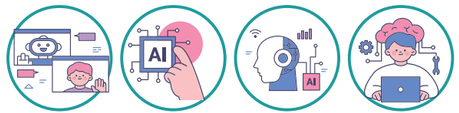
인공지능이 펼칠
혁신에 믿음을 더합니다.

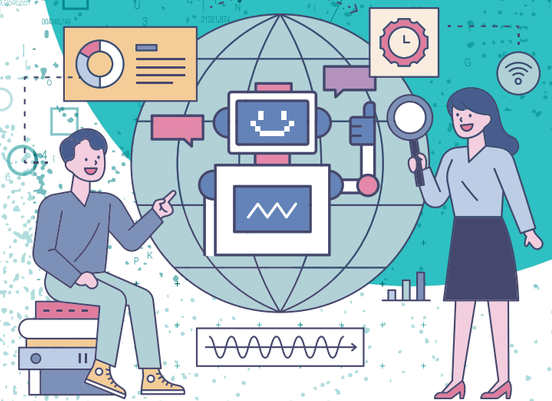
TTA AI융합시험연구소

자세한 정보는
QR코드를 촬영하세요

TOWARD TRUSTWORTHY AI



신뢰할 수 있는 인공지능 개발 안내서



TTA 한국정보통신기술협회
Telecommunications Technology Association

경기도 성남시 분당구 분당로 47(서현동 267-2)
TEL. 031)724-0001 trustworthyai@tta.or.kr

과학기술정보통신부
Ministry of Science and ICT

TTA 한국정보통신기술협회
Telecommunications Technology Association

신뢰할 수 있는 인공지능 개발 안내서

인공지능 제품 및 서비스를 개발하는 과정에서 신뢰성 확보를 위한 참고자료로 활용할 수 있도록, 15개의 개발 요구사항과 67개의 검증항목을 제공하는 인공지능 신뢰성 확보 길잡이

개발 안내서 활용 취지

신뢰성 확보를 위한 기술 지식 습득

생명주기별 요구사항 및 검증항목을 통한 필요성 인식 및 관련 기술 정보, 고려사항 습득

개발·서비스 중인 인공지능 시스템의 신뢰성 점검

인공지능 서비스 개발 및 운영 시 추가 고려가 필요한 신뢰성 요구사항에 대해 점검 및 확인

'신뢰할 수 있는 인공지능 개발 안내서 - 일반 분야'

1. 최신 국제 권고안 및 표준의 지속적 반영
2. 기술적 타당성 및 효용성을 고려한 본문 수정
3. 의견 수렴 및 자문 결과 반영을 통한 활용도 제고

분야별 '신뢰할 수 있는 인공지능 개발 안내서'

1. 최신 국제 권고안 및 표준의 지속적 반영
2. 기술적 타당성 및 효용성을 고려한 본문 수정
3. 실무 활용도 제고를 위한 검증항목의 수정·추가 및 각 항목 내 분야별 특화 사례 수록

의료 | 자율주행 분야 | 공공·사회 분야 | 생성AI | 스마트 치안 | 채용

기업/기관 자체 신뢰성 확보 지침 및 점검표 마련

인공지능 신뢰성의 정보 격차를 완화하고 서비스 특성 및 개발 목적에 따른 항목 선별을 통한 자체 신뢰성 점검표 마련

인공지능 신뢰성 담론의 수집 및 논의의 장 마련

사회/산업 구성원의 다양한 의견 검토와 논의를 통한 합의와 공감대 마련의 초석

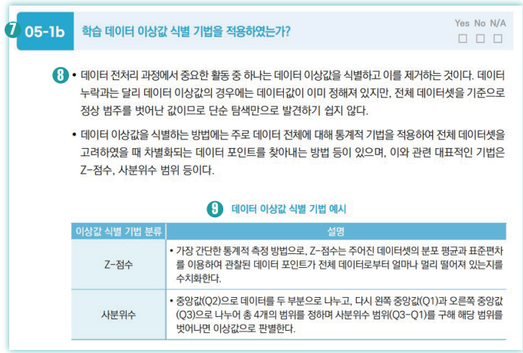
인공지능 신뢰성 확보를 위한 국제 요구사항 반영

표준화기구, 기술단체, 국제기구, 주요국에서 인공지능 신뢰성 확보를 위해 발표한 정책, 권고안, 그리고 표준을 기반으로 기술적 요구사항을 도출하고 구체화

요구사항	다양성존중	책임성	안전성	투명성
요구사항01 인공지능 시스템에 대한 위험관리 계획 및 수행		✓		✓
요구사항02 인공지능 거버넌스 체계 구성	✓	✓	✓	✓
요구사항03 인공지능 시스템의 신뢰성 테스트 계획 수립			✓	✓
요구사항04 데이터의 활용을 위한 상세 정보 제공		✓		✓
요구사항05 데이터 강건성 확보를 위한 이상 데이터 점검			✓	
요구사항06 수집 및 가공된 학습 데이터의 편향 제거	✓	✓		✓
요구사항07 오픈소스 라이브러리의 보안성 및 호환성 확보		✓	✓	
요구사항08 인공지능 모델의 편향 제거	✓			
요구사항09 인공지능 모델 공격에 대한 방어 대책 수립			✓	
요구사항10 인공지능 모델 명세 및 추론 결과에 대한 설명 제공		✓		✓
요구사항11 인공지능 시스템 구현시 발생 가능한 편향 제거	✓			
요구사항12 인공지능 시스템의 안전 모드 구현 및 문제발생 알림 절차 수립		✓	✓	✓
요구사항13 인공지능 시스템의 설명에 대한 사용자의 이해도 제고				✓
요구사항14 인공지능 시스템의 추적가능성 및 변경이력 확보		✓		✓
요구사항15 서비스 제공 범위 및 상호작용 대상에 대한 설명 제공		✓		✓

개발 안내서 구성

1. **핵심 속성** — 요구사항을 통해 확보할 수 있는 윤리 속성 제시
2. **요구사항** — 인공지능 신뢰성 확보를 위한 생명주기 단계별 기술 요구사항 제시
3. **대표 행위자** — 요구사항별 신뢰 확보 활동 수행을 위한 대표 행위자 제시
4. **요구사항 해설** — 요구사항의 필요성, 기술 해설 등 요구사항 이해를 위한 설명 제공
5. **세부 요구사항** — 요구사항 만족을 위한 세부항목 제시
6. **세부 요구사항 해설** — 세부 요구사항의 필요성, 기술 해설 등 세부항목 이해를 위한 설명 제공



7. **자가검증항목** — 세부 요구사항 충족을 위해 확인해야 할 자가검증항목 제시
8. **자가검증항목 해설** — 인공지능 신뢰성 확보를 위한 생명주기 단계별 기술 요구사항 제시
9. **참고 사항** — 세부 요구사항과 관련된 예시, 사례 등 참고 정보 제공