

No.02

WEF 보고서로 본 AI 사이버보안 회복력 전략

AI가 먼저 뚫어본다, 미토스 쇼크 그 이후!

기업은 어떻게 탐지하고, 막고, 회복해야 하는가

취약점은 숨어 있는 것이 아니라, 아직 AI가 찾지 않았을 뿐이다. AI는 취약점 탐색에 드는 시간과 비용을 낮추고, 피싱·악성코드·계정탈취 시나리오를 더 빠르게 변형시킨다. 그래서 기업의 목표는 완전한 무사고 선언이 아니라, **먼저 발견하고 피해 범위를 작게 가두며 실제로 복구되는 운영 능력**을 갖추는 것이다.

84

orgs · 15개 산업 커뮤니티 인사이트
인용 자료의 최신성 표기를
웹 발행본 그대로 옮겼다

WEF 「AI and Cybersecurity」
Strategic Intelligence Briefing
WEF × KPMG 「Empowering Defenders:
AI for Cybersecurity」 기반 인사이트 브리프

— 인쇄판 —

— 이번 호의 한 문장

뚫리지 않는 기업이 아니라, 먼저 발견하고 작게 가두고 빠르게 회복하는 기업이 살아남는다.

AI가 공격의 준비 시간을 압축하면서, 기업 보안의 질문이 바뀌었다. “우리 회사는 뚫리지 않을 수 있는가”에서 “AI가 먼저 취약점을 찾기 전에, 우리는 먼저 발견하고, 작게 막고, 빠르게 회복할 수 있는가”로. 이 브리프는 2026년 5월에 발행된 WEF 보고서 두 건 — 전략 브리핑 「AI and Cybersecurity」와 WEF×KPMG 백서 「Empowering Defenders」 — 을 기업 실행 관점에서 재구성했다. NIST CSF 2.0의 여섯 기능과 AI 도입 4단계, 백서에 실린 기업 방어 사례를 탐지·차단·대응·회복 네 개의 실행 축으로 다시 배치했다.

03	원전 지도 — 무엇을 읽고 썼는가	p.3
04	핵심 요약 — 다섯 개의 발견	p.4
05	문제의식 — 기업 보안의 질문이 바뀌었다	p.5
06	미토스 쇼크 — 사건이 아니라 전환점이다	p.6
07	핵심 인사이트 — 두 보고서에서 도출한 6가지 판단	p.7
08	위협 지형 — 기업이 직면한 세 가지 위협	p.8
09	새 기준 — 먼저 찾고, 작게 막고, 빠르게 회복하라	p.9
10	How to 01 · 탐지 — 어떻게 탐지할 것인가	p.10
11	How to 02 · 차단 — 어떻게 막을 것인가	p.11
12	How to 03 · 대응 — 어떻게 대응할 것인가	p.12
13	How to 04 · 회복 — 어떻게 회복할 것인가	p.13
14	도입 로드맵 · 90일 실행 계획	p.14
15	경영진 체크리스트 — 반드시 물어야 할 10가지 질문	p.15
16	앰버서더 Q&A — 이원태	p.16
17	닫는 글 — 회복하는 기업이 살아남는다 · 다음 호 예고	p.17
18	부록 — 용어 · 출처 · 발행 고지	p.18

— 이번 호의 원전

무엇을 읽고 썼는가

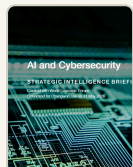
최신 Strategic Intelligence 보고서와 실행형 White Paper를 함께 읽고, 미토스 출현 이후 급부상한 AI 보안 이슈에 대해 “기업은 무엇을 해야 하는가”에 중점을 맞춰 재구성했다.

표지

확인

원전

이 보고서에서의 역할



생성
2026.05.11

AI and Cybersecurity
WEF · Strategic
Intelligence Briefing

AI 기반 공격과 방어의 동시 고도화, 미토스 쇼크, 회복력 격차, 거버넌스, 자율형 AI 방어를 다룬 전략 프레임 자료. 이 호의 문제의식과 위협 지형이 여기서 나온다.

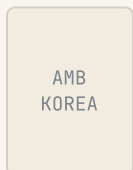
Generated: 11 May 2026 · Curated with
World Economic Forum



발행
2026.05.04

**Empowering Defenders:
AI for Cybersecurity**
White Paper · WEF × KPMG

NIST CSF 2.0, AI 도입 4단계, 20개 실제 기업 방어 사례와 CISO·경영진 실행 질문을 담은 실행형 자료. 이 호의 실행 축(탐지·차단·대응·회복)과 로드맵의 뼈대.



2026.06

앰버서더 Q&A — 이원태
전 KISA 원장 · 한국 현장
시각

글로벌 원전의 명제를 국내 정책·기업 실행의 언어로 옮기는 자리. 정부·민간의 패러다임 전환 논의, 경영진 거버넌스, 중소기업의 현실적 대응을 짚는다.

