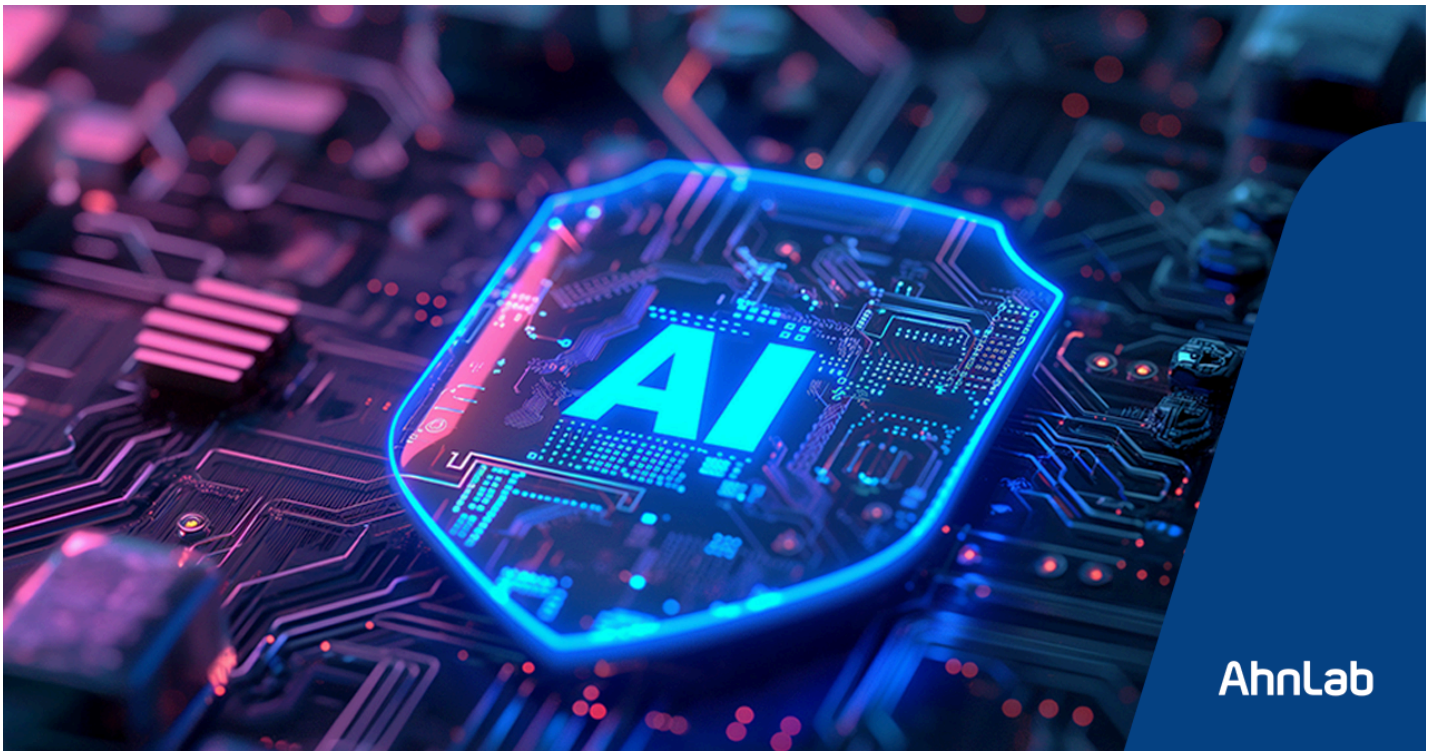
미토스(Mythos)란 무엇인가: AI 공격의 본질과 기업 보안 대응 전략

최근 글로벌 보안 업계에서는 앤트로픽(Anthropic)이 발표한 클로드 미토스 프리뷰(Claude Mythos Preview)와 프로젝트 글래스윙(Project Glasswing)이 주요 이슈로 부상하고 있다. 일부에서는 이를 “AI 기반 새로운 공격 기술의 등장”으로 해석하기도 하지만, 보다 정확한 관점은 다르다. 이번 사례는 새로운 공격 기법의 출현이라기보다, AI를 통해 취약점 탐색, 악용(Exploit), 공격 시나리오 수행의 속도와 자동화 수준이 크게 향상될 수 있음을 보여준 사건에 가깝다.

