PROJECT

AI와 자동화 중심 MDR 서비스 체계 구축

팀명 : 플랜IT조
연계기업 : (주) 지니언스

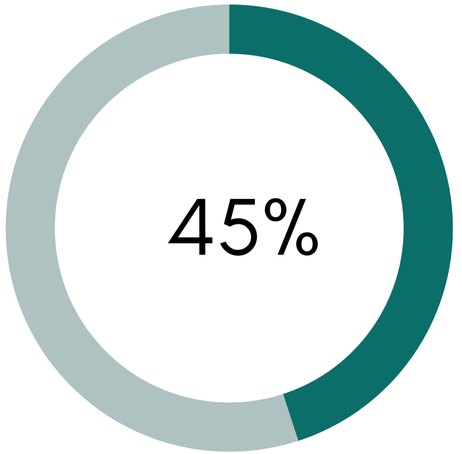
S-개발자 3기 2차 기업프로젝트 최종 발표

팀장 : 강서현
팀원 : 김동혁
김진솔
이시하

CONTENT

01. 제품 소개

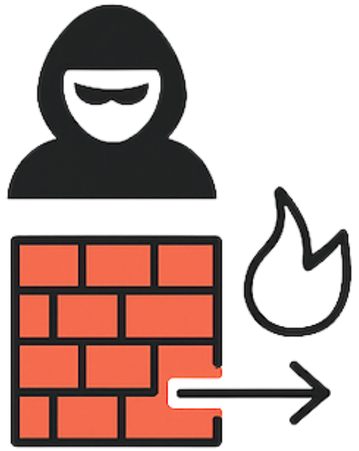
보안 전문인력 부족



수동검증이 필요한 EDR이벤트 비율

01. 제품 소개

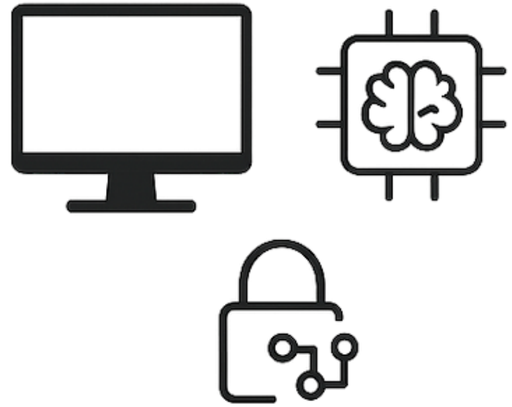
단말 중심 보안의 중요성



기존 네트워크 중심

- 알려진 위협만 탐지
- 신종 위협 대응 불가

보안 전략의 진화

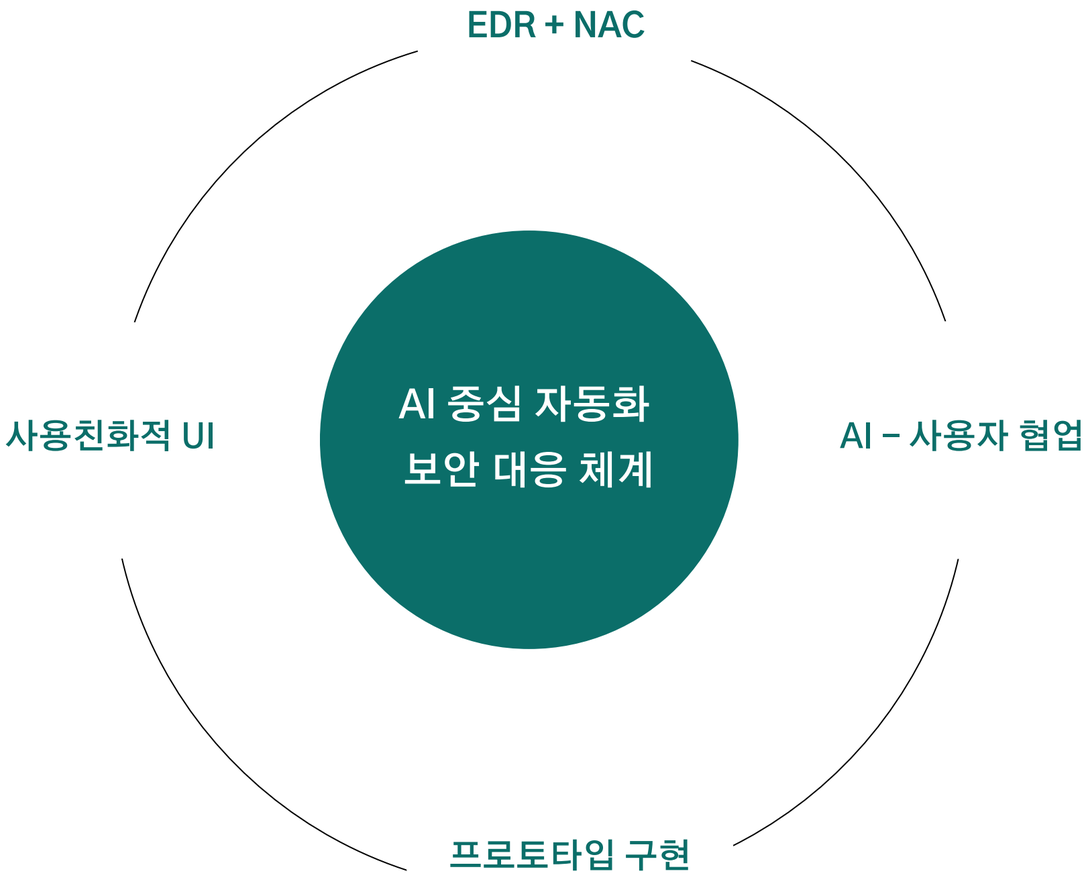


엔드포인트 중심

- 단말에서 위협 감지
- 즉각 대응 및 추적에 용이

01. 제품 소개

Planet MDR 의 방향성



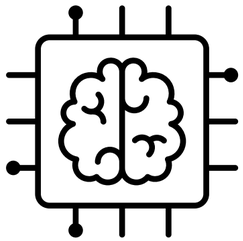
01. 제품소개

데이터 수집

NAC

EDR

AI 분석



- AI 기반 위협분석 자동화
- 위험도 기반 자동 의사결정
- 정책 준수형 자동 대응 실행

자동 대응

프로세스 차단

네트워크 격리

알림

안전

- 위험도 높은 티켓은 즉시 격리
- 불확실한 경우, 담당자 분석 요청

보고

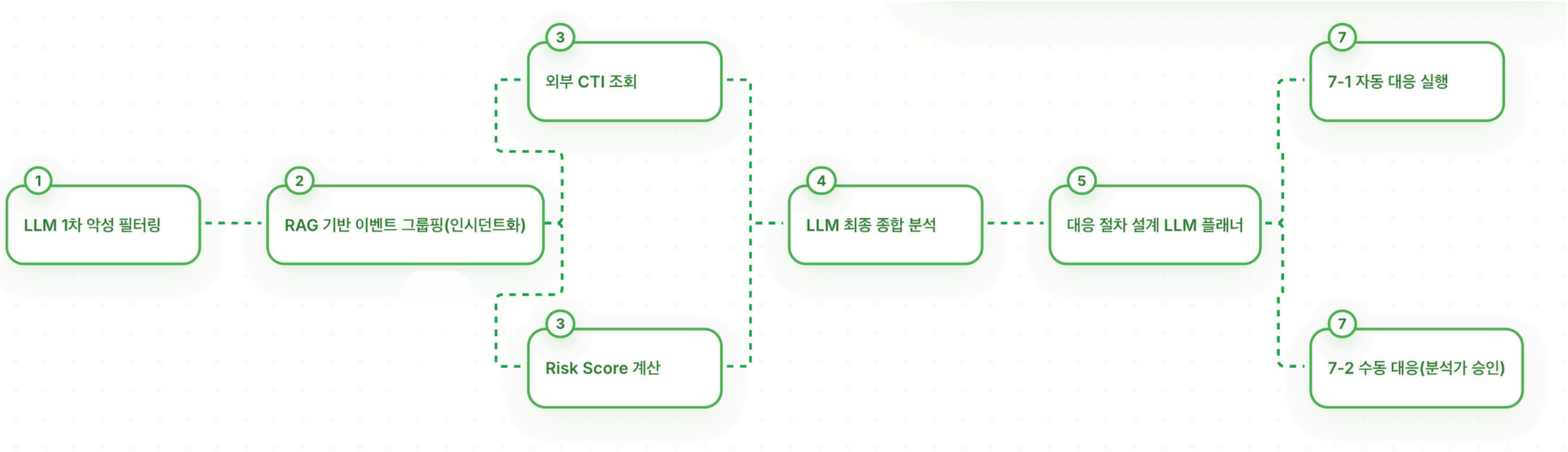
주간(월간) 보고서

티켓 알림

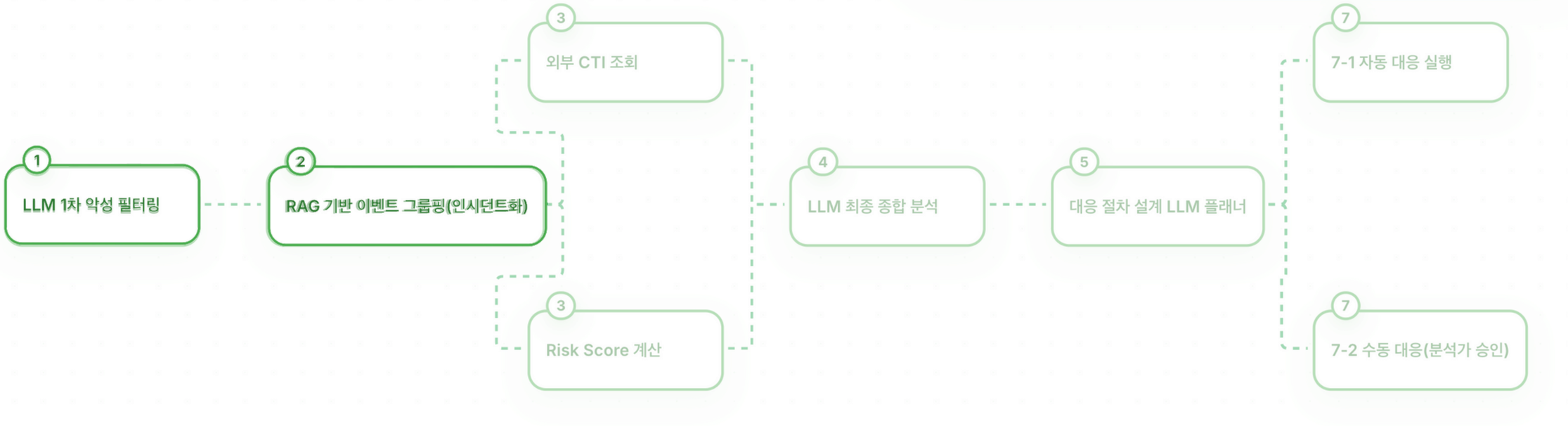
고객사 전달