11 May 2026 · AI and Cybersecurity 생성일

4 May 2026 · WEF 백서 공식 발행일

84 orgs · 15개 산업 커뮤니티 인사이트

읽기의 방법 · METHODOLOGY

본 호는 위 두 보고서의 정독을 기본으로 하고, 인용 수치는 원전 표기를 그대로 옮겨 범위값(예: 30~40%↓)은 범위 그대로 표기했다. 개별 기업 사례의 수치는 해당 사례의 조건에서 관찰된 값이며 산업 평균이 아니다. 도식은 원문 프레임워크를 인쇄 가독성에 맞춰 다시 그렸다. 본 발행물은 인용 보고서의 발행 주체와 무관하며 후원·제휴를 의미하지 않는다.

— 핵심 요약

다섯 개의 발견

8415개 산업 커뮤니티
인사이트 참여 조직**20**백서에 실린
실제 기업 방어 사례**6**NIST CSF 2.0의
보안 기능**4**AI 도입을 검증하는
단계

A 미토스는 사건이 아니라 전환점이다

AI가 취약점을 찾는 순간, 방어자는 “알고 나서 막는” 방식에서 “먼저 찾아 줄이는” 방식으로 이동해야 한다. 제로데이는 더 이상 극소수 전문가만의 영역이 아니다.

B 보안은 차단에서 회복력으로 이동한다

침입 가능성을 전제로 피해 범위를 줄이고, 핵심 업무를 복구하는 능력이 경쟁력이 된다. 백업은 보유가 아니라 복구 테스트 통과로 본다.

C 가장 위험한 기업은 자기 취약점을 모르는 기업이다

외부 노출 자산, SaaS 권한, 클라우드 설정, 협력사 계정을 모르면 AI가 먼저 약점을 찾는다. 탐지는 기술 과제가 아니라 경영진의 가시성 과제다.

D 중소기업과 공급망이 약한 고리가 된다

대기업은 AI 방어에 투자하지만 중소기업과 협력사는 뒤처질 수 있다. 협력사 보안 격차는 원청·고객·산업 생태계 전체의 운영 리스크다.

E AI 보안은 기술 구매가 아니라 경영 의사결정이다

어디에 자동화를 허용하고, 어디서 사람이 승인하며, 사고 시 누가 멈추는지 정해야 한다. 핵심은 자동화가 아니라 통제 가능한 자동화다.

— EXECUTIVE MESSAGE

기업 보안의 질문이 바뀌었다

AI가 공격의 준비 시간을 압축하면서 기업은 “완전 차단”이라는 오래된 목표만으로는 충분하지 않다. 경영진이 확인해야 할 질문은 더 직접적이어야 한다.

OLD QUESTION

우리 회사는 뚫리지 않을 수 있는가?

이 질문은 여전히 중요하지만 충분하지 않다.

연결된 자산, 협력사 계정, 클라우드 설정, SaaS 권한까지 포함하면 완전한 폐쇄형 방어는 현실적인 목표가 아니다.

NEW QUESTION

AI가 먼저 취약점을 찾기 전에, 우리는 먼저 발견하고, 작게 막고, 빠르게 회복할 수 있는가?

탐지는 빠르게, 차단은 작게, 복구는 실제로 가능해야 한다. 이 질문에 답하지 못하면 보안 투자는 사고 이후의 설명 자료에 머물 수 있다.

공격자는 AI를 쓴다. 방어자가 수동 대응에 머물 수는 없다. 경영진은 보안팀에 더 많은 알림을 보라고 요구하는 대신, **어떤 알림을 먼저 보고, 어떤 권한을 즉시 줄이고, 어떤 업무를 먼저 복구할지 정하는 운영체계를 만들어야 한다.**

기업은 AI 보안 솔루션을 사는 것이 아니라, **AI 시대의 보안 운영체계를 만들어야 한다.**

FIRST

먼저 발견하라

외부 노출 자산과 고위험 취약점을 공격자보다 먼저 찾아야 한다.

SMALL

작게 가둬라

침입을 전제로 권한, 계정, 네트워크 이동 경로를 좁혀야 한다.

REAL

실제로 복구하라

백업 보유가 아니라 복구 테스트 통과가 회복력의 기준이다.

— MYTHOS SHOCK

미토스는 사건이 아니라 전환점이다

WEF Strategic Intelligence Briefing이 짚은 핵심은 AI가 공격자와 방어자 모두의 능력을 끌어올린다는 점이다. 문제는 공격자의 속도가 먼저 빨라질 수 있다는 데 있다.

01 취약점 발굴이 자동화된다

제로데이는 더 이상 극소수 전문가만의 영역이 아니다. AI는 코드, 설정, 노출 자산을 빠르게 훑고 악용 가능성이 높은 지점을 좁힌다.

공개된 코드 · 인터넷 노출 서비스 · 오래된 라이브러리 · 잘못 열린 관리 포트가 탐색 대상이 된다

02 공격 준비가 기계화된다

초개인화 피싱, 변형 악성코드, 계정탈취 시나리오가 더 빠르게 만들어진다. 방어자가 수동 목록으로 대응하는 동안 공격자는 실험을 반복한다.

