

보도시점 2024. 12. 24.(화) 12:00
(2024. 12. 25.(수) 조간)

배포 2024. 12. 24(화) 09:00

한-미 ‘인공지능 신뢰성 체계(프레임워크)’ 간 높은수준의 상호호환성 확인

- 과기정통부한국정보통신기술협회 ‘신뢰할 수 있는 인공지능 개발안내서’, 미 상무부
국립표준기술연구소(NIST) ‘인공지능 위험관리프레임워크(AI RMF) 상호 교차분석
(crosswalk) 결과 발표(12.24.)

과학기술정보통신부(장관 유상임, 이하 ‘과기정통부’)와 한국정보통신기술협회(회장 손승현, 이하 ‘TTA’)는 우리나라 ‘신뢰할 수 있는 인공지능 개발안내서’(이하 ‘AI개발안내서’)와 미 상무부 국립표준기술연구소(이하 ‘NIST’)의 인공지능 위험관리체계(프레임워크)(이하 ‘AI RMF’) 간 상호 교차분석*(Crosswalk)을 완료하였다고 밝혔다.

* 한국정보통신기술협회(‘인공지능 개발안내서’ 연구기관)와 미국 국립표준기술연구소(‘인공지능 위험관리체계<AI RMF>’ 연구기관)가 교차분석 진행(‘24.2.~12.)

※ 상호 교차분석 세부결과는 한국정보통신기술협회 누리집(www.tta.or.kr/tta), 미국 국립표준기술연구소 누리집(airc.nist.gov/AI_RM_F_Knowledge_Base/Crosswalks), 한국 인공지능 안전연구소 누리집(www.aisi.re.kr/artide)에서 확인 가능

과기정통부와 한국정보통신기술협회는 ’21년부터 ‘인공지능 개발안내서’ 개발·보급을 통해 인공지능 신뢰성 확보를 위한 15개 기술 요구사항(67개 세부 검증항목)을 제시하고, 국내기업이 인공지능 체계를 개발하는 과정에서 자발적으로 인공지능 신뢰성을 확보할 수 있도록 지원해왔다. 또한, ‘인공지능 개발안내서’를 기반으로 ‘인공지능 체계 신뢰성 제고를 위한 요구사항’ 단체 표준 제정(’23.12.), ‘민간자율 인공지능 신뢰성 인증’(’23.11.~, 총7건 인증) 운영 등 지속가능한 인공지능 혁신생태계 조성을 뒷받침하기 위해 인공지능 신뢰·안전성 확보에 대한 정책적 노력을 지속해 왔다.

< 한국 '인공지능 개발안내서'('22.2.) 개요 >

- 인공지능 체계 개발과정에서 인공지능 신뢰성 확보 및 위험 저감대응을 위한 기술 요구사항을 인공지능 생명주기 5단계에 따라 구체적으로 제시하여 민간의 자율적인 인공지능 신뢰성 확보 지원

* 계획 및 설계 - 데이터 수집 및 처리 - 인공지능 모형 개발 - 체계 구현 - 운영 및 점검

미국 국립표준기술연구소는 인공지능 신뢰성 연구와 방침 개발을 주도하는 정부기관으로, 개인·조직이 인공지능 체계 설계·개발·배포·활용과 관련된 위험을 이해·관리·완화하는 것을 지원하기 위한 자발적 체계인 '인공지능 위험관리 체계(AI RMF)'를 '23년1월 발표하였다. 미국 국립표준기술연구소 '인공지능 위험관리 체계(AI RMF)'는 세계적 차원에서 민간·공공분야 주체들의 인공지능 신뢰성 확보방안 논의·발전과정에서 주요한 역할을 해왔으며, 전세계 많은 조직들이 인공지능 신뢰성 확보를 위한 기초 체계로 도입·활용하고 있다.

- * (미국 '인공지능 위험관리 체계(AI RMF)'와 상호 교차분석 사례) 국제표준(ISO-IEC-42001:AI경영시스템), 싱가포르 AI Verify('23.10.), 일본 AI사업자 가이드라인('24.4./'24.9., 용어·개념 한정) 등

< 미국 '인공지능 위험관리체계(AI RMF)'('23.1.) 개요 >

- 인공지능 체계를 설계, 개발, 배포 또는 사용하는 조직이 인공지능체계로 인한 위험을 관리하여 신뢰할 수 있고 책임감 있는 인공지능 체계를 개발·사용을 촉진시키는데 도움이 되는 지침문서로 국립표준기술연구소(NIST)가 「2020 국가 인공지능 추진 전략법(AI이니셔티브법)」에 근거하여 개발

이번 상호교차 분석(Crosswalk)은 국제적으로 인공지능 신뢰성 분야에서 가장 영향력 있는 체계 중 하나인 미국 국립표준기술연구소 '인공지능 위험관리 체계(AI RMF)'와 우리나라 '인공지능 개발안내서' 특성을 분석하고, 세부항목 간 정합성을 확인하여 상호호환성을 확보해 나가기 위한 목적으로 진행되었다. 상호교차 검증을 진행한 결과, 세부항목의 상당부분이 조화됨을 확인*하였으며, 인공지능 신뢰성을 높이고 관련 위험을 완화하기 위해 양 방침이 유사하게 활용될 수 있음을 확인하였다.