즉, 공격의 본질이 바뀐 것이 아니라, 공격이 이루어지는 방식의 ‘속도와 규모’가 달라지고 있다.



Mythos와 Glasswing은 무엇인가

앤트로픽(Anthropic)은 2026년 4월 7일, 자사의 최신 프론티어 모델인 클로드 미토스 프리뷰(Claude Mythos Preview)와 함께 프로젝트 글래스윙(Project Glasswing)을 발표했다.

클로드 미토스 프리뷰(Claude Mythos Preview)

- 대규모 언어모델(LLM, Large Language Model) 기반 보안 분석 모델
- 취약점 탐색(Vulnerability Discovery) 및 익스플로잇 생성(Exploit Generation)에 특화
- 제한된 환경에서 높은 수준의 자동화된 공격 시나리오 생성 가능성 확인

특히 이 모델은 단순 취약점 식별을 넘어, 실제 동작 가능한 공격 코드(Working Exploit)를 자동으로 생성하는 능력을 보여주며 주목을 받았다.

프로젝트 글래스윙(Project Glasswing)

- AWS, 마이크로소프트, 구글, 시스코, 크라우드스트라이크 등 주요 기업이 참여한 협력 프로그램
- 핵심 소프트웨어 및 인프라의 취약점을 조기에 발견하고 대응하기 위한 보안 연구 프로젝트
- 일반 공개가 아닌 제한된 파트너 기반 운영

즉, 이 프로젝트는 공격 도구의 확산이 아니라 AI 기반 취약점 발견 역량을 방어 목적으로 활용하기 위한 시도라고 보는 것이 적절하다.

공격은 그대로, 속도와 자동화만 달라졌다

이번 이슈를 이해하기 위해 중요한 점은 다음이다. 현재까지 확인된 바에 따르면 AI 기반 공격 역시 다음과 같은 기존 공격 흐름을 그대로 따른다.

- 취약점 악용(Exploit)
- 악성코드 실행(Malware Execution)
- 권한 상승(Privilege Escalation)
- 내부 확산(Lateral Movement)

즉, 공격의 '형태'는 변하지 않았다. 다만 AI의 도입으로 인해 다음과 같은 변화가 나타나고 있다.

- 취약점 탐색 속도의 급격한 증가
- 반복 공격의 자동화
- 공격 시나리오의 정교화
- 공격 진입 장벽의 낮아짐

가트너(Gartner)는 이를 취약점 발견과 악용 사이의 시간 간격이 급격히 줄어드는 구조적 변화로 설명하며, 베인앤컴퍼니(Bain & Company) 역시 AI 기반 공격이 이미 현실화되었음을 지적하고 있다.

왜 '완전 차단' 전략은 한계에 도달했는가

전통적인 보안 전략은 '사전 차단(Prevention)'을 중심으로 설계되어 왔다. 그러나 공격의 속도와 자동화 수준이 높아지면서, 이 접근 방식은 점점 한계를 드러내고 있다.

대표적인 변화는 다음과 같다.

- 첫째, 취약점이 공개된 이후 패치(Patch)가 완료되기 전에 공격이 발생하는 사례가 증가하고 있다.
- 둘째, 공격이 자동화되면서 동일한 공격이 빠르게 반복되고 확산된다.
- 셋째, AI를 통해 발견되는 취약점의 수 자체가 크게 증가하고 있다.

이러한 변화는 보안의 초점을 근본적으로 바꾸고 있다.

보안의 기준은 '노출 관리(Exposure Management)'로 이동하고 있다

이제 보안의 핵심 질문은 달라지고 있다. 과거에는 "얼마나 빠르게 패치하는가(MTTR, Mean Time To Remediate)"가 중요했다.

하지만 현재는 ▶취약점이 노출된 시간(노출 시간, Exposure Window) ▶얼마나 빠르게 탐지하는가 ▶얼마나 신속히 대응하는가가 더 중요한 기준이 되고 있다.

가트너(Gartner)는 이를 노출 관리(Exposure Management) 중심의 보안으로 정의하며, 보안의 KPI 자체가 변화하고 있음을 강조한다.

안랩이 보는 변화의 본질: 방어 방식의 전환

안랩은 이번 미토스(Mythos) 이슈를 단순한 위기로 보지 않는다. 오히려 이는 보안 운영 방식이 전환되는 시점으로 해석한다. 특히 주목해야 할 변화는 다음과 같다.

