발간번호	2026-비정기-AGR-AI안전평가-700

프런티어 AI 보안 위협
금융분야 대응 요령
(버전 1.0)

2026년 7월

금 융 보 안 원

연 혁 표

구분	배포일	시행일	주요 내용
제정	2026.7.1.	2026.7.1.	프런티어 AI 보안 위협 금융분야 대응 요령 최초 제정(버전 1.0)

목 차

1 배경

□ **앤트로픽社 미토스 시스템 카드*와 글래스윙 프로젝트 중간 결과**를 통해 공개된 높은 성능은 금융 보안 체계에 **새로운 위협**으로 대두

* AI 모델의 성능 및 안전성 평가 결과 등이 포함된 기술 문서
** 한 달 만에 1만 건 이상의 신규 고위험 취약점을 발견

○ **각국 정부 및 NYDFS, CSA, FS-ISAC 등** 여러 기관에서 **미토스 쇼크에 대응**하기 위한 다양한 **전략 및 관리 항목 제시**

□ 이를 분석하여 **①거버넌스, ②취약점 관리, ③자산·공급망 식별/관리, ④AI 기술 활용, ⑤복원력, ⑥침해 확산 방지** 등 **6개 핵심 대응 요령** 도출

○ **금융회사 등이 참고**할 수 있도록 **국내 금융권에 특화**된 **상세 요령**을 안내

※ 본 대응 요령은 규제적 성격이 아니며, 프런티어 AI 보안 위협 대응을 위한 안내로서 참고

< 주요국 정부 및 기관별 프런티어 AI 보안 위협 대응 요령 비교 >

기관명	개요	핵심 전략 및 관리 항목
과학기술정보통신부 (대한민국)	□ 기업 전체 거버넌스와 기본기 강조 ㅇ 기술적 대응뿐만 아니라 경영진(CEO)의 역할 ㅇ 조직 문화 전반의 체질 개선 병행 요구 □ 국내 보안 프레임워크 제시 ㅇ 제로트러스트 2.0 모델 및 SIM3 모델 등 구체적인 측정 지표 활용 ※ AI 기반 사이버 공격 대비 기업 대응 요령(5.7.) 및 AI 기반 사이버공격 대비를 위한 CEO 행동 수칙(5.7.)	□ AI 기반 방어 전환 ㅇ 실시간 탐지·차단 자동 운영 체계 구축 ㅇ 코드 리뷰용 AI 에이전트 도입 □ 거버넌스 정비 ㅇ CEO-CISO 핫라인 및 정보보호 위원회 설치 ㅇ '완벽차단'에서 '신속회복'으로 목표 전환 □ 오픈소스 & 제로트러스트 ㅇ SW 자재명세서(SBOM) 작성 ㅇ 6대 요소별 제로트러스트 성숙도 진단 □ 민·관협력 ㅇ C-TAS, ISAC 채널을 통한 실시간 위협 정보 공유

기관명	개요	핵심 전략 및 관리 항목
과학혁신 기술부, 내각 사무처 (영국)	□ AI 보안 위협에 대한 공개서한 형태로 배포 ㅇ 위협 변화에 따른 대응 방식 변화 촉구 □ 조직적 대응과 기본기 강조 ㅇ 최고위층부터 조직의 보안 역량 강화에 의지를 가져야 하며, 기본적인 보안 요소부터 강화할 필요 ※ AI cyber threats: open letter to business leaders(4.22.)	□ 거버넌스 중심의 전사적 대응 ㅇ AI 보안 위협을 IT 조직의 단순한 실무 이슈로 치부하지 말고, 이사회와 최고경영진 수준의 핵심 경영리스크로 인식·대응 □ 기본기 중심의 방어 강화 ㅇ AI 기반 공격도 결국은 시스템의 기존 취약점을 악용해 침입, 고도화된 기술 도입에 앞서 기본적인 사이버 보안 위협 대응 조치를 완벽히 이행 필요
사이버 보안청 (싱가포르)	□ 골든타임의 극단적 축소 지적 ㅇ AI가 취약점을 찾아내고 악용하는 시간이 몇 달에서 몇 시간으로 단축되고 있으며 점점 더 짧아지는 추세 ㅇ AI 기반 공격은 인간이 인지하고 대응하는 속도보다 훨씬 더 빠른 '시스템 속도' 수준으로 수행 □ AI의 양면성을 인정·활용 ㅇ 프런티어 AI는 공격자에게만 유리한 것이 아닌, '보안 전문가 수준의 방어 기능'도 제공하기 때문에 방어 도구로 적극 활용 필요 ※ Advisory on Risks associated with Frontier AI Models(4.15.)	□ 중요 취약점 패치 ㅇ 인터넷에 노출된 시스템에 존재하는 중요도·심각도 높은 취약점 즉시 조치 □ 다중인증(MFA) 구현 ㅇ 클라우드 관리 콘솔을 포함한 모든 관리자 화면에서 다중인증 활성화 ㅇ 다중인증 적용이 어려울 경우 IP 주소 기반 접근통제 적용 □ 공격 표면 및 공격 경로 축소 ㅇ 경계 방어 및 시스템 강화 ㅇ AI 보안 위협이 현실화할 경우 조기 차단을 위해 마이크로 세그멘테이션 적용

