2026·06

SK 주식회사 AX Ackerton Partners

AI 플랫폼 설계/구축

AI가 취약점을 찾는 시대, 기업 보안 전략도 바뀌어야 한다

글 | Ackerton Partners, IE본부 오법영 상무

01 · Project Glasswing이 보여준 전환점

02 · AI는 방어 도구이자 공격 가속기다

03 · 기업 보안 전략의 세 가지 변화

04 · 우리 조직에 주는 시사점

05 · 결론: AI 보안은 선택이 아니라 필수 전략이다

Executive Summary

Project Glasswing

Anthropic의 Project Glasswing은 Claude Mythos Preview를 방어자 중심으로 제한 제공하여 핵심 소프트웨어의 취약점 식별과 검증에 활용하는 글로벌 협력 이니셔티브다. AI가 보안 문서 작성이나 관제 보조를 넘어 취약점 탐지와 공격 가능성 검증 영역으로 확장되고 있음을 보여준다.

AI 기반 취약점 탐지

AI는 대규모 코드, 바이너리, 오픈소스 구성요소를 빠르게 분석하고 복합 취약점의 연쇄 가능성까지 점검하는 단계로 진화하고 있다. 이는 방어자에게는 탐지 범위와 속도를 확대하는 기회이지만, 공격자에게도 취약점 발견과 공격 준비의 진입장벽을 낮추는 양면성을 가진다.

속도 중심의 취약점 관리

AI 시대의 보안 경쟁은 "누가 취약점을 먼저 찾는가"에서 "누가 더 빠르게 검증하고 조치하는가"로 이동한다. 기업은 취약점 목록 관리에 그치지 않고 자산 중요도, 외부 노출도, 악용 가능성, 비즈니스 영향을 종합한 우선순위 기반 대응 체계를 갖춰야 한다.

소프트웨어 공급망 보안

현대 기업 시스템은 오픈소스, SaaS, API, 컨테이너 이미지, 클라우드 서비스 등 외부 구성요소와 깊게 연결되어 있다. AI 기반 취약점 탐지 확산은 SBOM, 오픈소스 취약점 관리, 서드파티 리스크 관리, 패치 거버넌스를 기업 보안의 핵심 과제로 부상시킨다.

DevSecOps와 실행 거버넌스

NIST SSDF와 CISA Secure by Design 관점처럼 보안은 개발 이후 점검하는 활동이 아니라 기획, 설계, 개발, 테스트, 배포, 운영 전 과정에 내재화되어야 한다. 보안 조직은 단순 탐지자가 아니라 개발·인프라·운영 조직의 실행을 조율하는 보안 PMO 역할로 확장되어야 한다.

01 Project Glasswing이 보여준 전환점

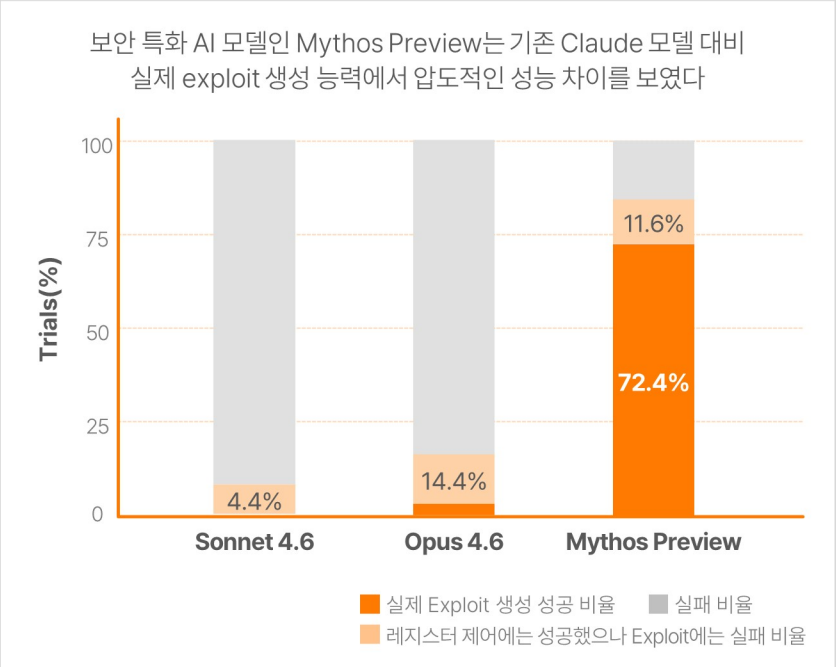
최근 글로벌 사이버보안 분야에서 중요한 변화가 나타나고 있다. AI가 단순히 보안 관제, 위협 인텔리전스 분석, 문서 자동화 등을 보조하는 수준을 넘어 핵심 소프트웨어의 취약점을 직접 탐지하고 공격 가능성까지 검증하는 단계로 진화하고 있기 때문이다.