1. **사람 중심 보안 운영의 한계:** 모든 경보를 사람이 직접 분석하고 대응하는 구조는 점점 비효율적이다.
2. **AI 기반 보안 운영 필요성:** 탐지(Detection), 분석(Analysis), 대응(Response) 전 과정에 AI를 활용하는 구조가 요구된다.
3. **AI를 행위 주체(Actor)로 인식:** AI는 단순 도구가 아니라 하나의 행위 주체로서 그 자체를 식별하고 통제해야 하는 대상이 된다.

즉, 앞으로의 보안은 ▶AI를 활용한 방어 고도화 ▶AI 자체를 통제하는 보안이 동시에 요구된다.

AI 공격 시대, 기업이 준비해야 할 것

이러한 변화 속에서 기업이 준비해야 할 방향은 비교적 명확하다. 자산 가시성(Asset Visibility)을 확보해야 하며, 취약점 관리(Vulnerability Management) 체계를 강화해야 한다.

또한 행위 기반 탐지(Behavior-based Detection)를 통해 이상 징후를 빠르게 식별하고, 확장 탐지 및 대응(XDR, Extended Detection and Response)을 통해 통합 분석과 대응을 수행해야 한다. 여기에 대응 자동화(Response Automation)와 위협 인텔리전스(Threat Intelligence) 연계를 통해 실제 대응 속도를 높이는 것이 중요하다.

안랩의 대응 방향: 연결된 보안 운영

안랩은 이러한 환경 변화에 대응하기 위해 탐지·분석·통제·대응을 연결하는 보안 운영 체계를 중심으로 전략을 발전시키고 있다.

- **AhnLab EDR:** 자체 행위기반 분석 엔진과 사용자 정의 탐지, MITRE ATT&CK 기반 상세 위협정보, MDR 통한 전문가 분석 및 가이드 제공
- **AhnLab XDR:** 다양한 시스템에서 위협정보를 수집해 분석·탐지·대응하는 보안 운영 플랫폼으로 자산과 리스크를 식별하고 종합적인 대응으로 연결
- **AhnLab TIP:** 악성코드, 침해사고, 위협행위자, 취약점, IoC, 보안 뉴스 등 다양한 위협정보를 빠르게 제공하는 위협 인텔리전스 플랫폼으로 EDR/XDR 제품과 연계하여 최신 위협과 취약점에 대한 조직의 영향도를 빠르게 파악
- **AhnLab XTG의 ZTNA:** 제로트러스트 원칙에 따라 사용자와 기기의 신원, 보안 상태, 접속 조건 등을 지속 검증하고, 애플리케이션과 네트워크 자원 등의 접근을 통제하여 초기 침투와 이동 등 피해 확산 가능성 감소

이러한 구조는 특정 공격을 차단하는 것이 아니라, 탐지 → 분석 → 통제 ☒ 대응을 유기적으로 연결하는 운영 체계를 목표로 한다.

결론: 미토스(Mythos)가 보여준 보안의 전환

이번 미토스(Mythos) 이슈는 새로운 공격 방식의 등장이라기보다, 기존 보안 체계의 한계를 드러낸 사건에 가깝다. AI의 도입으로 공격은 “더 빠르게 이루어지고, 더 자동화되며, 더 많은 취약점을 동시에 활용할 수 있다”는 방향으로 변화하고 있다.

그러나 중요한 점은 공격의 본질 자체는 변하지 않았다는 것이다. 취약점 악용, 실행, 확산이라는 기존 공격 구조는 그대로 유지된다. 이러한 변화는 보안의 초점을 근본적으로 이동시키고 있다. 이제 기업은 “모든 공격을 차단할 수 있는가”보다 “얼마나 빠르게 탐지하고, 정확히 분석하며, 신속하게 대응할 수 있는가”에 집중해야 한다.

결국 보안의 경쟁력은 다음 세 가지로 수렴된다.

- 탐지 속도 (Detection Speed)
- 분석 정확도 (Analysis Accuracy)
- 대응 실행력 (Response Capability)

미토스가 남긴 가장 중요한 메시지는 명확하다. 보안은 더 이상 ‘차단의 기술’이 아니라, 탐지·분석·대응이 유기적으로 연결된 ‘운영 체계(Security Operating Model)’로 진화하고 있다.