수동 분류 체계는 사고 초기에 병목이 되고, 대응 시간이 곧 피해 규모가 된다

03 회복력 격차가 커진다

대기업은 AI 방어에 투자하지만 중소기업과 협력사는 뒤처질 수 있다. 중소기업의 보안 격차는 공급망 전체의 리스크가 된다.

협력사 계정 · 외주 개발 환경 · SaaS 공유 권한이 우회로가 된다

미토스 이후, 사이버보안은 속도의 전쟁이 됐다.

— STRATEGIC INSIGHTS

두 보고서에서 도출한 6가지 판단

기술의 유행보다 중요한 것은 보안 운영의 기준이 어떻게 바뀌는가다. 다음 여섯 가지 판단은 경영 회의의 의제로 올라와야 한다.

TURNING POINT

미토스는 사건이 아니라 전환점이다

AI가 취약점을 찾는 순간, 방어자는 “알고 나서 막는” 방식에서 “먼저 찾아 줄이는” 방식으로 이동해야 한다.

RESILIENCE

보안은 차단에서 회복력으로 이동한다

침입 가능성을 전제로 피해 범위를 줄이고, 핵심 업무를 복구하는 능력이 경쟁력이 된다.

VISIBILITY

가장 위험한 기업은 자기 취약점을 모르는 기업이다

외부 노출 자산, SaaS 권한, 클라우드 설정, 협력사 계정을 모르면 AI가 먼저 약점을 찾는다.

SUPPLY CHAIN

중소기업과 공급망이 약한 고리가 된다

협력사 보안 격차는 단일 기업 문제가 아니라 원청, 고객, 산업 생태계 전체의 운영 리스크다.

GOVERNANCE

AI 보안은 기술 구매가 아니라 경영 의사결정이다

어디에 자동화를 허용하고, 어디서 사람이 승인하며, 사고 시 누가 멈추는지 정해야 한다.

CONTROL

사람의 판단을 살리는 자동화가 필요하다

AI 보안의 핵심은 자동화가 아니라 통제 가능한 자동화다. 속도와 책임이 함께 설계되어야 한다.

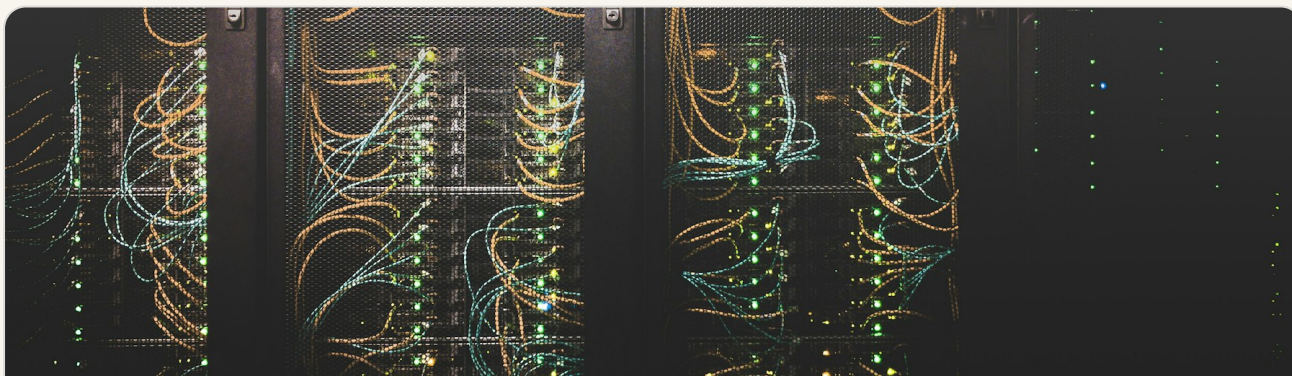
여섯 판단은 따로 놓인 항목이 아니다 — 먼저 보고(가시성), 작게 막고(회복력), 누가 멈추는지 정하는(거버넌스) 하나의 운영체제로 이어진다.

— THREAT LANDSCAPE

기업이 직면한 세 가지 위협

AI 기반 공격의 위험은 “새로운 해킹 도구” 하나로 설명되지 않는다. 더 큰 변화는 **취약점 발견, 공격 실행, 공급망 확산의 시간이 동시에 줄어든다**는 점이다.

구분	위협	무엇이 달라지는가	기업 영향
T1	취약점 발굴의 자동화	공개된 코드, 인터넷 노출 서비스, 오래된 라이브러리, 잘못 열린 관리 포트가 AI의 탐색 대상이 된다.	패치 우선순위를 놓치면 공격자가 먼저 사업 영향이 큰 취약점을 골라낸다.
T2	공격 속도의 기계화	피싱 문구, 악성코드 변형, 계정탈취 시나리오가 자동으로 실행된다. 보안관제는 알림 수보다 우선순위가 중요해진다.	수동 분류 체계는 사고 초기에 병목이 되고, 대응 시간이 곧 피해 규모가 된다.
T3	중소기업과 공급망의 약한 고리	협력사 계정, 외주 개발 환경, SaaS 공유 권한은 공격자가 우회로로 활용하기 쉽다.	한 회사의 보안 격차가 거래망 전체의 데이터 유출과 업무 중단으로 번진다.