02. 분석·대응 자동화 구조

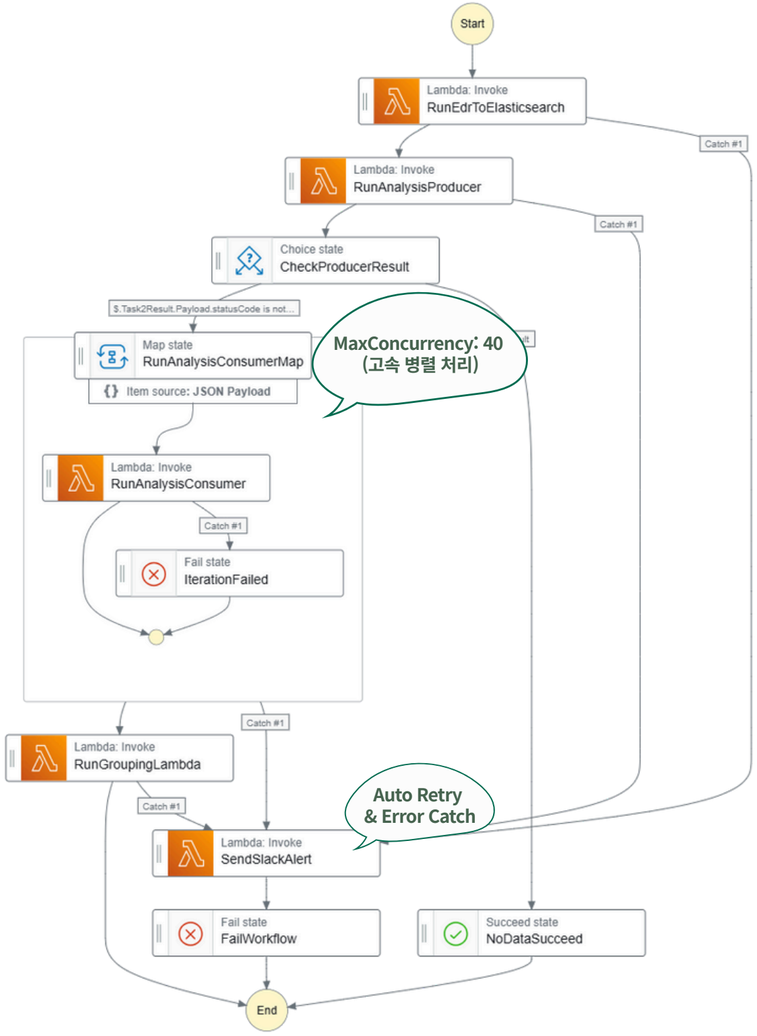
전체 구조


1
LLM 1차 악성 필터링
2
RAG 기반 이벤트 그룹핑(인시던트화)
3
외부 CTI 조회
3
Risk Score 계산
4
LLM 최종 종합 분석
5
대응 절차 설계 LLM 플래너
7
7-1 자동 대응 실행
7
7-2 수동 대응(분석가 승인)

전체 구조


1
LLM 1차 악성 필터링
2
RAG 기반 이벤트 그룹핑(인시던트화)
3
외부 CTI 조회
3
Risk Score 계산
4
LLM 최종 종합 분석
5
대응 절차 설계 LLM 플래너
7
7-1 자동 대응 실행
7
7-2 수동 대응(분석가 승인)

Step Function 전체 구조



Ingestion
EDR API
→ S3/ES 저장

Scheduling
분석 대상 분배

Analysis
AI 개별 분석
(악성 or 정상)

Grouping
분석된 위협 이벤트를
티켓으로 묶기

EDR Collector(Ingestion)

Analysis Producer(Scheduling)

Source Index
전체 수집된 로그

Dest Index
이미 분석 완료된 로그

이미 처리된 문서 제거
→ Cost Efficiency

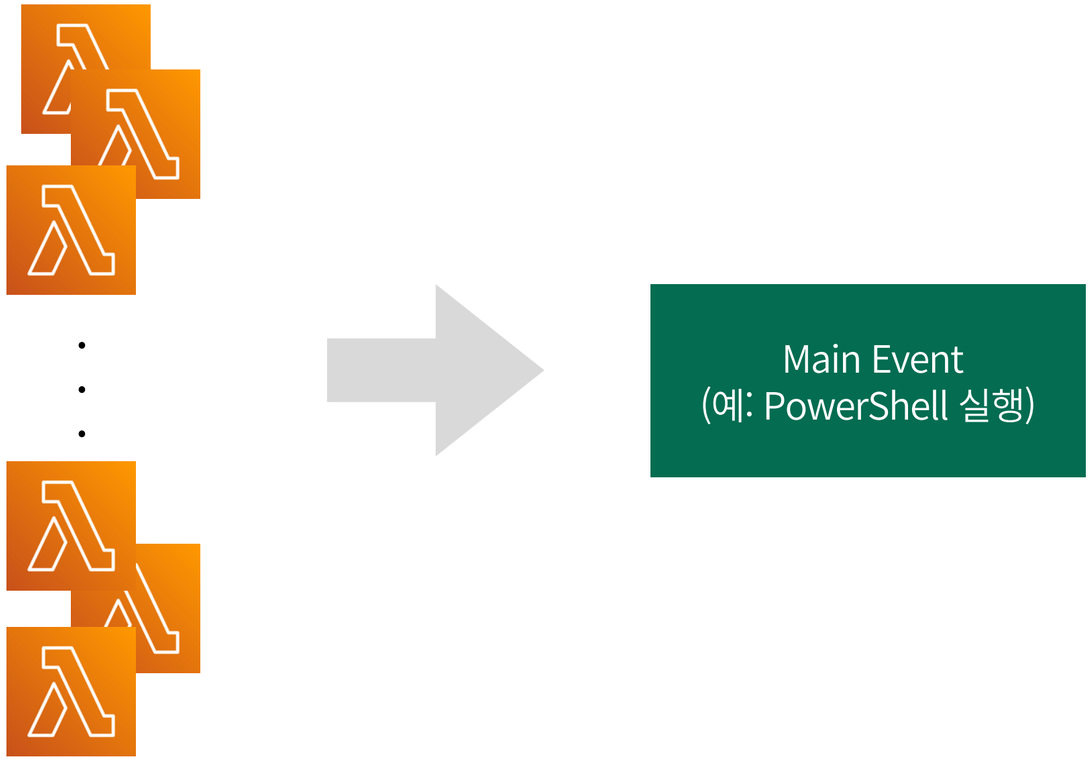

```
[
  { "UniqueID": "A-101", "Index": "..." },
  { "UniqueID": "B-205", "Index": "..." },
  { "UniqueID": "C-309", "Index": "..." }
]
```