기관명	개요	핵심 전략 및 관리 항목
금융 감독청 (NYDFS, 미국 뉴욕)	□ 규제 기관 관점의 리스크 관리 ㅇ 기존 보안 규정(23 NYCRR Part 500) 준수를 기반으로 한 위험 평가 업데이트 강조 □ 공급망 및 프로그래밍 보안 ㅇ AI가 생성한 코드의 위험성과 하위 제3자 의존성 관리에 집중 ※ Industry Letter: Heightened Cyberseucrity Risks Associated with Frontier AI Models(5.21.)	□ 신속한 취약점 관리 ㅇ 위험 평가 결과를 바탕으로 빠른 취약점 탐지 및 조치 프로세스 수립 □ 하위 공급망 의존성 맵핑 ㅇ 제3자 서비스 제공업체 및 하위 공급망의 취약점과 코드 검증, 모니터링 강화 □ 모니터링 강화 ㅇ 의심스러운 활동에 대한 로깅 및 보안 이벤트 경고 역량 진단 ㅇ 신속한 보고 체계 점검
CSA (Cloud Security Alliance)	□ AI 취약점 폭풍에 초점 ㅇ 공격자가 AI를 통해 단시간에 대량의 취약점을 찾고 복합 공격을 수행 □ 지속 가능한 개발(DevSecOps) 강조 ㅇ 개발 단계부터 LLM 기반 에이전트 적극 도입 권고 ※ The "AI Vulnerability Storm": Building a "Mythos-ready" Security Program(4.12.)	□ 기본기 및 보안환경 강화 ㅇ 네트워크 격리(Segmentation), 송신(Egress) 필터링, 강력한 다중 인증(MFA) 도입 □ 방어용 AI 에이전트 전면 도입 ㅇ 공격자의 자율 제어 속도에 대응하기 위해 사이버보안 인력 전반에 AI 에이전트 배치 □ 의사결정 패러다임 전환 ㅇ 보안패치 적용으로 인한 운영 다운타임(Downtime)에 대한 위험 허용 범위 재조정
FS-ISAC	□ 금융 부문 특화 리스크 관리 ㅇ 프런티어 AI 모델이 금융서비스에 미치는 위험 경고 ㅇ 전통적인 취약점 관리 가정을 전면 재수정 □ 공격 방어 중심 대응 ㅇ 단순히 탐지하고 사후 대응하는 것을 넘어, 공격의 확산을 사전에 '차단'하는 구조적 변화 요구 ※ Sector Risk Advisory: Preparing the Enterprise for AI-Enabled Vulnerability Discovery(4.19.)	□ 공격적인 취약점 해결 ㅇ 기존 보안 패치 지연을 컴플라이언스 부채가 아닌 '실질적 운영 리스크'로 취급할 필요 ㅇ 패치 기간을 주→일 단위로 단축 □ 경계 방어 현대화 ㅇ CDN, 클라우드 에지 제어, 웹 방화벽(WAF) 확장, 내부 트랩(tripwire)/기만(Deception) 기술 적용 □ 신규 오픈소스 통제 ㅇ 새로운 오픈소스 SW나 AI 모델 도입 시 검증 시간 확보를 위해 '의도적인 도입 지연(Controlled delay)' 검토

2 금융분야 대응 요령

가. 경영진 책임 강화

□ AI 보안 위협은 기술적 위험 외에도 **금융회사 등의 신뢰, 업무 연속성, 금융안정, 규제 준수 등 경영 전반에 영향을 미치는 사항**

○ **이사회 및 CEO 등 경영진 중심의 전사적 대응**이 필요

< 거버넌스 및 경영진 책임 관련 사례 >

(미국) 패치 속도와 플랫폼 최신성 등을 이사회와 거버넌스 위원회 등에 보고되는 운영리스크 지표로 취급할 것을 제안

(영국) 최고위층에서 AI 보안 위협을 심각하게 다룰 것을 촉구

(과기부) 정보보호는 CISO만이 아닌 경영진과 이사회가 다룰 핵심 안건

○ **이사회 내 정보보호위원회 등**에서 **AI 보안 위협 대응**을 논의하고, **CISO는 패치 관리, 외부 노출 자산, 서비스별 복원력 수준, 제3자 리스크, AI 방어 자동화 수준 등을 보고**

○ 최종 책임자인 **이사회와 CEO**는 **CISO에게 실질적인 예산 편성권 및 인력 운영의 권한**을 부여하고 **CIO도 AI 보안 위협 대응에 적극 참여**

