링하는 권고안 개발을 시작하였고, AI 기술 키워드를 처음으로 사용한 것은 2023.8월, 한국 주도로 AI 시스템에서 보안 요구사항 권고안 개발을 시작으로 본격적으로 국제표준화 활동을 착수하였다. 현재는 SG17 정기회의마다 AI 보안에 대한 신규 표준화 아이템 제안 비율이 가장 높고, 최근에 개최된 2025.4월, 국제회의에서 개발중에 있는 표준화 아이템과 신규 표준화 아이템들이 승인되어 AI 보안 관련 총 18건의 권고안과 기술보고서들이 개발되고 있다.



[그림 2] AI 기술의 보안 표준화 환경 분석

[그림 2]는 ITU-T SG17에서 AI 보안 국제표준을 개발할 때 표준화 아이템들 간에 중복성을 방지하기 위한 환경 분석을 나타내고 있다. AI 보안 기술은 인프라 영역, 핵심기술 영역, 제품 및 서비스 영역, 산업화 응용 영역, 그리고 4개의 영역에서 공통으로 사용할 수 있는 관리 및 평가 영역, 보안 및 거버넌스 영역으로 구성하였다.

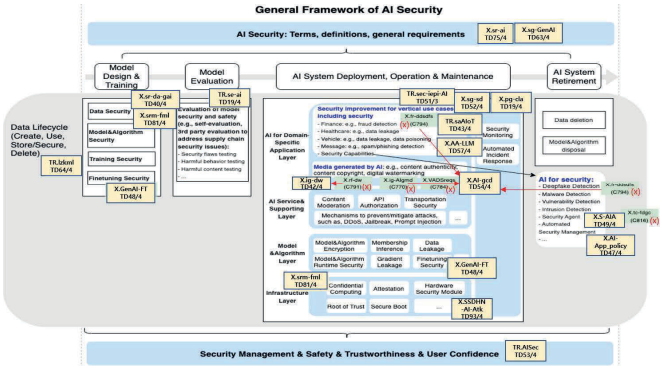


[그림 3] AI 기술의 보안 위협

[그림 3]은 AI 기술의 보안 위협으로 AI 기술 자체에 내재된 보안 위협과 AI 기술을 활용함에 따라 새롭게 발생되는 보안 위협을 나타내고 있다. [표 1]은 ITU-T SG17에서 개발되고 있는 AI 보안 표준화 아이템들의 현황이다.

[표 1] AI 보안 표준화 아이템 현황

No.	아이템 번호	권고/기술보고서 제목	제안 국가	완료 시점
1	TR.AISec	Technical Report: Artificial intelligence security standardization strategies	중국	2025.12
2	TR.Izkml	Technical Report: Landscape analysis of Zero-Knowledge machine learning	한국	2026.9
3	TR.saAIoT	Technical Report: Security Threat Analysis for Artificial Intelligence of Things on Devices	중국	2025.12
4	TR.se-ai	Technical Report: Security Evaluation on Artificial Intelligence Technology in ICT	중국	2025.12
5	X.AA-LLM	Guidelines for Preventing and Mitigating Adversarial Attacks on LLMs in Metaverse and Digital Twin Environments	인도	2026.12
6	X.AI-App_policy	Reference architecture for AI-assisted analysis of consistency between App's data usage behaviourand its privacy policy	중국	2026.9
7	X.AI-gcd	Guidelines for Artificial Intelligence-generated content detection	중국	2027.1
8	X.GenAI-FT	Security guidelines for fine-tuning generative AI model	중국	2027.3
9	X.S-AIA	Security requirements and guidelines for Artificial Intelligence agent	중국	2027.3
10	X.sg-GenAI	Security Guidelines for Generative Artificial Intelligence Application Service	중국	2027.6
11	X.sg-sd	Security guidelines for synthetic data in the context of AI systems	중국	2027.9
12	X.sr-ai	Security requirements for AI systems	한국	2026.6
13	X.sr-da-gai	Security threats and requirements for data annotation service of generative artificial intelligence	중국	2027.3
14	X.SSDHN-AI-Atk	Security Guidelines for Software-Defined Heterogeneous Networks Architecture against AI generated Attacks and Threats	인도	2026.9
15	X.pg-cla	A guideline for continual learning to actively respond to network attacks	한국	2027.9
16	X.ig-dw	Implementation guidelines for digital watermarking	중국	2026.9
17	X.srm-fml	Security requirements and measures of federated machine learning	중국	2026.6
18	TR.sec-iepi-AI	Technical Report: Security guideline for artificial intelligence applications of IoT-based electric power infrastructure monitoring	중국	2027.10



[그림 4] AI 보안 표준화 아이템들의 관계도

[그림 4]는 AI 보안 표준화 아이템들을 데이터 생명주기 관점에서 각 아이템들이 어느 영역 및 기술에 중점을 두고 있는지에 대한 관계도를 나타내고 있다. 이는 ITU-T SG17에서 향후 새롭게 제안되는 아이템들 간에 중복성 배제를 위해 활용될 수 있다.

Ⅲ. 결론

본 논문에서는 ITU-T SG17에서 논의되고 있는 AI 보안에 대한 국제표준화 활동 현황에 대해 살펴보았다. 현대 사회에서 AI 기술은 선택이 아닌 모든 산업과 일상적인 생활에서 활용되는 기술로 빠르게 자리매김하고 있어, AI 기술 발전과 함께 보안 관점에서 연구가 진행되어야 한다. 또한, 국가 간에 개인정보 교환과 능동적인 사이버보안 공동 대응 등 글로벌한 연구가 진행되고 있어 국제표준 개발에 산학연 전문가들의 적극적인 관심과 참여가 요구되고 있다. ITU-T SG17에서는 AI 보안 국제표준 개발을 시작하는 단계에 있어 한국에서도 적극적인 대응이 요구되는 시점이다.

ACKNOWLEDGMENT

본 논문은 2025년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(No. 2022-0-00009, ICT 국제공식 표준화 대응 및 국가표준 연구).

참 고 문 헌

[1] ITU-T SG17 Homepage, https://www.itu.int/en/ITU-T/studygroups/2025-2028/17/Pages/default.aspx

[2] ITU-T SG17 TD40/PLEN, Draft CG-AISEC Deliverable "Artificial intelligence security strategies and standardization landscapes" in the inter-regnum (September 2024 - March 2025).