“ 중복 분석 방지로 API 비용 절감
처리 누락/중복 없는 정확한 필터링 ”

Analysis Consumer(AI Analysis): 문맥 기반 판단

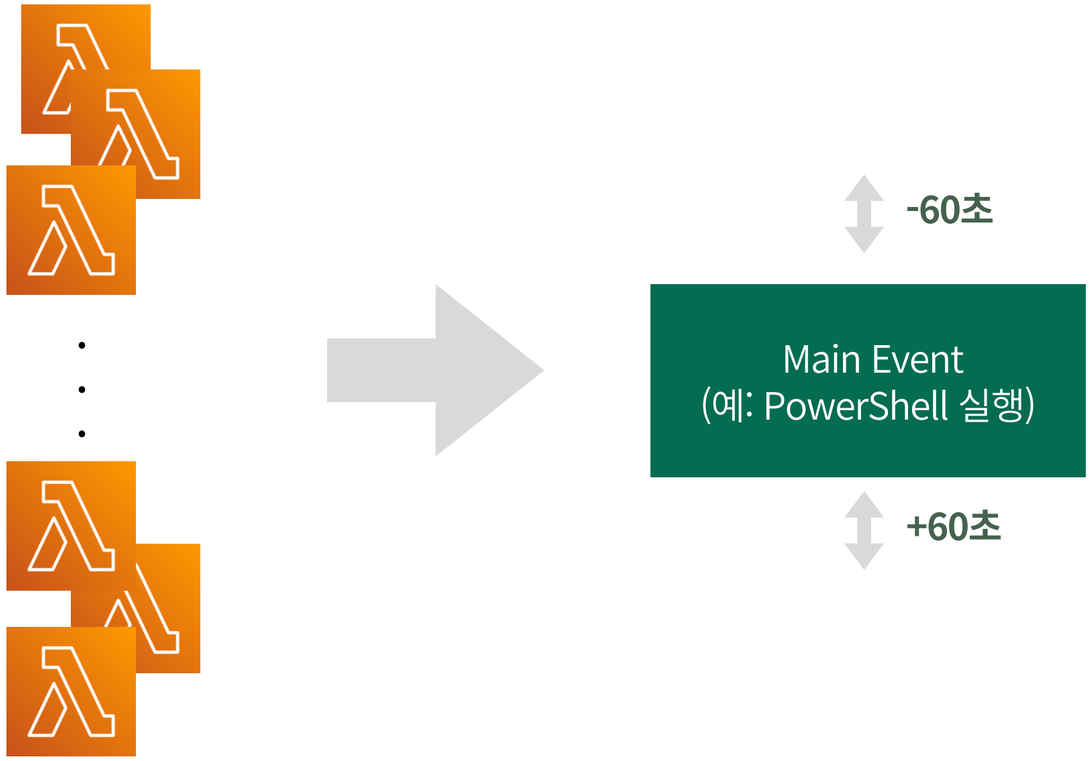
Max Concurrency: 40

Analysis Consumer(AI Analysis): 문맥 기반 판단



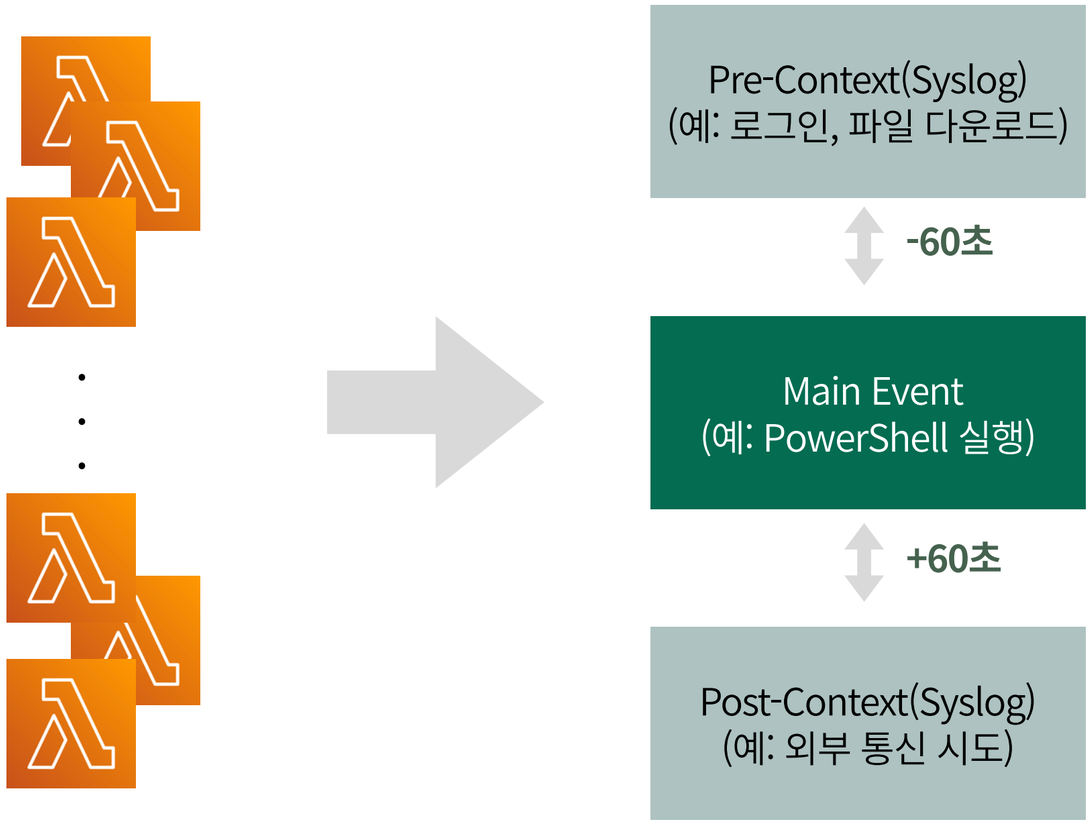
Max Concurrency: 40

Analysis Consumer(AI Analysis): 문맥 기반 판단



Max Concurrency: 40

Analysis Consumer(AI Analysis): 문맥 기반 판단



Max Concurrency: 40

Analysis Consumer(AI Analysis): 문맥 기반 판단

Max Concurrency: 40

Pre-Context(Syslog)
(예: 로그인, 파일 다운로드)

-60초

Main Event
(예: PowerShell 실행)

+60초

Post-Context(Syslog)
(예: 외부 통신 시도)

단일 로그는 정상이지만,
파일 다운로드 직후 실행
되었으므로 악성입니다.