□ 아울러, **AI 위협을 모니터링**하고 **대응 방안을 수립 · 추진**하기 위한 **CISO 직속 대응반 구성 · 운영**

○ **AI 위협 상황을 모니터링**하여 **신속하게 조치**, 취약점 인지 및 패치 개발을 포함한 **조치기간 단축을 위한 다양한 방안 도출 · 적용**

○ 네트워크 · 시스템 인프라 전문가, 보안 전문가, 개발 · 배포 파이프라인(CI/CD) 관리자, 서비스 개발자, 데이터 엔지니어 등으로 구성

□ **다수의 심각한 보안 사고가 동시에 발생**하는 **상황을 가정**하여 **정기적*** **모의훈련 실시**

* 매주 또는 매월, 취약점 대응이 안정화될 때까지 진행

○ **모의훈련 결과**는 중요 사고 대응을 위한 **매뉴얼 마련, 취약점 관리와 개발·배포 통합, 인프라 강화 등에 반영**

※ (CSA) 동시다발적 침해사고 발생에 대비, 대응 역량 강화를 위한 모의훈련 및 훈련 결과의 매뉴얼·인프라 반영 중요성 강조

적용 예시

▷ **이사회 보고용 '운영리스크 평가지표(KRI, Key Risk Indicator)' 대시보드 구성 예시**

○ (지표 설정)

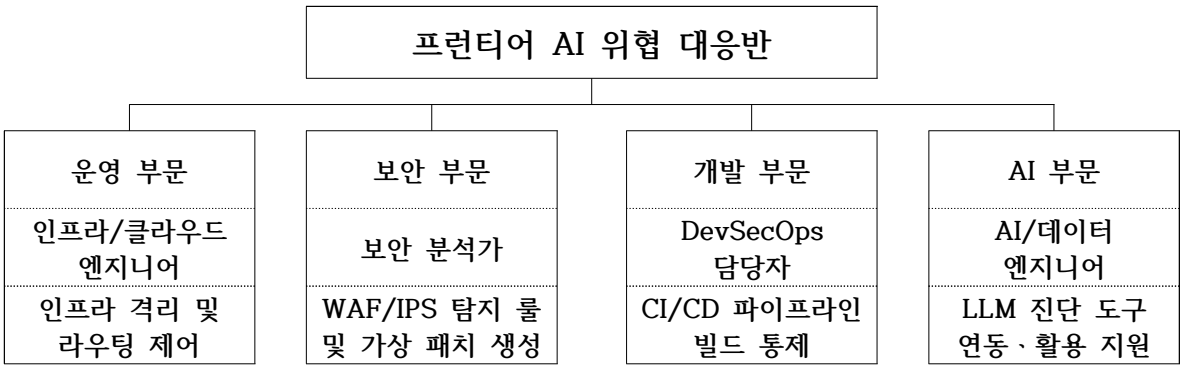
지표명	지표 설명
패치지연율	핵심 자산 중 24시간 초과 미패치 취약점 수
공격표면지수	대외 접점(인터넷 노출) 자산 중 미인증 접근 가능 포트 수
복원복구력	모의훈련 시 측정된 핵심 대고객 서비스의 실제 복구 시간(RTO)

○ CISO가 매월 지표를 취합하여 최고 경영자(CEO)에게 상시 보고, 분기별로 '정보보호위원회' 정식 안건으로 상정·보고

○ 지표를 참고, CISO에게 부여되는 예산·인력 연동

▷ **CISO 직속 '프런티어 AI 위협 대응반' 조직 구성 및 절차 예시**

○ (조직 구성)



○ (가상패치 조치 절차) ① 인지 → ② 가상패치 → ③ 코드수정/패치적용

① 위협 정보 채널 및 AI 진단 도구 등을 통해 신종 고위험 취약점 전파 접수

② 정식 패치 적용 전, 24시간 이내에 WAF 및 IPS 등 정보보호솔루션에 공격 차단 규칙·패턴 선반영

③ 취약점이 존재하는 소스코드를 수정하거나 프로그램에 패치를 적용, 근본 원인 제거

나. 취약점 및 패치 관리

□ **취약점 발견이 급증**할 것으로 전망됨에 따라 **패치 관리의 중심**을 **취약점 제거**에서, **공격 성공 가능성을 감소**시키는 것으로 **전환 필요**

< 취약점 및 패치 관리 관련 사례 >

(미국) 외부 시스템부터 패치하고 내부로 이동, 예외 제거, 취약점 SLA(Service Level Agreement) 일 단위 관리, 취약점 백로그(미조치 취약점)의 운영리스크 취급
(싱가포르) 고위험·중요 취약점의 즉시 패치외 취약점 노출 기간 최소화
(과기부) 자산 전수조사, 외부 접점 자산 우선 패치, 지원종료 시스템 격리