세 위협의 공통점 — 새로운 도구가 아니라 시간이다. 발견·실행·확산에 걸리던 시간이 함께 줄어들면, 같은 취약점도 훨씬 큰 피해로 번진다.

한 협력사의 작은 계정 하나도 **공급망 전체의 업무 중단**으로 번질 수 있다.

— NEW SECURITY STANDARD

먼저 찾고, 작게 막고, 빠르게 회복하라

NIST Cybersecurity Framework 2.0의 여섯 기능은 AI 시대에도 유효하다. 다만 기업은 이를 문서가 아니라 운영 리듬으로 바꿔야 한다.

탐지는 빠르게, 차단은 작게, 복구는 실제로.

NIST CSF 2.0 · 여섯 기능을 운영 리듬으로

01 Govern	보안 책임, AI 사용 기준, 중단 권한을 경영 의사결정으로 둔다.
02 Identify	자산, 데이터, 권한, 공급망 의존성을 한 장의 지도처럼 파악한다.
03 Protect	권한 최소화, MFA, 설정 기준, 보안 개발 체계를 실행한다.
04 Detect	AI로 이상 징후와 악용 가능성이 큰 취약점을 먼저 찾는다.
05 Respond	우선순위, 차단 범위, 보고 체계를 플레이북으로 실행한다.
06 Recover	복구 목표, 백업 테스트, 사후 개선을 실제 훈련으로 검증한다.

공격을 완전히 막겠다는 선언보다 중요한 것은 침입 가능성을 전제로 피해 범위를 줄이고, 복구 가능성을 반복 검증하는 체계다.

Framework adapted from World Economic Forum, Empowering Defenders: AI for Cybersecurity, May 2026, and NIST Cybersecurity Framework 2.0.

01

DETECT

— HOW TO 01 · DETECT

어떻게 탐지할 것인가

AI보다 먼저 우리 취약점을 찾아라. 탐지는 보안팀의 기술 과제가 아니라 경영진의 가시성 과제다. 모르는 자산, 모르는 계정, 모르는 취약점이 가장 위험하다.

경영진 이 요구할 산출물

외부 공격 표면 지도, 고위험 취약점 Top 20, 소유자별 조치 기한.

1 외부 노출 자산부터 전수 파악한다

도메인, 클라우드, SaaS, VPN, 원격접속, 개발 서버를 목록화하고 소유자를 지정한다.

2 취약점은 개수가 아니라 악용 가능성으로 관리한다

CVSS 점수만이 아니라 인터넷 노출 여부, 권한 상승 가능성, 핵심 업무 영향도를 함께 본다.

3 개발 단계부터 AI로 보안을 심는다

코드 작성, 리뷰, 배포 전 설정 점검에 AI 탐지를 연결하고, 적용 전에는 인간 검토를 거친다.

CASE · Accenture - Agent Oliver

수십만 개 인터넷 노출 사이트를 사람이 하나씩 점검하던 문제를 AI 에이전트 기반 공격표면 점검으로 전환한 사례다.

문제 사이트당 15분 점검이면 전체 공격표면을 보려면 사실상 수년이 걸렸다. AI 적용 크롤러, 설정 점검, 취약점 탐색, 리포터 에이전트를 묶어 자동 점검했다.

10만+

인터넷 노출 사이트 점검

15분 → 1분↓

사이트당 분석 시간 단축

93%↓

수작업 감소

기업이 배울 점 보안 점검의 출발점은 도구가 아니라 “우리가 밖에 무엇을 열어두었는지”를 매일 갱신하는 체계다.

CASE · Google - Big Sleep / CodeMender

대규모 코드베이스에서 사람이 놓치기 쉬운 취약점을 AI가 먼저 찾고, 패치 후보까지 제안하되 인간 연구자가 검토하는 방식이다.

문제 수동 코드 리뷰만으로는 미지의 취약점을 충분히 빠르게 찾기 어렵다. AI 적용 Big Sleep은 취약점을 찾고 CodeMender는 보안 패치 후보를 만든다. 성과 CodeMender가 중요 보안 이슈 100개+ 패치, 적용은 인간 연구자 검토 후.

기업이 배울 점 AI 자동화는 “자동 적용”이 아니라 “빠른 발견 + 통제된 인간 검토”로 설계해야 한다.

경영진 확인 질문

우리 회사는 외부에 노출된 모든 자산과 그 소유자를 알고 있는가?