Analysis Consumer(AI Analysis): 정밀도 향상 전략

Context Noise Reduction

단순 상태 정보 등 불필요한 로그를 배제
Process, Network, File, Registry 4대 핵심 이벤트 집중

Devil’s Advocate

단편적인 악성 지표가 발견되더라도, AI가 스스로 반론을 제기하여 정상 행위인지 재검증

Override Rule

모의 훈련 키워드가 감지되면, 악성으로 분류하되 신뢰도(Confidence)를 낮추어 실제 위협과의 혼동 방지

Analysis Consumer(AI Analysis)

t ai_analysis.analysis_summary	의심스러운 상위 프로세스(C:\Users\Public\splunkd.exe "-group red")에서 실행된 PowerShell(ExecutionPolicy Bypass)이 동일 호스트(포트 5985)로 WinRM Invoke-Command를 실행했습니다. 같은 세션에서 AMSI가 포착한 파일리스 실행 이벤트와 PowerShell이 정책 테스트 아티팩트 및 PowerShell 로그 폴더에 접근/삭제하는 동작이 확인되어, 스크립트 기반의 메모리 내 실행과 로그 변조/방어 회피 가능성을 시사합니다. 5985 포트로의 반복적인 연결 시도는 대상이 로컬호스트이더라도 원격 실행/수평 이동(lateral movement) 기법과 부합합니다.
ai_analysis.analyzed_at	Dec 3, 2025 @ 08:02:38.294
# ai_analysis.confidence	85
# ai_analysis.context_events_count	20
t ai_analysis.counter_evidence	관리자나 모니터링 도구가 신뢰성을 위해 PowerShell Remoting을 일괄적으로 사용(로컬호스트 포함)하고 ExecutionPolicy를 Bypass로 설정하는 경우가 있습니다. __PSScriptPolicyTest* 파일은 무해한 PowerShell 엔진 산출물일 수 있습니다. Splunk 유사 에이전트가 스크립트형 입력을 실행할 수 있으며, "group red"는 내부 라벨일 수도 있습니다. Temp 경로에서의 명시적 다운로드/실행이나 자격 증명 덤핑 정황은 관찰되지 않았습니다.
t ai_analysis.reason	맥락상 비표준 바이너리(Users\Public 내 splunkd.exe)가 ExecutionPolicy Bypass로 PowerShell을 실행하여 WinRM을 통해 로컬호스트 대상으로 원격 실행을 수행했고, 여기에 AMSI 파일리스 실행과 PowerShell 로깅 관련 의심스러운 파일 조작이 동반되었습니다. 이는 일반적인 관리자 활동과는 거리가 있는 강한 방어 회피 및 LOtL(정상 도구 악용) 지표입니다.
t ai_analysis.result	malicious

Grouping Lambda(RAG Grouping)

Step 1: Embedding & Search

LLM Summary

이벤트 로그를 “공격 행위 요약문“으로 변환
(IP, ID 등 노이즈 제거)

Vector Embedding

AWS Bedrock Titan 으로 벡터 변환

[예시]

`ai_group_id:` 163 `ticket_vector:` -0.05190053, 0.00272339, 0.152577, 0.041076355, 0.090478994, -0.059116647, 0.07660185, -0.472525, -0.00001517813, 0.011240489, 0.061614532, 0.07160608, 0.06688785, 0.010824175, -0.0034172474, -0.03608058, 0.0688306 -0.020954492, 0.019150462, 0.04995773, 0.033860236, 0.05079036

Vector Search

OpenSearch에서 유사한 과거 사례(k-NN) 검색

Result

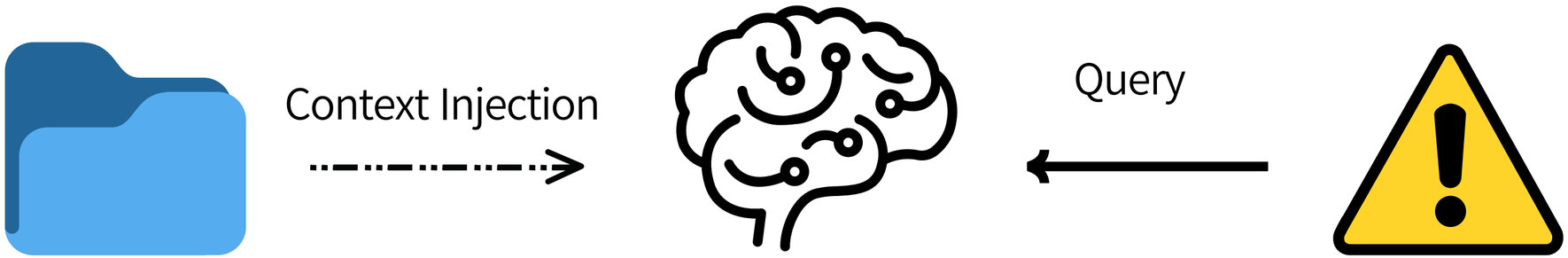
[판정 결과]

- 유사도: 과거 '랜섬웨어 사례 #45'와 90% 일치
- 조치: 동일 Case ID 부여 및 병합
- **Adaptive Threshold (적응형 임계값)** 적용

batch_classifiy_events()

Grouping Lambda(RAG Grouping)

Step 2: Context Injection



Reference Context

유사도가 검증된 과거 공격
시나리오의 전체 시퀀스 로딩

Contextual Alignment

과거 시나리오 기준으로
현재 이벤트를 재정렬할 준비 완료

Target Events

시간/순서가 뒤섞인 채 탐지된
개별 악성 이벤트 목록

Grouping Lambda(RAG Grouping)

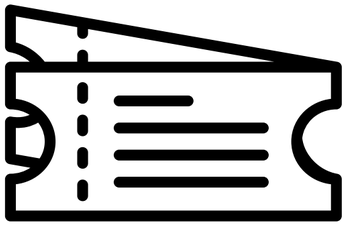
Step 3: Generative Grouping

system_prompt.txt

```
[Goal]
Target Events를 Reference Story(과거 사례) 흐름에 맞춰 그룹핑하라.

[Grouping Criteria (MUST)]
1. Time Proximity: 공격 단계(Step 1→2)가 연속적인가?
2. Host Match: 동일 호스트에서 발생했는가?
3. Context Chain: Reference의 순서와 문맥이 일치하는가?

[Negative Rules (Exception)]
흐름이 끊기거나 단순 유사도만 높은 경우 → "Separate Ticket" (단독 티켓)
공격 흐름과 무관한 이벤트 → "Exclude" (제외/분리)
```



group_events_by_context()

Grouping Lambda(RAG Grouping)

t ai_analysis.analysis_summary	splunkd.exe가 C:\Users\Public에서 "-server http://192.168.105.143:8888 -group red" 인자를 사용해 실행되었고, 대화형 사용자 계정으로 포트 8888에 아웃바운드 HTTP 연결을 수행했습니다. 이 경로, 명령줄 패턴, 사용자 컨텍스트는 정상 Splunk 구성요소로서는 이례적이며, 내부 호스트와 C2 채널을 수립하려는 위장된 에이전트를 시사합니다.
ai_analysis.analyzed_at	Dec 4, 2025 @ 07:03:20.166
# ai_analysis.confidence	88
# ai_analysis.context_events_count	20
t ai_analysis.counter_evidence	- 192.168.105.143:8888은 비표준 포트를 사용하는 내부 구성 Splunk HEC(HTTP Event Collector)나 관리용 엔드포인트일 수 있음. - 일부 조직은 사용자 컨텍스트에서 텔레메트리 포워더나 커스텀 도구를 테스트하기도 하며, "group red" 플래그가 내부 파라미터명일 수 있음. - 직전 60초 관찰 창에서 의심스러운 부모 프로세스가 없고, OS 활동(TiWorker, TrustedInstaller)도 정상으로 보이며, 이는 침해를 직접적으로 시사하지는 않음.
t ai_analysis.reason	여러 강한 이상 징후: (1) splunkd.exe가 Program Files가 아닌 C:\Users\Public에 위치해 있고 Windows 서비스로 실행되지 않음; (2) 명령줄 플래그(-server, -group red)가 Splunk의 일반적 사용과 일치하지 않음; (3) 비표준 포트 8888의 HTTP로 연결; (4) LocalSystem/Splunk 서비스 계정이 아닌 DESKTOP-PJUQHNC\SeohyeonKang 사용자 컨텍스트에서 실행. 이를 종합하면 정상적인 관리 작업이라기보다 프로세스 위장 및 C2/에이전트 행위일 가능성이 큽니다.
t ai_analysis.result	malicious
# ai_analysis.result_code	1
# ai_group_id	171
ai_grouped_at	Dec 4, 2025 @ 07:04:27.471