○ **신규 취약점 및 패치 배포가 크게 증가**할 것으로 예상되며, 이에 따라 **패치 우선순위에 대한 고려 필요성 또한 증가**

○ **고위험 취약점 일 단위 관리, 패치 예외를 승인한 경우에 대해 재검토, 패치 불가 시스템을 조치하는 방안*** 등 명확화

* 네트워크에서 격리, WAF·IPS 차단(가상 패치), 서비스 제한, 조기 교체 등

□ **취약점 점검에 활용할 수 있는 프런티어 AI 모델 중 금융회사 여건에 적합**한 모델을 활용하여 **내부 SW 점검**

○ **영향도 · 노출도가 높은 시스템 · 서비스**를 **우선적으로 점검을 수행**하고 **발견된 취약점**에 대해 **검토 후 조치**

○ **소스코드 외부 반출 불가 등의 사유**로 인해 외부 프런티어 AI 모델 활용이 어려운 경우, **SAST* 또는 온프레미스 AI 모델 활용** 가능

* Static Application Security Testing, 정적 소스코드 분석 도구

< 패치 우선순위 선정 관련 고려사항 및 예시 >

□ **금융회사 현황**에 맞는 **다양한 가점 요인과 감점 요인을 고려**해 패치 **우선순위 선정**을 위한 **기준 마련**

○ 아래 요소들의 경우 **상대적으로 높은 순위 배정 고려**

가점 고려	예시
취약점 심각도 및 악용 가능성이 높은 경우	알려진 취약점 점수(CVSS)가 높거나, 이미 시장에서 이를 악용한 공격 사례가 보고
핵심 비즈니스 및 자산 영향도가 높은 경우	패치 대상 시스템이 회사의 핵심 서비스(결제, 로그인 등)이거나 민감한 개인정보·기업기밀을 다루는 자산
대외 노출도가 높아 넓은 공격 표면을 허용하는 경우	외부 인터넷망에 직접 노출되어 있어 공격자가 쉽게 접근할 수 있는 시스템(외부 웹서버, VPN 등)
규제 및 컴플라이언스에 해당하는 경우	전자금융거래법, 신용정보법, 개인정보보호법, 전자금융감독규정, ISMS-P 등 법적·제도적 준수를 위해 필수적으로 기한 내 조치
패치 적용의 용이성	시스템 재부팅이 필요 없거나, 서비스 중단 시간 없이 즉시 적용 가능한 간단한 패치

○ 아래 요소들의 경우 **상대적으로 낮은 순위 배정 고려**

감점 고려	예시
대체 보안 통제 적용 가능한 경우	웹방화벽(WAF), 침입방지시스템(IPS) 규칙 적용, 또는 특정 포트 차단 등의 우회책으로 공격 차단 가능
노출도가 낮고 격리된 환경인 경우	내부 폐쇄망에 존재하거나, 외부의 접속이 격리되어 실질적인 공격 도달 가능성이 매우 낮은 시스템
종속성 및 시스템 영향도가 높은 경우	패치 적용 시 연동된 다른 레거시 시스템이 중단되거나, 서비스 환경에 심각한 부작용을 초래할 가능성이 높아 충분한 검증이 선행이 필요한 시스템

□ 내외부 SW **취약점점검 · 모니터링 · 조치 체계(VulnOps)**를 **통합 · 배포(CI/CD) 파이프라인에 적용**

○ 형상관리, 소스코드 빌드 시스템 등으로 구성된 **파이프라인**에 **취약점 점검 및 패치**를 위한 **기능*** **또는 플러그인 적용**

* 프런티어 AI 기반 취약점 점검, 신규 패치 모니터링·개발, 패치 사전 테스트·배포 등 적용

○ **AI 에이전트** 등을 이용한 **레드티밍 및 취약점 검증 자동화** 적용도 고려할 필요

○ 폐쇄망을 일부 운영하는 경우, 폐쇄망 내 **온프레미스 AI 모델** 기반 **취약점 점검 및 패치 생성 프레임워크** 필요성을 검토 후 필요시 마련

※ (CSA) 내부 프로그램 소스코드와 종속성 검사 및 CI/CD 파이프라인 상에서의 보안 검토 시 AI(LLM) 기술을 활용하여 수행할 필요

□ **안전한 패치 적용**을 위한 **사전 테스트**와 실패 시 **원활한 복구**를 위한 **시스템 백업 확보 등**은 여전히 중요

○ **패치 자동화 시 안전한 패치**를 위한 **기존 절차들도 함께 고려** 필요

적용 예시

▷ **자산 중요도 기반 패치 우선순위 점수 산정 기준 마련 및 적용 예시**

○ (우선순위 산식 적용) 내부 취약점 관리시스템(VMS)에 아래 연산 로직을 자동화 매핑

패치 우선순위 점수 = CVSS 점수 + 가점 요소 – 감점 요소

○ (세부 배점표 예시)