02

CONTAIN

— HOW TO 02 · CONTAIN

어떻게 막을 것인가

뚫리지 않는 보안보다 작게 막는 보안. 침입 가능성을 전제로 권한과 이동 경로를 줄여야 한다. 차단은 한 번의 벽이 아니라 여러 개의 작은 격리선이다.

경영진이 요구할 산출물

핵심 계정 MFA 현황, 관리자·외주 계정 목록, 권한 회수 완료표.

1 권한 최소화와 MFA를 핵심 계정부터 적용한다

관리자, 재무, 인사, 개발, 외주 계정부터 MFA와 조건부 접근을 강제한다.

2 외주 계정과 협력사 접근을 분리한다

공유 계정, 장기 미사용 계정, 프로젝트 종료 후 남은 권한을 우선 정리한다.

3 피싱과 데이터 유출은 AI로 조기 탐지한다

문장 패턴, 심리적 설득 요소, 비정상 전송 경로를 함께 본다.

사례	무엇을 바꿨는가	성과
Santander Phishing Detection	피싱이 문법 오류나 수상한 링크만으로 판별되지 않는 시대에, 문장의 의도와 심리적 설득 구조까지 분석했다. 다국어 LLM이 메시지의 의미, 브랜드 사칭, 설득 패턴을 함께 봤다. 배울 점 · 이메일 보안은 링크 차단을 넘어 “사람을 속이는 방식”을 탐지해야 한다.	10%+ 피싱 탐지 효과 항상 심리 분석 - 설득 요소까지 위험 신호 포착
ING Data Loss Prevention	대량의 데이터 유출 경고를 사람이 모두 볼 수 없을 때, AI가 데이터 흐름과 정책 위반 신호를 분석해 조사할 사건을 좁혔다. 배울 점 · DLP의 목표는 알람을 많이 만드는 것이 아니라, 실제 유출 가능성이 큰 사건을 먼저 보게 하는 것이다.	500만 알람 처리 20%↑ 분석 정확도 향상
AXIS Capital Secure Workflow	보안을 배포 직전 검사로 두지 않고, 개발자의 워크플로우 안에 코드와 클라우드 설정 점검을 삽입했다. 배울 점 · 보안은 개발을 멈추는 게 아니라, 개발 흐름 안에서 위험한 결정을 줄이는 방식이어야 한다.	Code - 코드와 클라우드 설정을 AI로 분석 Shift Left - 개발 흐름 안에 보안 조치 삽입

경영진 확인 질문

MFA가 적용되지 않은 핵심 계정과 외주 계정은 없는가?

03

RESPOND

— HOW TO 03 · RESPOND

어떻게 대응할 것인가

사람이 못 따라가는 속도는 AI로 보완하라. 보안관제의 핵심은 알림 개수가 아니라 우선순위다. AI는 분석가를 대체하는 것이 아니라, 분석가가 먼저 볼 사건을 정확하게 도와야 한다.

경영진이 요구할 산출물

1시간 대응 플레이북, 자동 차단 승인 기준, AI 판단 중지 절차.

1 알림을 줄이는 것이 아니라 우선순위를 정한다

업무 영향, 공격 단계, 계정 중요도, 확산 가능성을 기준으로 사건을 정렬한다.

2 1시간 안에 차단할 범위를 미리 정한다

계정 잠금, 네트워크 격리, 토큰 폐기, 외부 공유 중단 등 조치를 플레이북으로 둔다.

3 AI 판단 오류를 멈출 절차를 둔다

자동 차단의 승인 기준과 예외 처리 책임자를 사전에 지정한다.

사례	무엇을 바꿨는가	성과
IBM ATOM	전 세계 보안관제에서 반복되는 조사 업무를 AI가 처리해, 분석가가 더 위험한 사건에 집중하도록 만들었다. 위협 정보 수집, 사건 요약, 조사 보조를 자동화해 24시간 대응 체계를 보강했다.	95% 일일 위협 조사 처리 37%↓ 조사 시간 850h+ 월간 분석가 시간 절감
PETRONAS Incident Response	분석가 워크플로우 안에 AI 요약, 다음 조치 안내, 자연어 쿼리 변환을 넣어 사고 대응 품질을 표준화했다. 주니어 분석가도 빠르게 판단하도록 지원했다.	30~40%↓ 사고 대응·해결 시간 50%↑ 신규 분석가 온보딩
Standard Chartered Classification	사건 유형과 위험 신호를 분류해 분석가의 초기 판단을 보조했다. 고위험 사건의 조사 착수가 빨라졌다.	25~35%↓ 수동 분류 업무 20~30%↑ 분류 시간 개선
Dream Group Malware Guidance	악성코드 분석 후 조치 가이드를 작성하는 시간을 줄여, 현장 대응자가 바로 실행할 수 있는 지침을 빠르게 만들었다.	95%↓ 악성코드 조치 가이드 작성 시간

기업이 배울 점 · AI 관제의 목적은 사람을 줄이는 것이 아니라 사람이 먼저 볼 사건을 더 빨리 정하는 것이다 · 대응 속도는 “누구나 같은 품질로 대응하는 표준화”에서 나온다 · 대응 문서는 사후 기록이 아니라 사고 중 바로 실행되는 작전 지시서여야 한다.