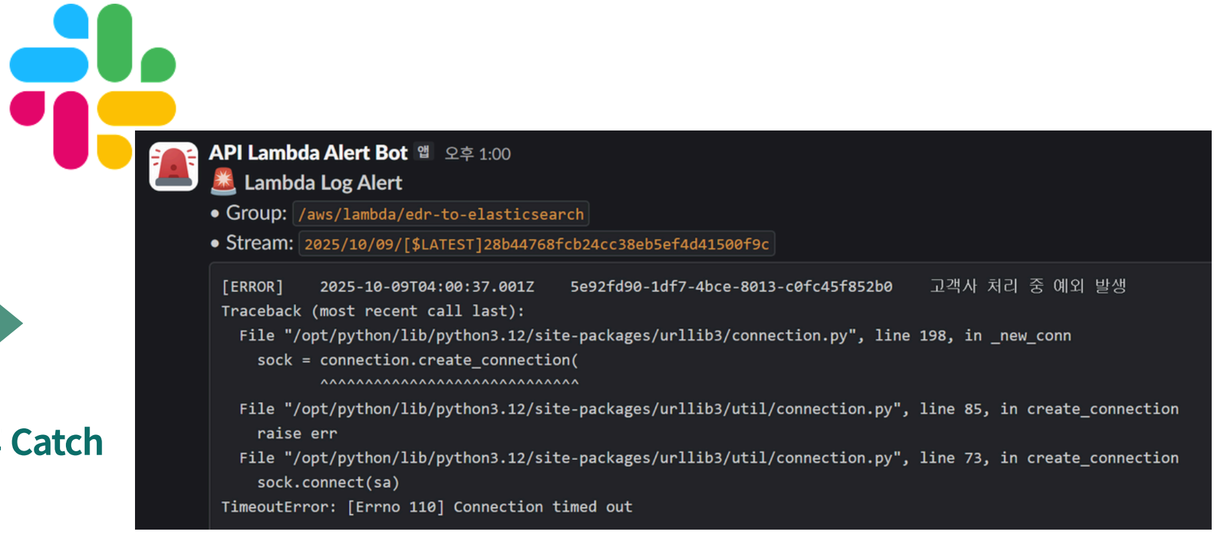
group_events_by_context()

Slack Alert & Reliability

step function 실행 중
에러 발생

Slack 알림 전용
Lambda

Retry 3회 후 Catch



분석

```
- CTI(reasons by source):
- - VT:
  - VT_hash: malicious=64
  - VT_hash: suspicious=0
  - VT_hash: total=76
  - VT_ip(3.33.130.190): malicious=1 total=95
- - AbuseIPDB:
  - AbuseIPDB(3.33.130.190): reports=19
  - AbuseIPDB(3.33.130.190): distinctSources=14
  - AbuseIPDB(3.33.130.190): confidence=27
  - AbuseIPDB(3.33.130.190): lastReportedAt=2025-11-30T14:03:44+00:00
  - AbuseIPDB(3.33.130.190): minutesSinceLast=4749
- - KISA: (none)
- - SecurityTrails: ST_core(mojobiden.com): A=15.197.148.33(Amazon.com, Inc.),
3.33.130.190(Amazon.com, Inc.) | NS=ns04.domaincontrol.com(GoDaddy),
ns03.domaincontrol.com(GoDaddy) | MX=-
- - host.io: host.io(mojobiden.com): ok status=200, title=, links=[]
```

외부 CTI 조회

Flow Guide

이벤트에 등장하는 해시·IP·도메인을 외부 CTI에서 자동으로 조회해주는 단계입니다.

· VirusTotal, AbuseIPDB, SecurityTrails, host.io에서 악성 여부/정보 조회
- 입력 데이터: 이벤트에 포함된 해시, IP, 도메인 등 IOC
- 출력 데이터: CTI 검색 결과

분석

Risk Score 계산

Flow Guide

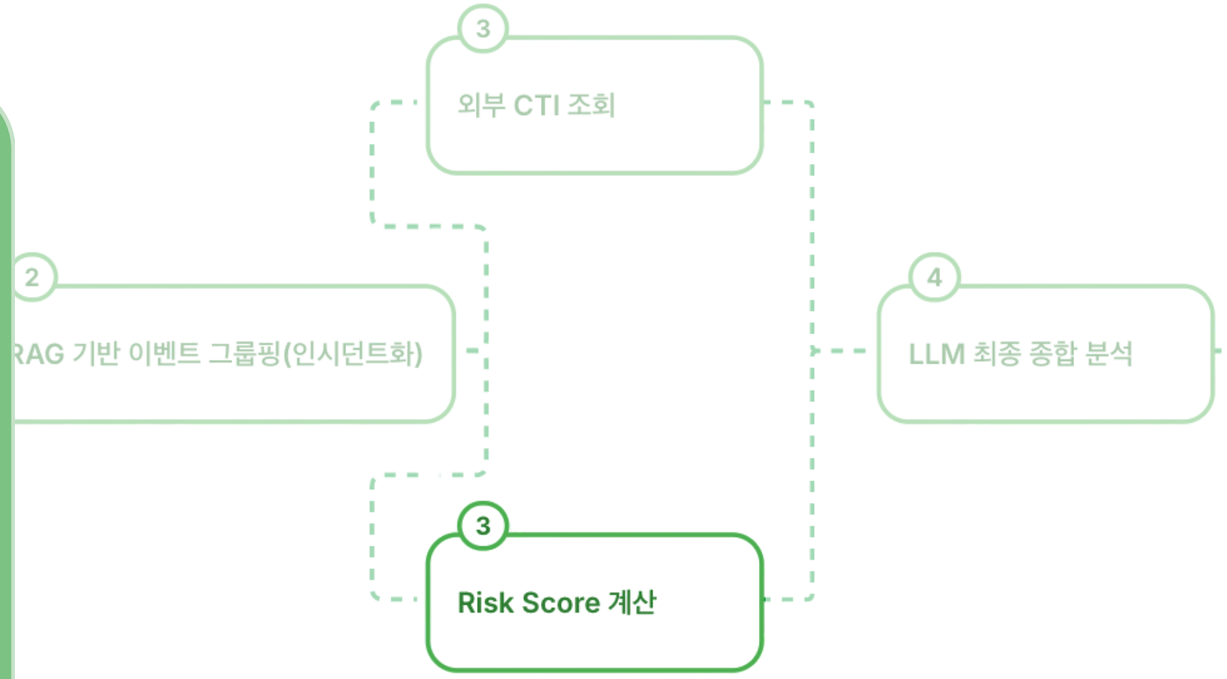
이벤트의 신뢰도 및 영향도를 점수화하는 단계입니다.신뢰도와 영향도를 각각 공식으로 계산해 종합 위험 점수를 산출합니다.

1) 신뢰도 계산
· ML 모델의 confidence와 EDR 정책 기반 신뢰도를 조합해 계산합니다.
· 신뢰도 = 0.8 × ML confidence + 0.2 × EDR 정책 신뢰도

2) 자산 중요도
· 부서/직급의 중요도를 가중합한 값입니다.
· 자산 중요도 = 0.6 × 부서 가중치 + 0.4 × 직급 가중치

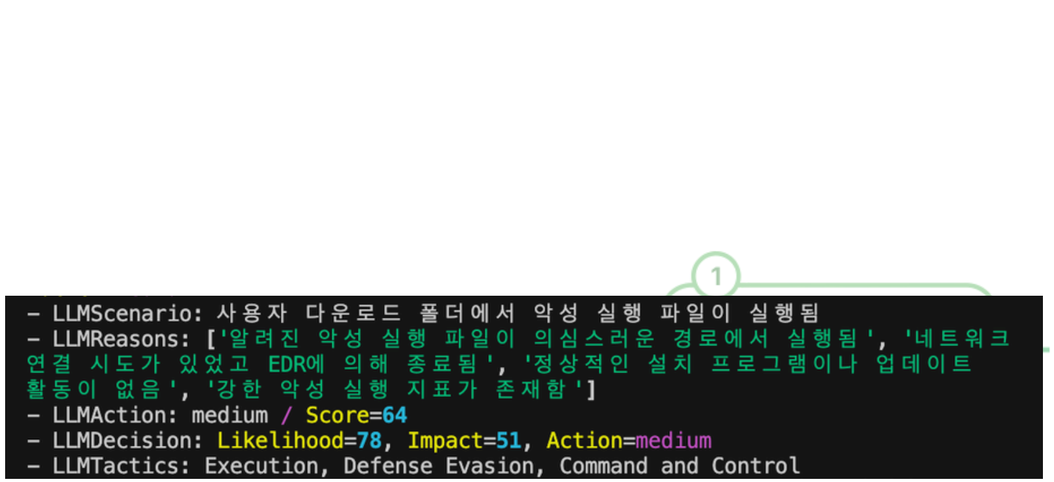
3) 영향도
· 공격 전술(Tactic)의 위험도와 자산 중요도를 결합합니다.
· 영향도 = 0.3 × 전술 가중치 + 0.7 × 자산 중요도

4) 종합 위험 점수
· 신뢰도와 영향도를 동일 비율로 조합해 위험 점수로 계산합니다.
· 종합 위험점수 = 0.5 × 신뢰도 + 0.5 × 영향도




```
- RiskCalc: Likelihood = 0.8*ML(0.95) + 0.2*Policy(0.10) -> 78; AssetCrit =
0.6*DeptW(0.50) + 0.4*RoleW(0.55) -> 0.52; Impact = 0.3*TacticW(0.50) +
0.7*AssetCrit(0.52) -> 51; Risk = 0.5*Likelihood + 0.5*Impact -> 64;
```

분석



LLM 최종 종합 분석

Flow Guide

앞에서 모은 정보(원본 로그, CTI, 종합 위험 점수)를 한 번에 모아서, 최종 분석을 하는 단계입니다.