구분	평가 항목	세부기준	배점
기본	CVSS 점수	CVE 기반 취약점 고유 위험도 점수	0 ~ 10
가점	대외 노출도	인터넷망 직접 연결(외부 웹, VPN, 메일 서버 등)	+5.0
	핵심 비즈니스	계정계, 로그인 및 인증, 결제 처리 시스템	+4.0
	규제 대상	전자금융거래법 및 신용정보법상 고유식별정보 처리 시스템	+2.0
	패치 용이성	시스템이나 서비스 영향 없이 패치 적용이 가능한 경우	+3.0
감점	보완 통제 적용	WAF, IPS에서 해당 취약점 공격 패턴 차단 가능여부	-3.0
	폐쇄망	외부망과 완전히 분리된 독립 내부 폐쇄망	-4.0
	패치 영향도	패치시 연동된 다른 서비스·시스템 영향 여부	-5.0

○ (최종 조치 기준 예시) 합산 점수 15점 이상은 24시간 이내 무조건 패치 적용

▷ **CI/CD 파이프라인 내 AI 소스코드 자동 진단 절차 예시**

① (커밋/병합 요청) 개발자가 소스코드 저장소에 소스코드 병합을 요청

② (AI 보안 리뷰) 소스코드 저장소와 연계된 자동화 프로그램을 통해 프런티어 AI API를 호출

③ (정밀진단) 프런티어 AI 에이전트가 소스코드 내 취약점*을 자동 점검

* 하드 코딩된 API키, SQL 구문삽입 취약점, 공급망 오염 등

④ (자동 승인/거절) 취약점 발견 시 커밋/병합을 실패 처리하고, 개발자에게 수정 코드 예시를 포함한 보고서 발송

다. 자산 · 공급망 관리 강화

□ AI는 **금융회사 외부 공격 표면을 빠르게 탐색**할 수 있으므로, 금융회사는 **자산·공급망에 대해 정확한 식별 · 관리** 수행

< 자산·공급망 관리 강화 관련 사례 >

(미국) 핵심 자산 및 제3자 현황 최신화
(싱가포르) 제3자 리스크를 포함한 엄격한 공급망 보안
(과기부) SBOM(Sofware Bill Of Material) 및 SCA(Software Composition Analysis) 등 강조

○ 시스템, 애플리케이션, API, 클라우드, 오픈소스, 위탁 업체, 제3자 소프트웨어 등 **모든 자산을 통합적으로 관리**할 필요

○ 특히, **인터넷 노출 여부, 업무 중요도, 데이터 민감도, 패치 상태 등을 종합적으로 관리**해야 유사시 즉시 대응이 가능

< 참고 : 금융보안원 공격표면관리(ASM) 서비스 개요 >

□ 공격자가 네트워크를 통해 **전산시스템을 공격할 수 있는 경로**를 **사전에 모니터링해 대응**하는 **예방 중심**의 **보안관제 서비스**

○ **대표적 고위험 자산**인 네트워크 장치, IoT 기기 등 **엣지 디바이스***를 **목록화**하고, 제품군별 **위험 이상징후 포착**

* 사람이나 사물과 상호작용하는 현장에서 서비스가 실제로 작동하는 단말 기기로서, 관리가 어렵고 물리적으로도 쉽게 접근할 수 있어 공격 표면이 넓음

□ 금융회사 등이 실시간으로 **노출 자산, 식별된 소프트웨어 정보 및 취약점 정보 등을 직접 확인**하고 **능동적으로 조치**할 수 있도록 지원

□ 금융회사 자체적으로 활용중인 **오픈소스 SW를 조사**하여 **명세(SBOM 등)를 작성**하고, **신규 취약점 모니터링 · 조치 강화**

○ 공급자 제공 패치 급증에 대비, **패치 우선순위 기준 마련** 및 **패치공개 → 패치테스트 → 패치적용 프로세스 자동화 · 개선**

< 참고 : 금융보안원 금융권 SW공급망보안 플랫폼 SBOM 기능 개요 >

□ 금융회사에서 자체 개발한 SW에 대하여 구성요소(오픈소스)를 목록화하고 취약점을 식별·매칭할 수 있도록 SBOM 추출·분석 및 모니터링 기능 제공

- ㅇ 플랫폼에서 제공하는 SCA(Software Composition Analysis) 스캐너* 또는 자체 보유한 SBOM 추출 도구를 이용하여 표준 포맷의 SBOM 추출

  * 조직 내 소스코드 빌드 가능 환경 또는 소스코드가 들어있는 컨테이너에서 이용 가능

- ㅇ 추출된 SBOM을 플랫폼에 업로드하고 분석하여 SW 구성요소 정보(오픈소스 종류·버전 정보 등)을 식별 및 목록화
- ㅇ SW 구성요소 정보와 최신 CVE 취약점 DB를 매칭하여 취약점을 찾고 해당 취약점이 해결된 안전한 버전 정보 제공
- ㅇ 분석한 SBOM을 모니터링 자산으로 등록 시 신규 고위험 취약점 발생 알림 수신