[표 1] Antropic 취약점 발견 사례

사례	주요 내용
OpenBSD	27년 동안 잠재되어 있던 취약점을 발견, 원격 공격자가 운영체제를 실행 중인 시스템을 중단시킬 수 있는 가능성이 제시됨
FFmpeg	16년간 발견되지 않았던 취약점 발견, 자동화 테스트가 해당 코드 라인을 수백만 회 실행했음에도 탐지하지 못했던 문제로 설명됨
Linux Kernel	여러 취약점을 자율적으로 발견하고 연결하여 일반 사용자 권한에서 시스템 전체 제어 권한으로 상승할 수 있는 공격 가능성을 제시

Anthropic이 발표한 Project Glasswing은 이러한 변화를 상징적으로 보여주는 사례다. Project Glasswing은 Claude Mythos Preview라는 보안 특화 AI 모델을 방어자 중심으로 활용하기 위한 협력 이니셔티브이며, 참여 조직은 이를 핵심 시스템의 취약점 탐지와 보안 검증에 적용한다. Anthropic은 이 프로젝트를 통해 AI를 책임 있게 활용하여 보안 위험을 대규모로 줄일 수 있는 기회가 열렸다고 설명한다.

[그림 1] Firefox JS Shell exploitation



출처: https://red.anthropic.com/2026/mythos-preview/FRT-Blog-Chart-CMP-Firefox-exploit@2x.png

특히 주목할 점은 Claude Mythos Preview가 기존 자동화 테스트나 인간 전문가의 리뷰로는 찾기 어려웠던 복잡한 취약점을 탐지하고, 일부 경우에는 취약점의 연쇄 활용 가능성까지 제시했다는 점이다. Anthropic의 기술 공개 자료에 따르면 Mythos Preview는 Linux 커널에서 여러 취약점을 결합하여 권한 상승 가능성을 검증했고, 주요 브라우저와 암호화 라이브러리, 웹 애플리케이션 논리 취약점에 대해서도 강한 분석 역량을 보였다.

이는 기업 보안에 매우 중요한 의미를 갖는다. **지금까지 취약점 진단과 모의해킹은 전문 인력, 시간, 예산이 많이 필요한 영역이었다. 그러나 AI 기반 보안 모델이 발전하면서 취약점 탐지, 악용 가능성 검증, 패치 우선순위 판단의 속도가 크게 빨라질 가능성이 높아지고 있다.**

	기존 보안 진단	AI 기반 보안 진단
진단 범위	정해진 대상과 체크리스트 중심	**대규모 코드·바이너리·오픈소스 구성요소까지 확장**
검증 방식	도구 결과와 전문가 수동 검증 중심	**취약점 후보 탐지, 악용 가능성 추론, PoC 검증 보조**
운영 의미	정기 점검·프로젝트성 진단	**상시적 취약점 탐지와 우선순위 기반 조치 체계**

02 AI는 방어 도구이자 공격 가속기다

2.1 방어자에게 주는 기회

AI 기반 취약점 탐지 기술은 방어자에게 강력한 기회를 제공한다. 보안팀은 제한된 인력으로 더 많은 코드, 시스템, 바이너리, 오픈소스 구성요소를 점검할 수 있고, 기존 도구가 놓친 복합 취약점이나 논리 취약점까지 분석 범위를 넓힐 수 있다. 대규모 소프트웨어 공급망을 운영하는 기업, 클라우드 기반 서비스를 제공하는 기업, AI·디지털 플랫폼을 확대하는 기업에는 특히 중요한 변화다.