· 공격 요약 정리 및 위험 등급 산정
  MITRE ATT&CK 전술/기술 검토
  경우에 따라 Risk score를 ±10 재조정
- 입력 데이터: 원본 이벤트, CTI 검사 결과, Risk score
- 출력 데이터: 공격 요약 및 근거, 재조정된 Risk score/TTPs

대응


```
- Planner: nodes=2 edges=1 reason=의심스러운 프로세스를 종료하고 알림을
전송합니다. reasons=['악성 실행 파일이 사용자 다운로드 폴더에서 실행됨.', '강한
악성 실행 지표가 존재함.', 'EDR에 의해 프로세스가 종료됨.', '정상적인 설치
프로그램이나 업데이트 활동이 없음.']
```

대응 절차 설계 LLM 플래너

Flow Guide

종합 분석 결과를 바탕으로 무엇을, 어떤 순서로, 자동으로 할지 대응 방안을 설계하는 단계입니다.

· 프로세스 강제종료, 네트워크 격리, 알림 전송 조치 후보 산정
· 위험 등급에 따라 선조치 후보고 / 담당자 확인 후 조치 로 나뉨
- 출력 데이터: 대응 방안, 판단근거, 추천 조치

대응

7

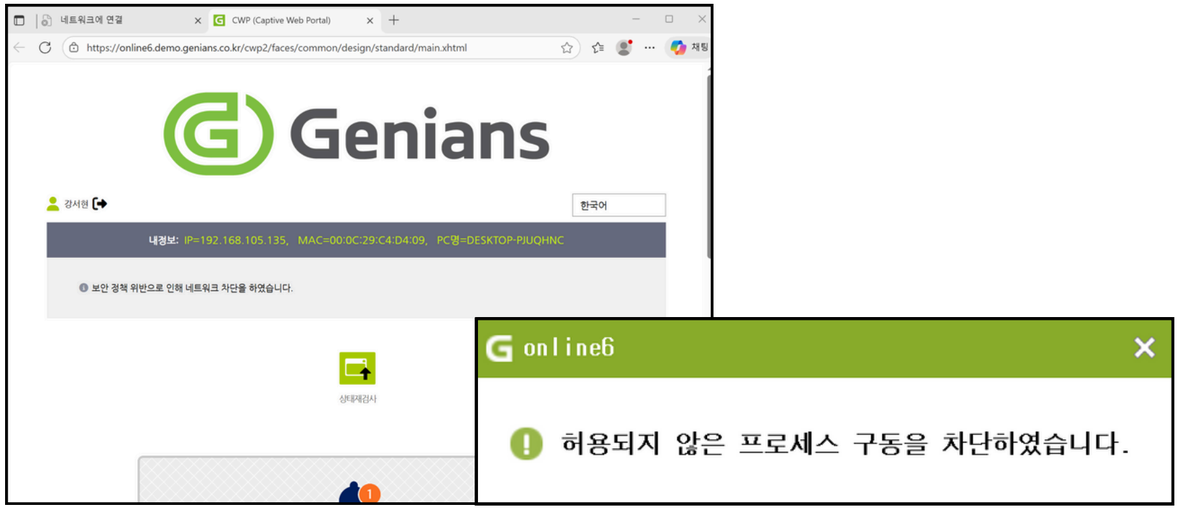
7-1 자동 대응 실행

5

대응 절차 설계 LLM 플래너

7

7-2 수동 대응(분석가 승인)



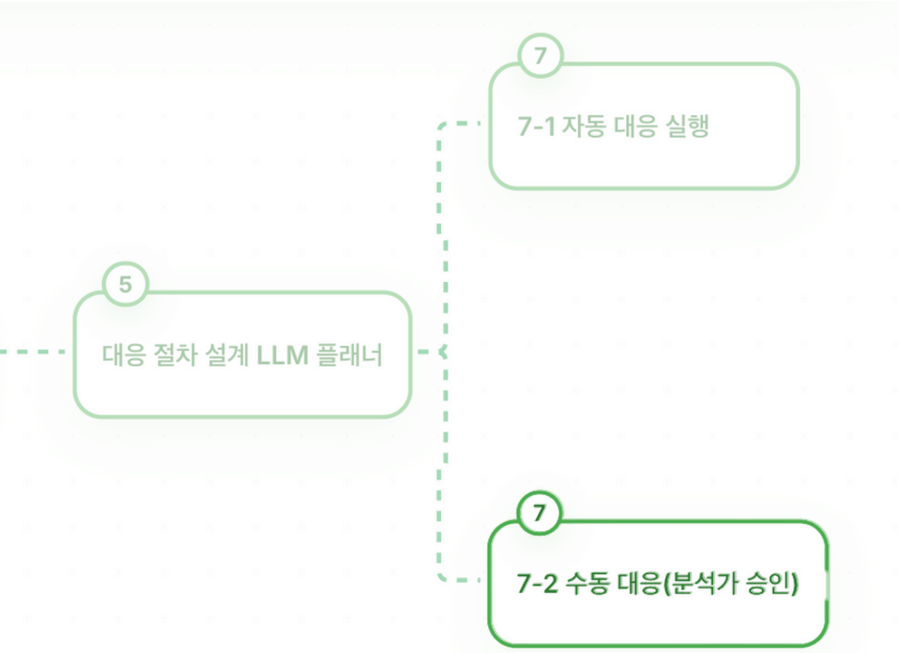
7-1 자동 대응 실행

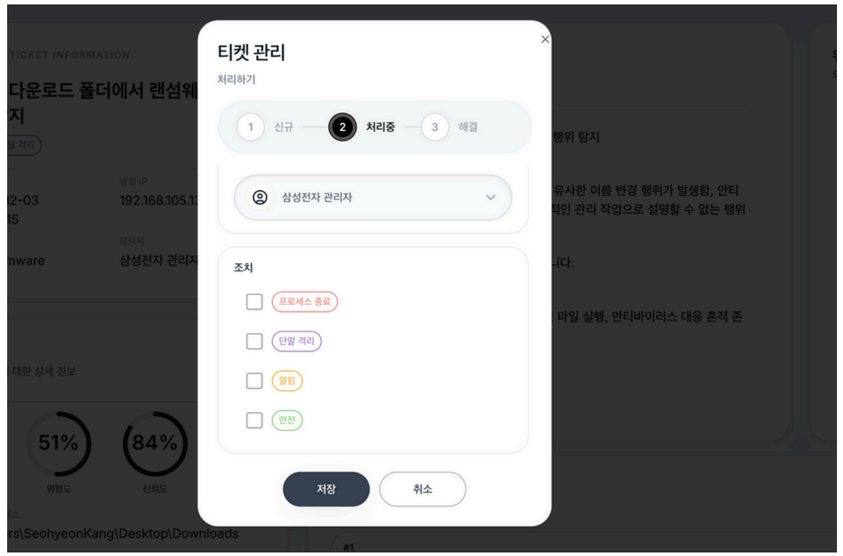
Flow Guide

위험도가 높고, 고객사 정책 상 자동 조치가 허용되는 경우 실제 조치를 수행하는 경로입니다.