□ SBOM을 통해 SW 구성요소 가시성을 확보하고, 신규 취약점 발생 시 SW에 대한 영향을 신속히 파악 및 대응할 수 있도록 지원

< 참고 : AI 기반 사이버공격 대비 기업 대응 요령(과기부, KISA) 중 오픈소스 식별・관리 >

❸ (오픈소스 식별・관리) 공격자가 미리 취약점을 파악할 수 있는 오픈소스를 통한 AI 사이버공격 위협에 대비 필요

- ㅇ 기업이 사용 중인 오픈소스 리스트(SBOM)*를 작성하여 현황 파악

  * SBOM(Software Bill of Material) : 소프트웨어를 구성하고 있는 다양한 정보에 대한 상세 내역

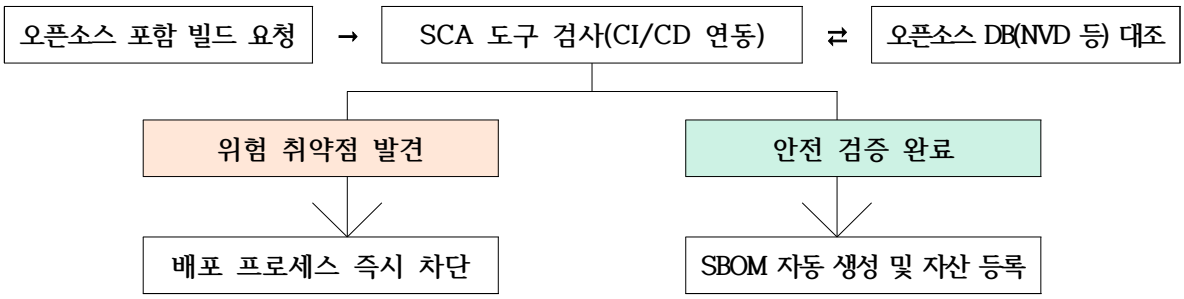
  - (생성 및 관리) 현재 사용 중인 라이브러리, 프레임워크, 종속성 등을 면밀히 조사하고 SBOM 생성 및 신규 업데이트 정보 관리
  - (분석 및 검증) 생성된 SBOM을 바탕으로 취약점 데이터베이스와 대조하여 잠재적 위협을 식별하고 오픈소스 사용 여부 검토

- ㅇ 공개되어 알려진 취약점이 포함된 오픈소스 사용이 자동 차단될 수 있도록 SCA(Software Composition Analysis) 도구를 적극 도입

적용 예시

▷ **오픈소스 SBOM 및 SCA 자동화 구성 예시**

○ (구성 예시)



○ (SCA 설정) 자사 소프트웨어 빌드 프로세스에 SCA 도구를 연동하고, 배포 시 SBOM이 생성되도록 구성

▷ **3rd 제품 도입 시 검증 강화 예시**

○ (격리 검증 구역) 외부 업체 솔루션 및 신규 오픈소스 소프트웨어 도입 시, 내부망과 격리된 샌드박스 검증 환경 구축·운영

○ (격리 분석 수행 예시)

- (1~7일차) 무결성 검증 및 SBOM 기반 하위 컴포넌트 취약점 전수조사
- (8~14일차) 악성 행위(비인가 아웃바운드 통신, 백도어 등) 유무 모니터링

라. AI 기반 방어 자동화

□ AI 기반 공격에 대비해 **보안시스템 자동화 체계**를 마련할 수 있으며, 보안 자동화 과정의 오탐지, 과잉대응, 서비스 장애 가능성에 대비하기 위해 **업무 위험도** 등을 고려하여 **자동화 수준**을 **차등화**할 필요

< AI 기반 방어 자동화 관련 사례 >

(미국) AI를 보안 경보 분류, 취약점 탐지·개선, 레드팀, 개발자 보안 검토 등에 활용하되 인간 감독과 거버넌스를 포함할 필요

(싱가포르) AI기반 취약점 탐지 시스템과 전사 인프라의 지속 스캔 권고

(과기부) AI 기반 상시 탐지·평가·대응 체계와 자동 운영체계 필요. 다만, 실제 서비스 시스템에 미치는 영향에 대해서는 사전 분석 필요

○ **(자동화 적극 적용)** 로그 분석, 취약점 우선 순위 산정, 코드 리뷰, 피싱 탐지, 보안 이벤트 요약, 위협 인텔리전스 분석 등

○ **(일부 자동화+인간개입)** 계정 잠금, 트래픽 차단, 시스템 격리, 패치 배포 등은 사전 시뮬레이션과 인간 승인을 전제로 운영

□ **엔드포인트 및 인프라***에서 **수집된 정보를 분석**, 여러 시스템에 걸친 광범위하고 **신속한 자동 대응** 수행 역량 확보

* 네트워크, 클라우드 워크로드, 이메일 게이트웨이 등

○ 신속하고 **능동적인 대응**을 위해 **EDR/XDR 등 정보보호 대응 인프라 강화**

< 엔드포인트/확장 탐지·대응 비교 >

구분	엔드포인트 탐지·대응 (EDR, Endpoint Detection&Response)	확장된 탐지·대응 (XDR, eXtended Detection&Response)
보호대상	PC, 서버, 노트북, 모바일 등	엔드포인트+네트워크+클라우드+이메일+계정 활동 등 모든 영역
데이터 소스	단말 내부 프로세스, 파일, 레지스트리 로그	기업 내 다양한 보안 솔루션의 이기종 로그 및 텔레메트리
분석 방식	단말 관점의 행위 기반 분석	이기종 데이터 간의 상관관계 분석
구축 목적	단말 내부로 침투한 위협 추적	기업 전체 인프라에 걸친 복합적 공격 식별 능력 확보 및 대응 자동화