경영진 확인 질문

사고 발생 시 1시간 안에 무엇을 차단하고 누구에게 보고할지 정해져 있는가?

04

RECOVER

— HOW TO 04 · RECOVER

어떻게 회복할 것인가

복구 없는 탐지는 경보음에 불과하다. 사고 대응의 끝은 보고서 작성이 아니다. 핵심 업무가 실제로 돌아오고, 같은 경로가 다시 열리지 않도록 운영을 고치는 것이다.

경영진이 요구할 산출물

업무별 RTO/RPO, 복구 테스트 결과, 랜섬웨어 훈련 개선 목록.

1 핵심 업무별 RTO/RPO를 정한다

업무별 허용 중단 시간과 데이터 손실 허용 범위를 경영진이 승인한다.

2 백업은 보유가 아니라 복구 테스트 통과로 본다

랜섬웨어 상황을 가정해 격리 백업과 실제 복구 시간을 검증한다.

3 협력사 장애가 우리 회사에 미치는 영향을 분석한다

대체 공급, 데이터 접근 중단, 계정 폐기, 임시 업무 절차를 사전에 정한다.

4 사고 후 설정과 권한, 탐지 룰을 개선한다

사고 원인을 문서화하는 데서 끝내지 않고 동일 경로를 닫는 변경을 배포한다.

네 개 모드 · 이번 주 만들 산출물 한 장

탐지 먼저	외부 공격 표면 지도 — 도메인, 클라우드, SaaS, 원격접속, 개발 서버를 한 표로 묶고 책임자를 지정한다.
피해 작게	권한·계정 통제표 — 관리자·재무·개발자·외주 계정부터 MFA를 적용하고, 프로젝트 종료 후 남은 권한과 공유 계정을 제거한다.
대응 빠르게	1시간 대응 플레이북 — 계정 잠금, 토큰 폐기, 네트워크 격리 절차를 사전 승인하고, AI 판단이 틀렸을 때 사람이 멈출 수 있는 절차를 둔다.
복구 실제로	복구 테스트 결과표 — 격리 백업을 실제로 복원해 복구 시간을 측정하고, 사고 후 권한·설정·탐지 룰을 고치는 개선 루프를 운영한다.

경영진 확인 질문

백업은 실제 복구 테스트를 통과했는가?

— AI ADOPTION ROADMAP

AI 보안 도입은 네 단계로 검증한다

WEF 백서의 AI 도입 경로는 기술 구매 순서가 아니라 신뢰 가능한 운영체계를 만드는 순서다.

01 ALIGN	02 READY	03 VALIDATE	04 SCALE
맞춘다	준비한다	검증한다	확장·유지한다
어떤 보안 문제를 AI로 풀 것인지 경영 목표와 맞춘다.	데이터, 권한, 로그, 정책, 보안팀 역량을 준비한다.	고위험 영역 하나에서 정확도와 통제 가능성을 검증한다.	검증된 것만 확장하고 정기적으로 성능과 위험을 재평가한다.
우선순위 업무 선정 · 책임자와 승인 기준 정의	로그 품질 점검 · 접근 권한 정리	파일럿과 인간 검토 · 오탐·미탐 대응 절차	운영 지표 반영 · 감사와 개선 루프

Roadmap adapted from World Economic Forum, Empowering Defenders: AI for Cybersecurity, May 2026.

90일 안에 무엇을 바꿀 것인가

AI 보안은 장기 전환 과제지만, 첫 90일에는 보이는 위험을 줄이고 검증 가능한 파일럿을 만드는 데 집중해야 한다.

90-DAY ACTION PLAN	
0~30일	먼저 보이는 것부터 정리한다 — 외부 노출 자산과 소유자 목록화 · 핵심 계정 MFA 적용 현황 점검 · 백업 복구 테스트 계획 수립
31~60일	고위험 영역 하나를 골라 AI 파일럿을 한다 — 피싱, 취약점, 보안관제 중 하나 선택 · 오탐·미탐과 인간 검토 기준 설정 · 경영진 보고 지표 정의
61~90일	검증된 것만 확장한다 — 효과가 확인된 탐지 룰과 플레이북 배포 · 협력사 계정과 외주 접근 기준 강화 · 복구 훈련과 사후 개선 루프 실행

— EXECUTIVE CHECKLIST

경영진이 반드시 물어야 할 10가지 질문

보안 회의에서 기술명보다 먼저 확인해야 할 질문이다. 답이 모호한 항목은 즉시 90일 계획에 넣어야 한다.

AI 보안 회복력 자가진단 · 10문항

점검일 20 . . .

☐

01. 우리 회사의 외부 노출 자산을 모두 알고 있는가?

도메인·클라우드·SaaS·VPN·원격접속·개발 서버와 소유자 — 탐지의 출발점

탐지

☐

02. 가장 중요한 데이터와 시스템은 무엇인가?