· NAC/EDR API를 통해 프로세스 종료, 호스트 격리, 태그 부여 등의 액션을 자동 실행합니다.
조치 결과를 요약해 고객사별 Slack/알림 채널에 전송합니다.
- 출력 데이터: 수행된 자동 조치 내역, 조치 결과 로그

대응



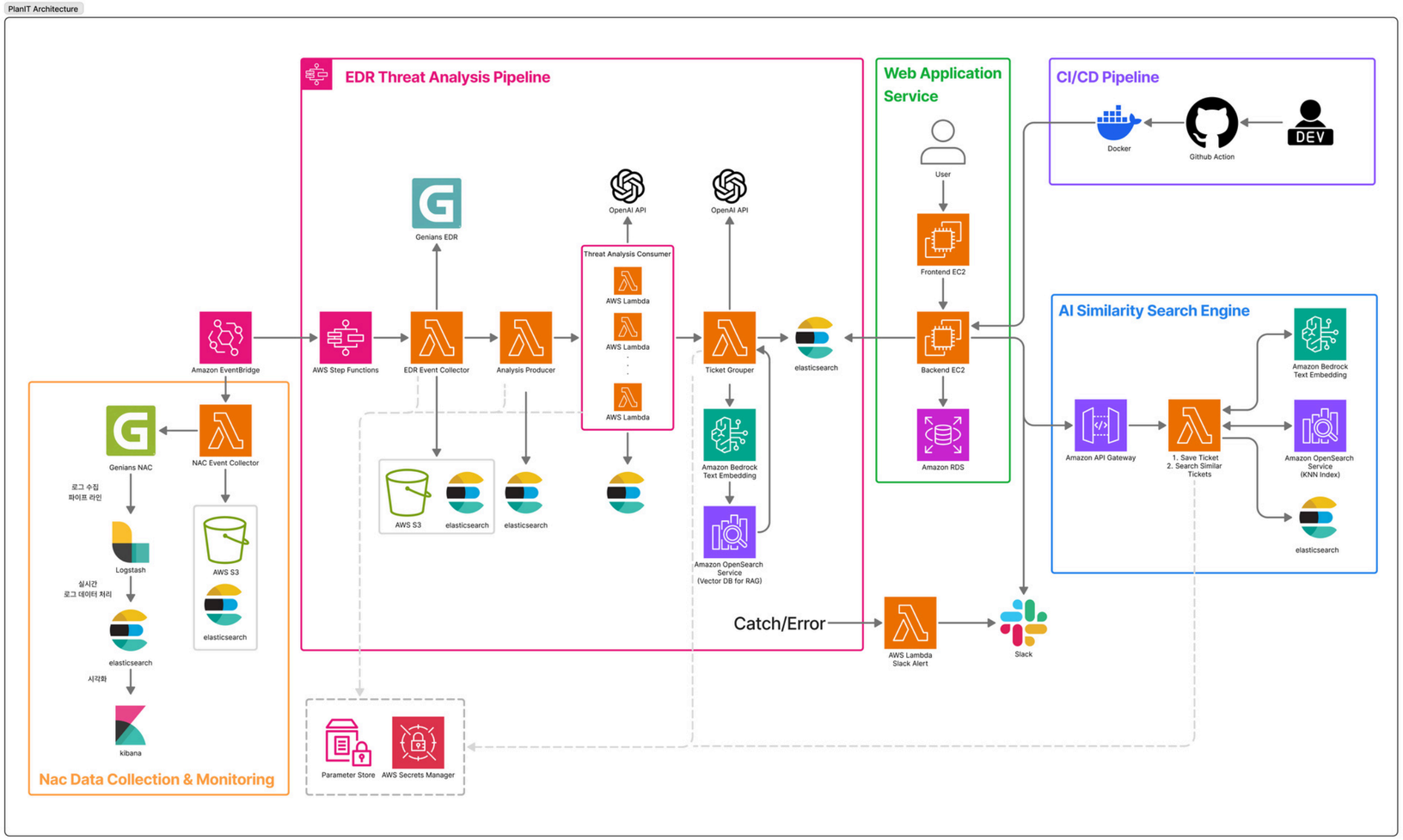


7-2 수동 대응(분석가 승인)

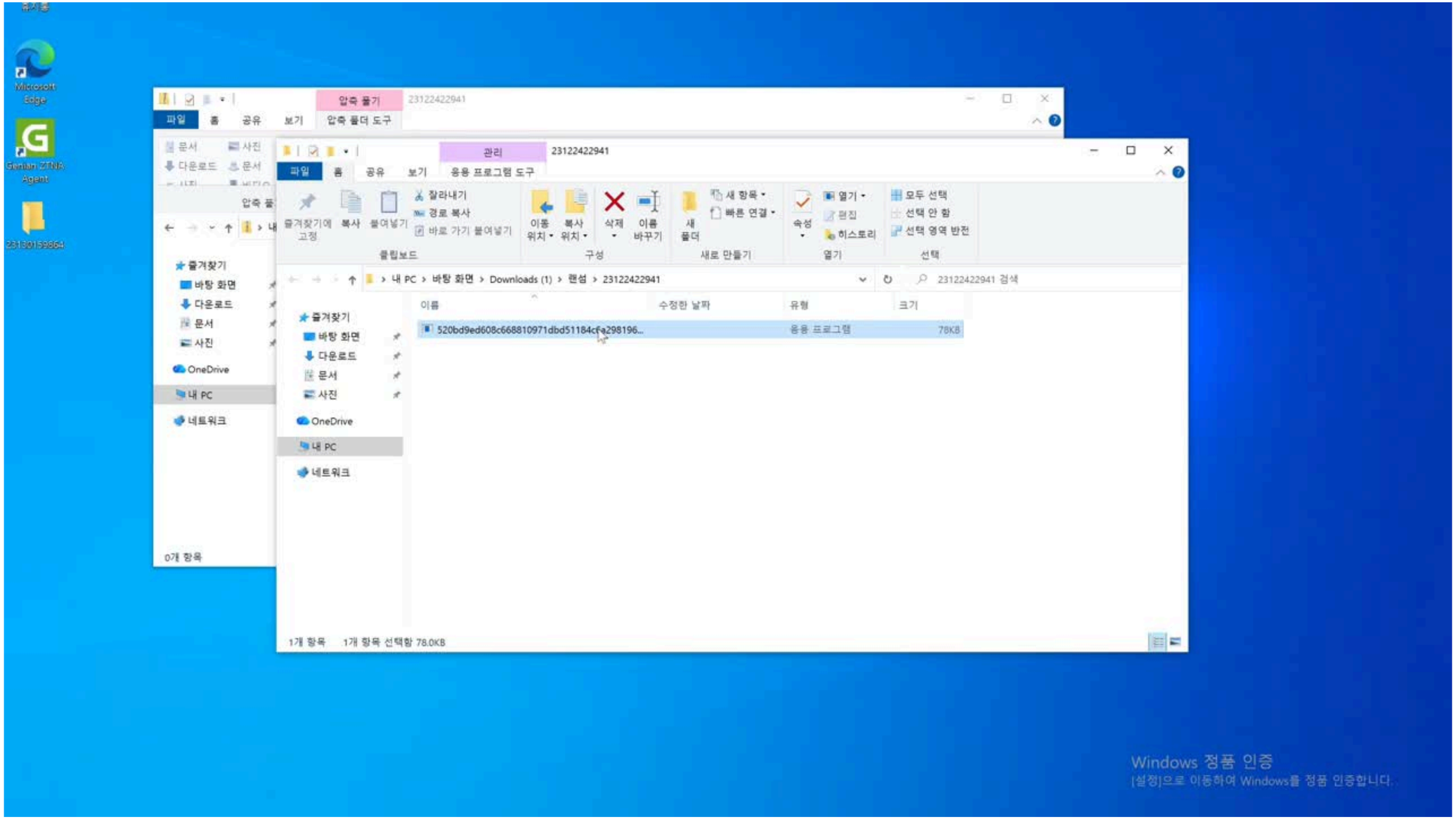
Flow Guide

위험도는 있지만 자동 조치가 부담되거나, 정책 상 반드시 사람 검토가 필요한 경우 수동 대응으로 넘기는 경로입니다.