- * 항목별로 구체성에 일부 차이는 있으나 대다수 항목이 상호 호환한 '인공지능 개발안내서' 기준 67개 세부 검증항목 중 63개 항목이 미국 국립표준기술연구소 '인공지능 위험관리체계<AI RMF>'와 호환

아울러, 미국 국립표준기술연구소 ‘인공지능 위험관리체계(AI RMF)’와 교차분석을 통해 세계적 차원에서 우리나라 인공지능 신뢰성 기술·검증체계의 수준·완성도를 확인하였으며, 이번 교차 분석결과를 토대로 과기정통부와 한국정보통신기술협회는 미국 국립표준기술연구소와의 기술협력을 더욱 강화하고, 인공지능 신뢰성 분야 국제 기술표준·규범과의 상호호환성 확보를 위한 노력도 확대할 계획이다. 또한, 미국 국립표준기술연구소 ‘인공지능 위험관리 체계AI RMF’와 차이가 발생하는 부분은 ‘인공지능 개발안내서’ 고도화 시 반영·보완하여 우리나라 인공지능 신뢰성 기술·검증체계의 국제적 정합성을 제고할 예정이다.

한국정보통신기술협회 손승현 회장은 “이번 상호 교차분석 작업을 통해, 우리나라의 인공지능 신뢰성 확보체계를 세계적 수준으로 끌어올리는 계기가 되었으며, 앞으로도 국제 협력 확대를 통해 인공지능 신뢰성 기술·검증체계의 공신력을 강화하고, 국내산업 지원체계를 공고히 하여 인공지능 신뢰성분야 전문기관으로서 역할에 최선을 다하겠다.”라고 밝혔다.

과기정통부 송상훈 정보통신정책실장은 “최근 법사위를 통과한 「인공지능 기본법」에도 인공지능 신뢰·안전성 확보를 위한 정부지원 근거를 폭넓게 마련했다”면서, “국내 인공지능 산업의 신뢰·안전성 확보를 위한 정책적 지원을 강화하고, 국내기업의 세계 진출이 확대될 수 있도록 인공지능 안전연구소, 한국정보통신기술협회 등 전문기관과 함께 인공지능 신뢰·안전성 분야 국제표준화도 적극 추진해 나가겠다”고 강조하였다.

담당 부서	과학기술정보통신부 인공지능기반정책과	책임자	과 장	남철기 (044-202-6270)
		담당자	사무관	이재호 (044-202-6262)
관련 기관	한국정보통신기술협회 AI신뢰성센터	책임자	센터장	신준호 (010-3439-5536)
		담당자	팀 장	곽준호 (010-5110-2693)



더 아픈 환자에게 양보해 주셔서 감사합니다
가벼운 증상은 동네 병·의원으로



◆ **(진행경과)** AI신뢰성 연구·표준화 업무 협의(‘24.2., TTA·NIST) → 상호 교차분석 보고서 초안 전달(TTA→NIST, ‘24.6.) → TTA·NIST 상호 교차 검토(‘24.8.~11., 3회) → 분석결과 최종합의(‘24.12.)

- **(개요)** 韓 과기정통부·TTA ‘AI개발안내서’, 美 NIST ‘AI RMF’는 AI신뢰성 확보와 AI위험 관리를 위해 필요한 요구사항을 정리한 가이드라인
- **(韓 AI개발안내서)** AI신뢰성 확보 및 위험 저감·대응을 위한 기술 요구사항을 AI의 생명주기 5단계에 따라 구체적으로 제시(총 67개)
* 5단계: 기획 – 데이터 처리 – 모델 개발 – 시스템 개발 – 운영·관리
 - **(美 AI RMF)** AI위험 저감·대응을 위한 기술 및 조직·운영 관점의 요구사항을 위험관리 절차 4단계에 따라 구체적으로 제시(총 72개)
* 4단계: 거버넌스(GOVERN)–매핑(MAP)–측정(MEASURE)–관리(MANAGE)
- **(결과)** 양 가이드라인 간 공통점과 차이점이 존재, 가이드라인에 제시된 항목의 구체성과 관점의 특성에 기인하며, 상호보완적으로 적용 가능
- **(공통점)** 항목별로 일부 구체성에 차이는 있으나 대다수 상호 호환 (AI개발안내서 67개 세부검증항목 中 63개 호환)
 - **(차이점)** 양 가이드라인 간 차별화되는 항목들은 아래와 같음
< 양 가이드라인 간 차이점 분석 >

韓 ‘AI개발안내서’	美 NIST ‘AI RMF’
■ 세부 검증항목 67개 中 4개항목은 美 ‘AI RMF’와 차이 존재 ○ 데이터 처리 부문(3개): 데이터 오류 처리, 데이터 다양성 확보방안, 과도한 데이터 특성 처리 ○ AI시스템 편향 제거 부문(1개): 사용자 인터페이스로 인한 편향 제거 등	■ 위험관리 항목 72개 中 19개 항목은 韓 ‘AI개발안내서’와 차이 존재 ○ GOVERN(4개): AI위험관리 세부절차 ○ MAP(5개): AI활용 맥락에 대한 이해 ○ MEASURE(6개): 평가방안 선택·문서화 ○ MANAGE(4개): 3자 검증 및 서비스 개시 후 반복적 위험관리 수행 등

※ 항목별 포함 관계 및 연계 여부는 상호 교차분석(Crosswalk) 결과보고서 참조

- **(향후계획)** ‘AI개발안내서’ 고도화 시, 양 가이드라인 간 차이점에 대해 반영·보완하여 글로벌 정합성 제고(‘25.1.~)