적용 예시

▷ **위험 수준별 인적개입 차등화 예시**

○ 보안 장애 및 오탐으로 인한 서비스 중단을 방지하기 위해 AI의 자율 제어 권한을 3단계로 차등화

위험수준	대응 자산 및 시나리오	자동화 조치 수준	인적 개입 여부
저위험	대내 직원 포털, 단순 로그 분석, 피싱 메일 수신 등	완전 자동화 (예시) 악성 이메일 즉시 격리, IP 주소 자동 차단	조치 결과 로그만 모니터링
중위험	일반 웹서버, 내부 정보 자산 권한 상승 시도 탐지 등	조건부 자동화 (예시) 의심 단말의 네트워크 격리	보안관제요원 확인 후 5분 내 미승인 시 자동 격리
고위험	코어 계정계, 실시간 결제·이체, 인증 DB 등	권고 및 대안 제시 (예시) 차단 규칙 추천	보안담당자/CISO 최종 승인 후 차단 실행

▷ **XDR 기반 침해 단말 자동 격리 절차 예시**

① (로그 상호연관 분석) XDR이 EDR, 방화벽, 클라우드 가상 네트워크 등의 이기종 로그를 통합 분석

② (AI 위협 판단) 단시간 내 다수 단말을 거쳐 이동한 징후를 AI 에이전트를 활용해 식별

③ (API 기반 강제 격리) 인적 개입 없이 XDR이 해당 단말의 백신 에이전트 및 스위치 장비 접근제어목록(ACL)을 제어하여 네트워크 통신을 차단

마. 금융권 공동 대응 및 시스템 복원력 강화

□ **(회복 중심 리스크 관리)** 프런티어 AI 기반 공격에 대한 완벽한 방어는 매우 어렵기 때문에, **신속한 회복력(Resilience) 중심 사후 관리 대책 수립**

○ **위협 인지 시** 사람의 개입 없이도 **공격 IP 및 오염 세션**을 **실시간 격리**할 수 있는 **자동화 대응 메커니즘 마련**

○ 업무연속성계획(BCP, Business Continuity Plan)에 의거 **오염되지 않은 백업본으로 서비스를 자동 복원**하는 체계 정비

□ 금융회사가 사용하는 **동일한 솔루션, 클라우드, 오픈소스, 위탁업체** 등의 경우 **하나의 취약점이 빠르게 확산**할 가능성

○ 특히, **AI의 대규모 보안 위협에 대한 개별 금융회사 차원의 대응에는 한계**가 있으므로 **복원력 중심의 공동 대응*** 필요

* 위협정보 공유, 공동 탐지룰, 취약점 선공개 대응, 공급망 공동 점검, 모의훈련 등

□ **금융AI보안연구소 지원센터**(aisupport@fsec.or.kr, 02-3495-9851)를 통해 **프런티어 AI 보안 위협 공동 대응** 및 **관련 안내 지원**

○ 아울러, **보안 위협 발생** 시에는 **금융ISAC**(isac@fsec.or.kr, 02-3495-9999)을 통해 **관련 정보를 전파**하고, **신속하게 대응**

< 금융권 공동 대응 및 시스템 복원력 강화 관련 사례 >

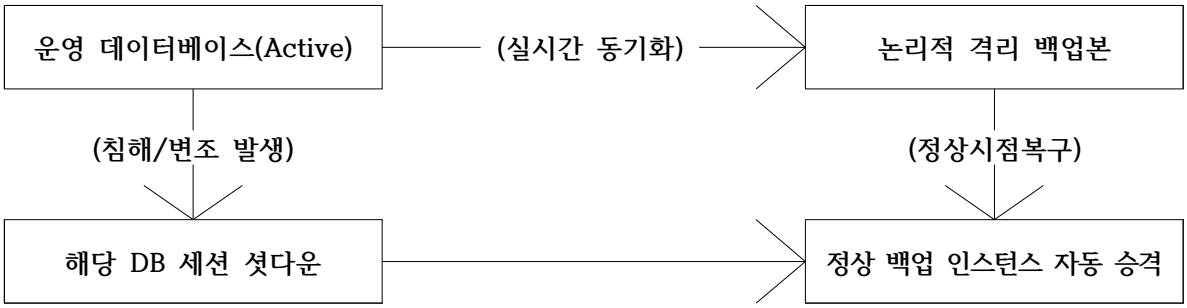
(미국) 개별 조직은 새로운 위협을 모두 식별 할 수 없으며, 공동 대응이 공격자에게 기울어진 우위를 금융시스템 복원력 쪽으로 되돌릴 수 있음

