

보도 시점 2025. 12. 10.(수) 14:00
(2025. 12. 11.(목) 조간) 배포 2025. 12. 10.(수) 09:00

과기정통부, 「인공지능(AI) 보안 안내서」 발표

- 모델 개발자, 서비스 제공자, 이용자를 대상으로 총 113개의 보안 요구사항 제시

과학기술정보통신부(부총리 겸 과기정통부 장관 배경훈, 이하 ‘과기정통부’)와 한국인터넷진흥원(원장 이상중, 이하 ‘KISA’)은 외부의 사이버 위협으로부터 인공지능 모델·서비스를 보호하기 위해 기업과 국민이 참조할 수 있는 「인공지능(AI) 보안 안내서」를 마련하였다고 밝혔다.

인공지능 대전환(AI) 시대가 본격 도래하면서 국민 일상을 넘어 국가 경제·사회 전 영역에서 인공지능 기반 서비스의 도입이 확대되고 있다. 이에 정부는 국민과 기업이 인공지능의 혜택을 안전하게 누릴 수 있도록 그간 지침, 방침(가이드라인) 등을 발간해 왔지만, 주로 인공지능 생성물, 추론 과정 등에 대한 신뢰성 및 윤리적 활용의 관점*에서 다루어지면서 외부 사이버 위협을 예방·대응하는 보안 관점에서의 상세한 가이드 또는 기준은 제시되지 않았었다.

* 신뢰할 수 있는 인공지능 개발 안내서(과기정통부), 생성형 인공지능 윤리 안내서(가이드북)(한국지능정보사회진흥원, '23) 등

이에 과기정통부는 산·학·연 전문가들과 함께 국내·외 인공지능 보안 정책 관련 사례 등을 분석하고 관계기관·업계 대상으로 의견 수렴을 거쳐 인공지능 모델 개발자, 서비스 제공자, 이용자 등이 보안 측면에서 참고할 수 있는 「인공지능(AI) 보안 안내서」를 마련하였다.

「인공지능(AI) 보안 안내서」는 인공지능 모델·시스템에 대해 보안 3대 요소(기밀성, 무결성, 가용성)를 기반으로 개발자, 서비스 제공자 대상 인공지능 생애주기 전 단계* 보안 요구사항과 이용자의 인공지능 보안 수칙 등 총 113개의 항목으로 구성되었다.

* 1) 개발자 생애주기 : ①계획 및 설계 → ②데이터 수집 및 준비 → ③모델 개발 → ④모델 배포 → ⑤점검(모니터링) 및 유지보수 → ⑥파기

2) 서비스 제공자 생애주기 : ①서비스 계획 및 설계 → ②서비스 개발 및 구축 → ③서비스 제공 및 운영 → ④서비스 유지보수 및 지원 → ⑤환류(피드백) 및 서비스 개선 → ⑥파기

구체적으로는 모델 개발 단계에서 인공지능 모델 위험 관리, 데이터 암호화, 실시간 점검(모니터링) 등 인공지능 보안 내재화(Secure by Coding)를 실현하기 위한 보안 요구사항을 제시하고, 서비스 제공 단계에서는 서비스 이상 행위 탐지, 애플리케이션 프로그래밍 인터페이스(API)·인터페이스 보안, 백업시스템 구축 등 서비스 복원력 관점에서의 보안 요구사항을 제시하였다. 마지막으로 중요 정보 입력금지, 인공지능 악용 금지 등 이용자를 위한 구체적인 보안 수칙을 제시하였다.

이번 보안 안내서는 국제 기준 및 방침(가이드라인)* 등을 참고하여 인공지능 보안 요구사항을 세계적 수준으로 정립함으로써, 국내 모델뿐만 아니라 해외 모델에도 적용할 수 있도록 범용성과 국제 규범 간 호환성을 확보하였다는 점이 특징이다.

* NIST AI RMF 1.0(미국, '23), OWASP Top 10 for LLM Applications(OWASP, '25) 등

과기정통부 임정규 정보보호네트워크정책관은 “이번 인공지능 보안 안내서는 고도화되는 사이버 위협 속에서 인공지능을 안전하게 개발·활용할 수 있도록 하는 중요한 이정표가 될 것”이라고 기대하면서, “인공지능에 대한 새로운 보안 위협을 점검(모니터링)하여 반영하는 등 보안 안내서를 지속 갱신(업데이트)해나갈 예정”이라고 덧붙였다.

「인공지능(AI) 보안 안내서」는 과기정통부, 한국인터넷진흥원(KISA) 누리집(홈페이지)을 통해 무료로 내려받아 사용할 수 있다.

담당 부서	정보보호네트워크정책관 정보보호기획과	책임자	과장	김연진 (044-202-6440)
		담당자	사무관	심재환 (044-202-6441)

내일을 만드는 과학기술
내일을 채우는 디지털·AI

대한민국
지책브리핑





AI 이용자를 위한 정보보호 수칙



01 AI 서비스 접속

1

공식 사이트에서만 다운로드
서비스 접근·가입 시, 공식 사이트를
통해 가입 및 프로그램 다운로드하기

2

안전한 비밀번호 설정 및 주기적 변경
특수문자를 포함한 강력한 비밀번호로
설정하고, 주기적(3개월)으로 변경하기

3

공개된 장소에서 이용 금지
카페 등 공공장소 및 공개된 네트워크에서
서비스 이용, 정보 입력 등 금지

4

법률 및 이용약관 확인
서비스 관련 법률, 이용약관에서 정하는
금지행위, 이용자 권리 등 확인하기

02 AI 서비스 이용

1

중요 정보 입력 금지
개인정보, 기밀정보 등 중요 정보 및 허위
콘텐츠 생성을 위한 허위 정보 등 입력 금지

2

결과에 대한 정확성 검증
AI로 생성되는 정보에 대한 정확성과
함께 정보의 최신성, 편향성 등 검증 필수

3

데이터 삭제
입력, 생성된 정보는 반드시 삭제하고,
필요 시에는 별도로 다운로드하여 저장하기

4

최신 보안 업데이트 적용
서비스의 안전성, 안정성 등 유지를 위한
보안 패치는 최신버전으로 업데이트 필수

03 AI 악용 피해 예방

1

항상 의심하고 확인
정보의 허위·조작 가능성 등을 항상 의심하고
생성물 활용 분야와 목적을 반드시 확인

2

AI 악용이 의심되면 삭제
AI를 악용하여 제작된 콘텐츠, 사이트,
프로그램 등이 의심되면 삭제하고 신고하기

3

AI 악용 결과를 공유 금지
악성코드, 허위 콘텐츠 등으로 의심되는
결과물을 소장하거나 다른 사람에게 공유 금지

4

허위 콘텐츠 매매 금지
AI를 악용해 생성된 허위 콘텐츠(가짜 뉴스,
영상, 음성 등)를 돈으로 사거나 팔지 않기