

보도시점 2026. 7. 8.(수) 12:00
(2026. 7. 9.(목) 조간)

배포 2026. 7. 8(수) 09:00

AI 보안 진단부터 대응까지, 공격자 관점으로 AI 보안 취약점을 잡는다.

- 프롬프트 인젝션에서 데이터 유출까지... AI 보안 위협 대응 가이드 2종 발간
- 경영진부터 보안실무자, 레드팀 운영진까지 현장에서 즉시 이용 가능한 진단·대응 기준 제시

【관련 국정과제】 23-4. AI시대를 지탱하는 견고한 디지털 보안·안전 체계 구축

과학기술정보통신부(부총리 겸 과기정통부 장관 배경훈, 이하 ‘과기정통부’)와 한국인터넷진흥원(원장 이상중)은 AI 서비스의 보안 위협에 효과적으로 대응하고 현장의 보안 점검 역량을 강화하기 위해 ‘AI 보안 위협 대응 매뉴얼’, ‘AI 보안 레드티밍 가이드’를 발간했다고 밝혔다.

최근 AI 기술이 다양한 산업과 서비스에 빠르게 확산되면서 생산성과 혁신을 견인하고 있으나 프롬프트 인젝션, 권한 오남용, 데이터 유출 등 새로운 유형의 보안 위협도 함께 증가하고 있다.

특히, AI 보안 위협은 기존 정보보호 체계만으로는 충분히 검증하고 대응하기 어려운 특성을 가지고 있어, 실제 공격자 관점에서 보안 취약점을 식별하고 대응 방안을 마련하는 AI 레드팀*(AI Red Team)의 중요성이 높아지고 있다.

* AI 모델에 의도적으로 문제가 되는 행동을 유도하여 취약점과 위험을 발견하는 팀

이번에 발간된 2종의 가이드는 AI 환경에서 발생 가능한 주요 보안 위협과 사례를 체계적으로 정리하고, 실제 현장에서 활용할 수 있는 점검 및 대응 방안과 레드팀 운영 방안을 제시하는데 중점을 두었다.

「AI 보안 위협 대응 매뉴얼」은 ▲AI 보안위협 분류* 및 진단, ▲산업별** 위협 시나리오, ▲AI 보안 위협별 대응 방안 등을 담고 있으며, AI에 특화된 보안 위협과 실무 중심의 대응 방안을 제공한다. 특히, C-레벨급 경영진에게는 AI 보안 위협의 유형과 사례를, AI 보안 실무자, IT 운영자 등 실무자에게는 AI 보안 위협을 어떻게 진단하고 대응할 것인지에 대한 구체적인 기준과 방안을 제시한다.

* 데이터 위협, 모델 위협, 에이전트 위협, 공급망 위협, 고성능 모델 위협 등으로 분류

** 금융, 의료, 공공·행정, 교육, 제조·에너지, 통신, 법률, IT 등 8개 도메인

「AI 보안 레드티밍 가이드」는 AI 보안 레드팀을 운영하고자 하는 실무진을 위한 가이드로, 여러 산업을 아우르는 AI 레드팀 운영 방법론을 체계적으로 정리한 정부 발간 AI 보안 레드팀 운영 기준서이다. 동 가이드는 ▲레드팀 기획 및 구성부터, ▲레드티밍 준비, ▲레드티밍 이행, ▲결과 보고까지 AI 보안 레드팀 운영의 전 과정을 안내한다. 또한, 레드티밍 체크리스트와 점검 도구, 레드팀 인력 직무기술서 등을 추가하여 실용성을 제고하였다.

특히 이번 가이드는 민간 기업의 AI 보안 담당 실무자들의 의견을 폭넓게 수렴하고 레드티밍 전문기업 관계자의 의견을 비롯해 현장의 수요를 폭넓게 반영하여 마련하였으며, 학계 전문가의 자문을 통해 현장 활용성과 전문성을 높였다. 아울러 「AI 보안 레드티밍 가이드」는 AI 레드티밍 국제표준안(ISO/IEC 42119-7)을 기반으로 구성하여 국제적으로 통용되는 기준에 부합하도록 하였다.

최근 AI 에이전트의 확산으로 새로운 유형의 보안 위협이 증가하고, 선제적 보안 대응과 안전한 AI 서비스 운영에 대한 수요가 커지고 있는 만큼, 이번에 발간된 가이드 2종은 AI 서비스를 개발·운영하는 기업과 AI 레드티밍을 수행하는 전문기업 등에서 폭넓게 활용될 것으로 기대된다.

임정규 과기정통부 정보보호네트워크정책관은 “AI 기술의 확산과 함께 보안 위협도 빠르게 진화하고 있다.”며, “이번 가이드가 AI 서비스의 보안 수준을 높이고 현장에서 활용할 수 있는 실질적인 기준이 되기를 기대

한다.”고 밝혔다.

이어 “앞으로도 AI 보안 레드팀 사업을 통해 공격 시나리오와 적용 도메인을 지속적으로 확대·발굴하고, 최신 AI 기술과 동향을 반영해 가이드와 검증 체계를 고도화하여 AI 환경 변화에 선제적으로 대응해 나가겠다.”고 말했다.

『AI 보안위협 실무 매뉴얼』과 『AI 보안 레드티밍 가이드』는 ‘과기정통부 (<https://www.msit.go.kr>)’ 및 ‘한국인터넷진흥원(<https://www.kisa.or.kr>)’ 홈페이지를 통해 무료로 내려받아 활용할 수 있다.

담당 부서	정보보호네트워크정책관 정보보호기획과	책임자	과 장	지은경 (044-202-6440)
		담당자	사무관	김 인 (044-202-6441)



내일을 만드는 과학기술
내일을 채우는 디지털·AI

대한민국
지·책·브리핑