(과기부) ISAC, C-TAS, 정보보호 포럼 등을 통한 위협정보의 대응 방향 공유

적용 예시

▷ **업무연속성계획(BCP) 기반의 데이터베이스 자동 롤백 및 서비스 복원 예시**

○ (아키텍처 구성)



○ (구현 절차) 변조 및 랜섬웨어 공격 확산 인지 즉시, 침해·변조된 채널을 차단함과 동시에 격리 보관 중인 최신 백업을 즉시 운영 서버로 승격

○ (주기적 점검) 업무연속성 자동화 스크립트를 주기적으로 실행·점검

바. 침해 확산 방지

□ AI 기반 공격은 **초기 침투 이후 권한 상승과 횡적 이동**을 통해 **금융회사 내부 시스템 전체로 빠르게 확산**할 우려

○ 내부의 구성원을 믿는 **경계 기반 보안이 아닌** 공격자가 내부에 침입했을 수 있다고 가정, **항상 검증하는 제로트러스트 개념 적용**

< 침해 확산 방지 관련 사례 >

(싱가포르) 네트워크 세그멘테이션, 최소권한, 지속 검증, 런타임 애플리케이션 보안 모니터링 등 장기 조치

(과기부) 제로트러스트가 핵심 과제 (식별자·신원, 기기·엔드포인트, 네트워크, 시스템, 애플리케이션·워크로드, 데이터의 6대 요소별 진단 권고)

○ **다중 인증, 최소 권한, 마이크로세그멘테이션**, 휴면계정 제거, API 접근통제, 중요 데이터 접근 검증 등 **핵심 통제 강화**

* 고객정보, 결제/인증, 계정계, 대외계, 관리자 콘솔 등은 별도로 분리하여 운영하고, 내부망이라는 이유만으로 무조건적 신뢰 금지

※ (FS-ISAC) 취약점 관리에서 익스플로잇 방어로의 인식 전환 필요, 침해 발생시 내부 이동 제한을 위한 네트워크 세분화, 접근통제 및 시스템 간 격리가 중요

□ **제로트러스트, EDR/XDR 등 솔루션** 도입 시 **신뢰성 확보**를 위해 **국내외 인증·평가*** 결과를 참고

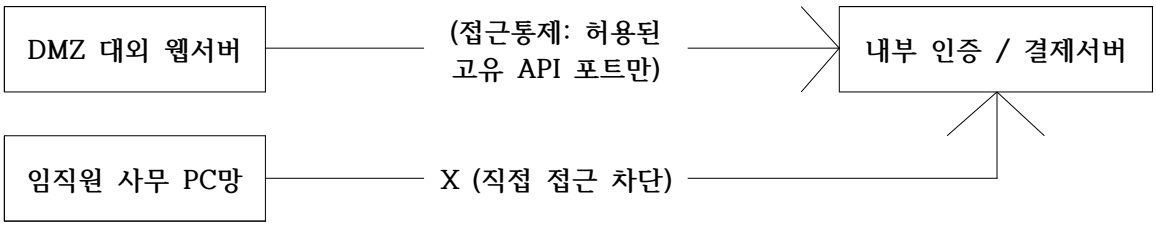
* (예시) 국내외 CC 인증, MITRE, SOC2, FedRAMP 등

적용 예시

▷ **마이크로세그멘테이션 적용 예시**

○ (원칙) 모든 프로그램(워크로드)간 통신을 기본 차단(Deny All)로 설정

○ (구체적 격리 구성 예시)



○ 웹서버가 프런티어 AI 공격으로 완전히 장악되더라도, 내부 계정계나 고객 정보 DB로 이동할 수 없도록 가상 네트워크 인프라(SDN) 수준에서 프로그램 간 통신 허용 범위를 단일 API 단위로 축소하고 접속을 매 순간 검증

▷ **피싱 저항 다중인증(MFA) 적용 및 휴면계정 관리 자동화 예시**

○ (MFA 강화) 중요 서버 접속, 중요 정보 조회, 이상 행위 탐지시 생체 인증, 하드웨어 보안토큰 등 강화된 추가 인증을 적용하여 침해사고 방지

※ SMS 인증 및 OTP의 경우 공격자가 AI 보이스, 딥페이크, 피싱 사이트 등을 악용하는 경우 우회 가능성 존재

○ (계정·권한 관리 자동화)

- 인사관리 시스템과 권한관리 시스템을 실시간 연동
- 인사정보 변경 시 즉시 권한 삭제
- 일정 기간* 이상 로그인 이력이 없는 계정은 스크립트를 통해 자동으로 비활성화 또는 권한 삭제 수행

* 업무별 특성 및 시스템 이용 빈도 등을 고려하여 비활성화 및 권한 삭제를 수행할 기간 설정