FAQ: 미토스(Mythos)와 AI 기반 공격에 대한 주요 질문

Q1. Anthropic의 미토스(Mythos)/프로젝트 글래스윙(Project Glasswing)은 무엇을 의미하나

Anthropic은 2026년 4월 7일 클로드 미토스 프리뷰(Claude Mythos Preview)와 프로젝트 글래스윙(Project Glasswing)을 발표했다. 이는 일반에 공개된 상용 모델이라기보다, 일부 파트너와 참여 기관이 방어 목적의 보안 작업에 활용하는 제한된 연구 프로그램에 가깝다. 이 프로그램의 목적은 핵심 소프트웨어와 인프라에서 취약점을 더 빠르게 식별하고 보완하는 데 있으며, 앤트로픽 역시 해당 모델을 광범위하게 일반 제공할 계획은 없다고 밝히고 있다.

따라서 이번 이슈는 새로운 공격 도구의 확산이라기보다, AI를 활용해 취약점 탐색과 대응 속도를 높일 수 있음을 보여준 사례로 이해하는 것이 적절하다.

Q2. 미토스(Mythos)는 실제 위협인가, 과장된 이슈인가?

Mythos는 현재 제한된 환경에서 운영되는 연구 모델이다. 다만 AI 기반 공격의 가능성과 방향성을 보여준 사례로, 보안 전략 관점에서는 중요한 의미를 가진다.

Q3. 미토스(Mythos)와 같은 AI가 나오면, 기존 보안 제품은 무력화되는가?

보안의 기본 원리가 바뀐 것은 아니다. 다만 공격의 속도와 자동화 수준이 높아지면서, 자산 가시성, 취약점 관리, 행위 기반 탐지, 통합 분석, 대응 자동화의 중요성이 더욱 커지고 있다. 앤트로픽(Anthropic)도 미토스 프리뷰(Mythos Preview)를 일반 공개하지 않고, 방어 목적의 제한된 연구 프로그램으로 운영하고 있다.

Q4. AI 공격을 완전히 막을 수 있는가?

현실적으로 모든 공격을 완전히 차단하는 것은 어렵다. 중요한 것은 탐지·분석·대응의 속도를 높이는 것이다.

Q5. 안랩 제품은 AI 공격 대응에 도움이 되는가

AI 기반 공격도 결국 엔드포인트 행위, 네트워크 이동, 명령 실행 등의 형태로 나타난다. 안랩은 EDR, XDR, 위협 인텔리전스, 대응 자동화를 통해 대응 역량을 높이는 방향으로 지원한다.

Q6. 안랩이 공급하는 제품 자체는 AI 공격에 안전한가?

안랩은 제품 개발부터 릴리즈, 운영에 이르는 전 과정에서 공급망 보안(Supply Chain Security)과 취약점 대응 체계(Vulnerability Management)를 기반으로 제품 신뢰성을 관리하고 있다. 핵심은 AI가 찾았든 사람이 찾았든, 새롭게 발견되는 취약점을 얼마나 체계적이고 신속하게 식별·판단·조치할 수 있는가에 있다. 이를 위해 안랩은 글로벌 표준 기반의 오픈소스 관리(Open Source Management)를 비롯해 취약점 식별 및 영향도 분석, 반영 여부 검증과 지속 점검, 시큐어 코딩(Secure Coding), 정적·동적 분석(Static/Dynamic Analysis), 취약점 검사(Vulnerability Scanning), SBOM(Software Bill of Materials) 기반 구성 요소 관리까지 개발 전 과정에 보안 검증을 내재화하고 있다. 이를 통해 취약점 대응 리드타임을 최소화하고 있다.

또한 AI로 인해 취약점 발견과 악용 속도가 빨라지는 환경에서도 고객이 안심하고 제품을 운영할 수 있도록, 지속적인 보안 검증과 대응 체계 고도화를 이어가고 있다.

Q7. AI 시대 보안에서 가장 중요한 요소는 무엇인가?

차단 중심 접근보다 탐지·분석·대응을 유기적으로 연결하는 운영체계가 중요하다. 특히 대응 속도와 자동화 수준이 핵심이다.

Q8. 앞으로 보안은 어떻게 달라지는가

AI를 활용한 방어 체계 고도화와 함께, AI를 하나의 행위 주체로 보고 이를 식별·추적·통제하는 보안 영역이 함께 발전할 것으로 예상된다.

AhnLab