SHIN&KIM
법무법인(유) 세종

과기정통부, 에이전틱 AI 이용 환경에서 제로트러스트를 적용할 수 있는 「자율형 인공지능 철통 인증 안내서」 발간

2026.11.11

과학기술정보통신부(이하 '과기정통부')와 한국인터넷진흥원(이하 'KISA')은 11월 6일, 국내 기업들이 에이전틱 AI(자율형 인공지능)를 도입·운영하는 환경에서 제로트러스트(철통 인증)를 적용할 수 있도록 하는 「자율형 인공지능 철통 인증 안내서」(이하 '안내서')를 발간하였다고 밝혔습니다. 안내서는 산·학·연 전문가들과 함께 국내외 주요 인공지능 보안·제로트러스트 가이드를 분석하고, 수요·공급기관을 대상으로 의견을 수렴하여 마련되었으며, '인공지능 시대를 지탱하는 견고한 디지털 보안·안전 체계 구축' 국정과제의 일환으로 추진되었습니다.

아래에서는 안내서의 발간 배경과 주요 내용 등에 대해 간략히 살펴보도록 하겠습니다.

1. 발간 배경

최근 클라우드 이용 확대, 인공지능 기반 업무 자동화 등으로 기업의 업무환경이 빠르게 변화하고 있습니다. 특히 인공지능이 단순한 정보 분석·생성을 넘어 스스로 판단하고 도구를 활용해 업무를 수행하는 에이전틱 AI로 발전하면서, 통제 대상도 기존의 사용자·단말·서버뿐 아니라 AI 에이전트가 접근하는 외부 서비스와의 상호작용, 물리적 장치에 대한 명령 등으로 확대되고 있습니다.

이에 정부는 기존 보안 체계로는 대응하기 어려운 보안 위협, 즉 에이전틱 AI의 자율성과 높은 권한에서 비롯되는 위협을 최소화하고자 안내서를 마련하였습니다. 안내서에는 프롬프트 오염, 위임된 권한을 넘어선 자율 행동 등 에이전틱 AI 고유의 위협을 제로트러스트 적용을 통해 예방·대응할 수 있는 보안 조치가 반영되어 있습니다.

※ 제로트러스트: 정보시스템 등에 대한 접속 요구가 있을 때 네트워크가 이미 침해된 것으로 간주하여 "절대 믿지 말고, 계속 검증하라(Never Trust, Always Verify)"는 보안 개념

2. 안내서의 주요 내용

(기존 가이드라인의 성숙도 모델 확장) 안내서는 기업의 제로트러스트 적용 절차와 도입 방향을 제시한 「제로트러스트 가이드라인 2.0」('24. 12.)을 토대로 합니다. 기존 가이드라인에서 정의한 '성숙도 모델'을 에이전틱 AI 이용 환경에 맞게 확장하고 구체적인 적용 기준을 제시하였습니다. 성숙도 모델은 기업이 현재 보안 수준을 진단·분석한 후 목표를 설정하고 검증하는 데 활용할 수 있도록 정의한 4단계 보안 분류체계입니다.

(AI 에이전트의 자율 행위 지점별 검증 및 인간 승인) AI 에이전트의 자율 행위(입력 처리, 도구 호출, 작업 위임, 외부 통신 등)가 이루어지는 여러 지점에서 해당 행위를 검증하고 정책을 집행하도록 하였습니다. 또한 위험 수준에 따라 인간의 승인이나 사전에 승인된 정책을 적용하도록 하였습니다.

(성숙도 단계별 실무 기준 및 도입·운영 절차) 조직이 현재 보안 수준을 자가 진단하고 우선순위에 따라 개선할 수 있도록 성숙도 단계별 실무 기준을 제시하였습니다. 아울러 에이전틱 AI 환경에서 제로트러스트를 도입할 때의 고려사항, 단계별 도입·운영 절차, 자체 적용이 가능한 활용 예시 등을 함께 담아, 기업이 도입부터 운영·개선 단계까지 안내서를 활용할 수 있도록 하였습니다.

안내서의 상세 내용은 KISA 누리집(www.kisa.or.kr)에서 확인할 수 있습니다. 과기정통부 임정규 정보보호네트워크정책관은 "기업 업무에 인공지능 활용이 증가하는 상황에서 자율형 인공지능에 철통 인증을 적용할 수 있는 안내서가 나온 것은 시의적절하다"라면서, "정부는 철통 인증 보안 모델 구축 사업을 수행하면서 철통 인증 적용 안내 등을 지속해서 발간함으로써 철통 인증의 민간 자율 확산 기반을 마련해 나가겠다"라고 밝혔습니다.

3. 시사점

(에이전틱 AI 도입 기업의 내부 통제 체계 점검 필요) 이번 안내서는 AI 에이전트의 자율 행위를 지점별로 검증하고 위험 수준에 따라 인간의 승인을 거치도록 하는 등, 에이전틱 AI 환경에 제로트러스트를 적용하기 위한 구체적인 기준을 처음으로 제시하였다는 점에서 의미가 있습니다. 에이전틱 AI를 도입·운영 중이거나 도입을 검토하는 기업은 안내서의 성숙도 단계별 실무 기준에 따라 자사의 보안 수준을 자가 진단해 볼 필요가 있습니다. 아울러 AI 에이전트에 부여된 권한의 범위, 외부 서비스 연동 및 도구 호출에 대한 통제, 인간 승인 절차 등 내부 통제 체계를 점검해 볼 필요가 있습니다.

(AI 보안 관련 안내서와의 종합적 활용 및 후속 동향 모니터링) 과기정통부와 KISA는 지난 7월에도 「AI 보안 위협 대응 매뉴얼」과 「AI 보안 레드티밍 가이드」를 발간하는 등 AI 보안 관련 실무 기준을 연이어 마련하고 있습니다. 기업으로서는 이번 안내서를 기존 안내서들과 함께 참고하여 AI 서비스 전반의 보안 체계를 정비해 나갈 필요가 있습니다. 또한 정부가 제로트러스트 적용 안내 등을 지속적으로 발간하겠다고 밝힌 만큼, 향후 후속 안내 등 관련 정책 동향도 지속적으로 모니터링할 필요가 있습니다.

About Shin & Kim's ICT Group

법무법인(유) 세종 ICT그룹은 ICT 분야에서 독보적인 전문성과 인적 네트워크를 보유하고 있으며, 고객들로부터 최근 수년간 가장 높은 평가를 받고 있습니다. 방송과 통신, 개인정보, 인터넷 IT 분야에서 축적된 역량을 바탕으로 방송·통신·ICT 규제 동향 파악 및 대관, 법제개선·입법컨설팅, 규제영향 분석과 기업의 전략 수립 등에 대한 종합적인 법률자문을 제공하고 있으며, AI Compliance, 정보보호 관리체계 구축, 침해사고 대응 등과 관련하여서도 다양한 업무경험과 전문성을 보유하고 있습니다. 보다 전문적인 내용이나 궁금하신 사항이 있으면 언제든지 연락 주시기 바랍니다.

관련구성원

강신욱
대표변호사

02-316-4059
sokang@shinkim.com

최광희
고문

02-316-4651
khchoi@shinkim.com

노진홍
변호사

02-316-1639
jhnoh@shinkim.com

강지현
변호사

02-316-1518
jhykang@shinkim.com

김가영
변호사

02-316-7910
gyekim@shinkim.com

Copyright SHIN & KIM LLC. All rights reserved.