핵심 업무를 먼저 정의하지 않으면 복구 우선순위도 정할 수 없다

식별

☐

03. 그 시스템에 누가 접근할 수 있는가?

권한 지도가 없으면 침입 이후 이동 경로도 좁힐 수 없다

식별

☐

04. MFA가 적용되지 않은 핵심 계정은 없는가?

관리자·재무·인사·개발·외주 계정부터 — 차단 of 최소 조건

차단

☐

05. 협력사와 외주 계정은 어떻게 관리되는가?

공유 계정, 장기 미사용 계정, 프로젝트 종료 후 남은 권한

공급망

☐

06. 취약점은 사업 영향 기준으로 우선순위화되고 있는가?

개수가 아니라 악용 가능성 — 노출 여부·권한 상승·업무 영향

탐지

☐

07. 보안 알림은 사람이 모두 보는가, AI가 우선순위를 정하는가?

수동 분류는 사고 초기의 병목이 된다

대응

☐

08. 사고 발생 시 1시간 안에 무엇을 차단하고 누구에게 보고할지 정해져 있는가?

계정 잠금·네트워크 격리·토큰 폐기·외부 공유 중단의 사전 승인

대응

☐

09. 백업은 실제 복구 테스트를 통과했는가?

보유가 아니라 통과 — 격리 백업의 복구 시간을 측정했는가

회복

☐

10. AI가 잘못 판단했을 때 멈출 수 있는 절차가 있는가?

자동 차단 승인 기준과 예외 처리 책임자를 사전 지정했는가

거버넌스

쓰는 법 · HOW TO USE

열 문항 중 “그렇다”가 나오지 않는 항목이 곧 다음 90일의 안건이다. 각 문항에 담당과 기한을 적어 두 주 뒤 같은 표로 다시 점검한다.

— AMBASSADOR Q&A

앰버서더에게 묻는 AI 보안의 다음 질문



이원태

AMBASSADOR

전 한국인터넷진흥원(KISA) 원장
국가AI전략위원회
보안특별위원회 위원장

정보통신정책연구원(KISDI) 연구위원을 거쳐 제6대(2021~2023) 한국인터넷진흥원 원장을 역임했다. 현재는 국가인공지능전략위원회 보안특별위원회를 이끌며 대한민국의 AI 전략 및 사이버보안 정책 수립에 참여하고 있다.

Q 미토스 등 AI 해킹 가능성이 대두되면서 정부에서도 많은 논의가 이루어지고 있다. 합의된 액션 플랜이 있나?

A 현재 정부와 민간에서는 “AI가 먼저 취약점을 찾는 시대”라는 문제의식 아래, 단순 차단 중심이 아니라 “먼저 탐지하고, 작게 막고, 빠르게 회복하는” 방향으로 보안 패러다임 전환 논의가 진행 중이다. 특히 AI 기반 취약점 탐지, 우선순위화, 대응 자동화, AI-ISAC 형태의 위협정보 공유, 그리고 AI Runtime Security 및 VulOps 같은 새로운 대응체계 구축 필요성에 공감대가 형성되고 있다.

Q 디지털 정보보안은 기술 솔루션이나 보안팀이 알아서 할 문제라고 생각할 수 있다. 그런 경영진에게 충고한다면?

A AI 시대의 사이버보안은 더 이상 IT부서만의 기술 문제가 아니라, CEO·CFO·COO·CISO가 함께 관리해야 하는 경영 리스크이자 기업 생존 인프라에 가까워지고 있다. 특히 AI 기반 공격은 공급망, 외주, 클라우드, 계정 전체를 따라 확산되기 때문에, 경영진이 직접 데이터·권한·협력사·복구 체계를 점검하는 거버넌스 관점 접근이 더욱 중요해지고 있다.

Q 해킹의 완벽한 차단이 불가능한 상황에서 신속하고 온전한 복구와 회복이 중요한 과제가 되고 있다. 비교적 영세한 중소기업에게 권고할 수 있는 현실적인 대응 방안은 무엇일까?

A 중소기업은 “완벽 방어”보다 외부 노출 자산 파악, MFA 적용, 백업·복구 테스트, 협력사 계정 관리 같은 기본기를 먼저 갖추는 것이 가장 현실적이다. 또한 KISA 보호나라, VirusTotal, 무료 LLM 기반 피싱 분석 등 공개된 무료 및 저비용 AI 보안 도구를 적극 활용하고, 작은 AI 보안 PoC부터 단계적으로 시작하는 접근이 중요하다. 아울러 AI 보안 투자 여력이 부족한 중소·영세기업들을 위해 AI-SecaaS(Security-as-a-Service) 기반 공동 대응 체계와 AI 보안 서비스 지원 정책도 정부 차원에서 함께 검토하고 활용할 필요가 있다.