기존 취약점 관리 체계가 "스캔 결과 수집"에 머물렀다면, AI 기반 접근은 취약점 후보의 맥락을 이해하고 실제 악용 가능성을 판단하는 방향으로 확장될 수 있다. 예를 들어 외부에 노출된 자산, 중요 데이터와 연결된 서비스, 운영 중단 영향이 큰 시스템은 동일한 취약점이라도 우선순위를 다르게 가져가야 한다.

2.2 공격자에게 주는 위험

동시에 AI는 공격자에게도 취약점 발견과 공격 준비의 속도를 높이는 도구가 될 수 있다. 취약점 발견부터 공격 시나리오 구성, PoC 작성, 우회 기법 탐색까지 걸리는 시간이 줄어들 경우 기업은 기존의 월간·분기 단위 점검 체계만으로는 대응이 어려워질 수 있다.

CrowdStrike는 Project Glasswing 참여 배경에서 AI로 인해 취약점 발견과 악용 사이의 시간이 급격히 짧아지고 있으며, 방어자 역시 같은 속도로 움직여야 한다는 점을 강조했다. 결국 AI 시대의 보안 경쟁은 "누가 취약점을 먼저 발견하는가"뿐 아니라 "누가 더 빠르게 검증하고 조치하는가"의 경쟁으로 전환되고 있다.

03 기업 보안 전략의 세 가지 변화

[그림 2] 기업 보안 전략의 세가지 변화

취약점 관리 체계의 속도 경쟁	공급망 보안 강화	Shift Left와 DevSecOps 내재화
목록화에서 우선순위 기반 End-to-End 운영으로 전환	**오픈소스, 클라우드, SaaS, API까지 관리 범위 확대**	**개발 이후 점검이 아니라 SDLC 전 과정에 보안 통합**
• 자산 중요도, 외부 노출 여부, 공격 가능성,비즈니스 영향도를 기준으로 우선순위 설정 • 취약점 스캔 이후 담당 조직 지정, 조치 기한,예외 승인, 보완통제, 경영진 보고까지 연결	• 자체 개발 코드만이 아니라 오픈소스 라이브러리,외부 솔루션, 컨테이너 이미지까지 관리 대상화 • AI 확산으로 공통 라이브러리의 취약점이 더 빠르게 발견·악용될 가능성 증가	• 기획, 설계, 개발, 테스트, 배포, 운영 전 단계에 보안 요구사항과 자동화 도구를 삽입 • 개발 조직이 실제로 보안 조치를 수행할 수 있도록 프로세스와 책임을 DevSecOps에 내재화

3.1 취약점 관리 체계의 속도 경쟁

첫째, 취약점 관리 체계는 속도 중심으로 재설계되어야 한다. AI가 취약점 탐지 시간을 단축시키면 공격자와 방어자 간 시간 격차도 줄어든다. 기업은 취약점 스캔 결과를 단순히 목록화하는 수준을 넘어 자산 중요도, 외부 노출 여부, 공격 가능성, 비즈니스 영향도를 기준으로 우선순위를 정하고 신속하게 조치해야 한다.

이를 위해서는 취약점 관리 프로세스가 자산관리, 형상관리, 패치관리, 변경관리, 사고대응 프로세스와 연결되어야 한다. 취약점이 발견되었을 때 담당 조직 지정, 조치 기한 설정, 예외 승인, 보완통제 적용, 경영진 보고까지 이어지는 End-to-End 운영 체계가 필요하다.

3.2 오픈소스 및 소프트웨어 공급망 보안 강화

둘째, 오픈소스 및 소프트웨어 공급망 보안의 중요성이 더욱 커진다. 현대 기업 시스템은 자체 개발 코드만으로 구성되지 않는다. 오픈소스 라이브러리, 외부 솔루션, 클라우드 서비스, SaaS, API, 컨테이너 이미지 등 다양한 외부 구성요소가 복합적으로 연결되어 있다.

더 많은 내용을 보시려면

파일 다운받기

버튼을 눌러주세요