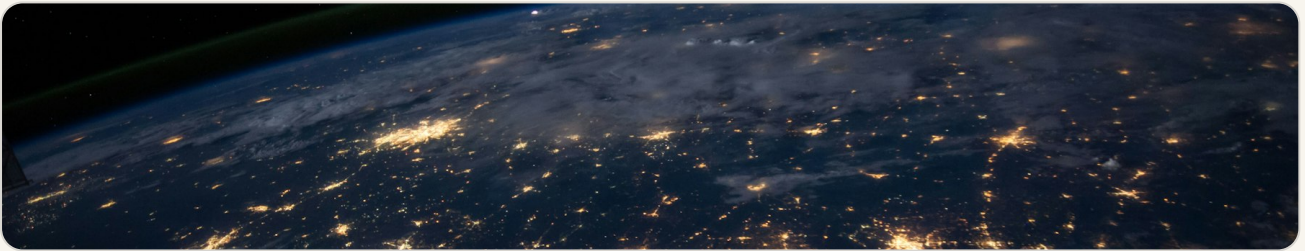
세 답이 가리키는 곳은 같다 — 패러다임은 탐지·격리·회복으로, 책임은 경영진으로, 중소기업은 기본기와 공동 대응으로. 앞선 10가지 질문(p.15)이 그 진단의 첫 장이다.

CLOSING · 닫는 글

뚫리지 않는 기업보다 회복하는 기업이 살아남는다

AI 시대의 사이버보안은 더 많은 경고음을 만드는 일이 아니다. 먼저 발견하고, 피해를 작게 가두고, 실제로 복구되는 운영 능력을 만드는 일이다.

**뚫리지 않는 기업이 아니라,
AI보다 먼저 발견하고, 피해를 작게 가두고,
빠르게 회복하는 기업이 살아남는다.**



한 협력사의 작은 계정 하나도 공급망 전체로 번진다 — 회복력은 연결의 문제다

NEXT ISSUE · 다음 호 예고

GPU 이후의 승부: 누가 한국 제조 데이터의 운영권을 가질 것인가

국내 반도체·완성차 대기업의 AI공장 투자가 던진 진짜 질문. GPU를 사는 일과 제조 데이터의 운영권을 갖는 일은 다르다 — 커스텀 브리프 3호에서 이어간다.



— 부록

용어 · 출처 · 발행 고지

이 호의 용어 · GLOSSARY

미토스 쇼크

AI가 취약점을 먼저 찾아내는 시대를 상징하는 사건. 본 호는 사건이 아니라 전환점으로 읽는다.

제로데이

아직 패치가 나오지 않은 미지의 취약점. 더 이상 극소수 전문가만의 영역이 아니다.

공격 표면

외부에서 접근 가능한 자산의 총합. 도메인·클라우드·SaaS·VPN·원격접속·개발 서버.

NIST CSF 2.0

Govern·Identify·Protect·Detect·Respond·Recover 여섯 기능으로 구성된 사이버보안 프레임워크.

MFA

다중 인증. 비밀번호 외 추가 수단으로 본인을 확인한다. 핵심 계정 차단인 최소 조건.

RTO · RPO

업무별 허용 중단 시간과 데이터 손실 허용 범위. 경영진이 승인하는 복구 목표값.

DLP

데이터 유출 방지. 유출 가능성이 큰 사건을 먼저 보게 하는 것이 목표다.

AI-ISAC · AI-SecaaS

AI 위협정보 공유 협의체, 그리고 보안을 서비스로 제공해 중소기업이 공동 대응하는 방식.

출처 · REFERENCES

World Economic Forum, AI and Cybersecurity, Strategic Intelligence Briefing, generated 11 May 2026.

World Economic Forum, Empowering Defenders: AI for Cybersecurity, White Paper, May 2026, in collaboration with KPMG.

World Economic Forum, New Report Shows How AI Gives Cybersecurity Competitive Advantage, press release, 4 May 2026.

본문의 사례·수치·프레임워크는 위 보고서를 기반으로 재구성했고, 도식은 인쇄 가독성에 맞춰 다시 작성했다.

C4IR KOREA · CUSTOM BRIEF No.02 — 인쇄판

방법론 · 본 호는 WEF 최신 공개 보고서 정독과 발행일 기준 글로벌 신호 교차 확인으로 제작되었습니다. WEF Strategic Intelligence 교차 확인 포함.

검증 · 본문의 모든 수치·인용·사실 관계는 1차 출처(논문 원문·공시·발표 기관 공식 자료)로 교차 확인했으며, 발행 전 독립 검수(레드팀) 심사에서 제기된 사항을 반영했습니다. 1차 출처로 확인되지 않은 수치와 주장은 실지 않았습니다.

고지 · 본 발행물은 인용 보고서의 발행 주체와 무관하며, 후원·제휴를 의미하지 않습니다.

※ 본 발행물은 특정 기업·제품에 대한 소개, 추천, 평가나 투자 권유를 목적으로 하지 않습니다. 국내 사례는 개별 기업명 대신 공개 IR·공시에서 확인되는 사례 유형으로 서술했습니다.