서비스 아키텍처


PlanIT Architecture
EDR Threat Analysis Pipeline
Web Application Service
CI/CD Pipeline
AI Similarity Search Engine
Nac Data Collection & Monitoring
Amazon EventBridge
AWS Step Functions
EDR Event Collector
Genians EDR
Analysis Producer
Threat Analysis Consumer
AWS Lambda
AWS Lambda
AWS Lambda
OpenAI API
OpenAI API
Ticket Grouper
elasticsearch
AWS S3
elasticsearch
elasticsearch
Amazon Bedrock Text Embedding
Amazon OpenSearch Service (Vector DB for RAG)
User
Frontend EC2
Backend EC2
Amazon RDS
Docker
Github Action
DEV
Amazon API Gateway
1. Save Ticket
2. Search Similar Tickets
Amazon Bedrock Text Embedding
Amazon OpenSearch Service (KNN Index)
elasticsearch
Genians NAC
NAC Event Collector
로그 수집 파이프 라인
Logstash
실시간 로그 데이터 처리
elasticsearch
시각화
kibana
AWS S3
elasticsearch
Catch/Error
AWS Lambda Slack Alert
Slack
Parameter Store
AWS Secrets Manager

03. 시연영상


Microsoft Edge
Genian ZTNA Agent
23130159864
압축 풀기
23122422941
파일 홈 공유 보기 압축 폴더 도구
관리
응용 프로그램 도구
잘라내기
경로 복사
바로 가기 붙여넣기
즐겨찾기에 고정
복사
붙여넣기
이동 위치
복사 위치
삭제
이름 바꾸기
새 폴더
새 항목
빠른 연결
속성
열기
편집
히스토리
모두 선택
선택 안 함
선택 영역 반전
클립보드
구성
새로 만들기
열기
선택
내 PC > 바탕 화면 > Downloads (1) > 랜섬 > 23122422941
23122422941 검색
즐겨찾기
바탕 화면
다운로드
문서
사진
OneDrive
내 PC
네트워크
이름
수정한 날짜
유형
크기
520bd9ed608c668810971dbd51184cfa298196...
응용 프로그램
78KB
0개 항목
1개 항목
1개 항목 선택함 78.0KB
Windows 정품 인증
[설정]으로 이동하여 Windows를 정품 인증합니다.

04. AI 기반 악성/정상 분류 모델 성능 평가

테스트 개요 및 데이터셋 구성

- 모델: GPT-5
- 데이터셋 구성: 총 100건 (악성 50건, 정상 50건)
  - 악성: Atomic Red Team 및 MITRE ATT&CK 기반 모의 공격 시나리오 로그
  - 정상: 사내 개발팀의 일상적인 업무 로그 및 시스템 활동
- 입력 데이터: EDR Syslog, Threat API(단일 위협 이벤트 및 전후 60초 Context Log)

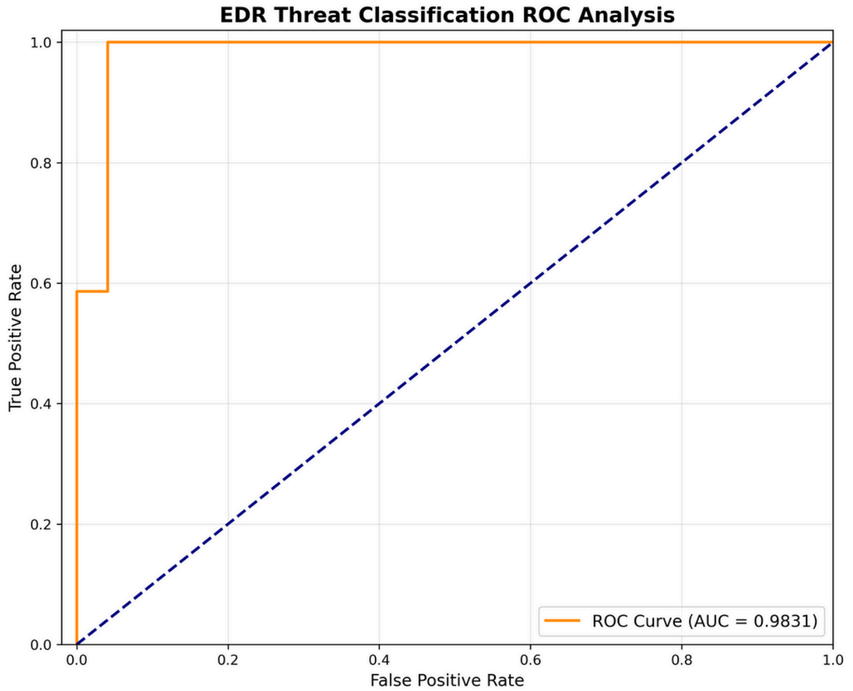
모델 성능 개선 및 최적화 과정

단계	적용 기법 (Technique)	정상 정확도 (Normal)	악성 정확도 (Malicious)	비고
Step 1	Logprobs 확률 계산	30.70%	98.30%	과도한 오탐 발생
Step 2	Self-Evaluation 및 프롬프트 최적화	89.00%	93.10%	오탐 획기적 개선
Step 3	논리적 검증 강화 (Devil's Advocate)	88.30%	94.80%	판단 논리 강화
Step 4	Syslog 데이터 추가 및 Context Filtering	96.80%	98.30%	핵심 성능 향상 구간
Step 5	최종 최적화 (Domain Knowledge 주입)	98.00%	100%	최종 모델 (미탐 0건)

성능 평가 결과

지표(Metric)	점수(Score)	비고
Accuracy	99%	전체적인 분류 정확도
Precision	98.04%	오탐 비율 제어 수준
Recall	100%	**공격 탐지 성공률(미탐 0건)**
F1-Score	99.01%	종합 성능 지표

[주요 성능 지표]



[ROC Curve]

05. AI 기반 위협 이벤트 그룹핑 성능 평가

테스트 개요

1. 테스트 목적 및 목표
   - Context 기반 자동 병합: 단일 탐지 이벤트를 '동일한 공격 맥락(Context)' 기준으로 식별하여 하나의 티켓으로 자동 그룹핑
   - 관제 효율성 제고: 보안 관제 요원이 분석해야 할 물리적 티켓 수 감소 확인
   - 공격 가시성 확보: 개별 이벤트가 아닌, 공격의 전체 흐름(Attack Flow) 파악 가능 여부 검증
2. 모델 아키텍처 (Model Architecture)
   - Embedding (벡터화): AWS Titan Embeddings v2
     - 텍스트 로그를 고차원 벡터로 변환하여 의미적 유사성 판단
   - Grouping Logic (병합 판단): GPT-5
     - RAG(검색 증강 생성) 기반 맥락 분석을 통해 최종 병합 결정
     -

시뮬레이션 데이터셋 기반 정량 평가

- 데이터셋: 총 100건 (Defense Evasion, Persistence, C&C 등 주요 공격 기법 포함)
- 검증 결과:
  - 총 62개의 실제 공격 사건을 63개의 그룹으로 분류

지표 (Metric)	점수 (Score)	의미 및 해석
ARI (Adjusted Rand Index)	**0.9889**	정답지와 AI 결과의 일치도 (1.0 만점). 무작위 수준을 훨씬 상회하는 완벽에 가까운 수치임.
Homogeneity (동질성/순도)	**1**	**하나의 AI 그룹 내에 다른 종류의 공격이 섞여 들어간 경우가 0%임. (오탐/혼재 없음)**
Completeness (완전성)	**0.9966**	하나의 실제 사건을 하나의 그룹으로 온전히 묶었는지 평가. 1건의 미세한 분할(Split) 발생으로 1.0 미달.

실제 운영 데이터 기반 검증

- 데이터셋: 실제 악성 공격(Trojan, Ransomware 등) 및 정상 개발자 활동이 혼재된 EDR 로그 41건
- 검증 결과:
  - 총 18개의 실제 공격 사건(True Cases)을 19개의 그룹으로 분류

지표 (Metric)	점수 (Score)	평가 및 해석
ARI (Adjusted Rand Index)	**0.9101**	정답과 AI의 그룹핑 결과가 매우 높은 수준으로 일치함.
Homogeneity (순도)	**1**	시뮬레이션 테스트와 마찬가지로 성격이 다른 사건이 섞인 사례가 **0건**임.
Completeness (완전성)	**0.9616**	단 1건의 공격 흐름이 끊긴 사례(Under-grouping)로 인해 일부 점수 하락.

결론

로그 검토 물량 약 42% 감소

05. 보안성 테스트

DAST 기반 소프트웨어 취약점 점검

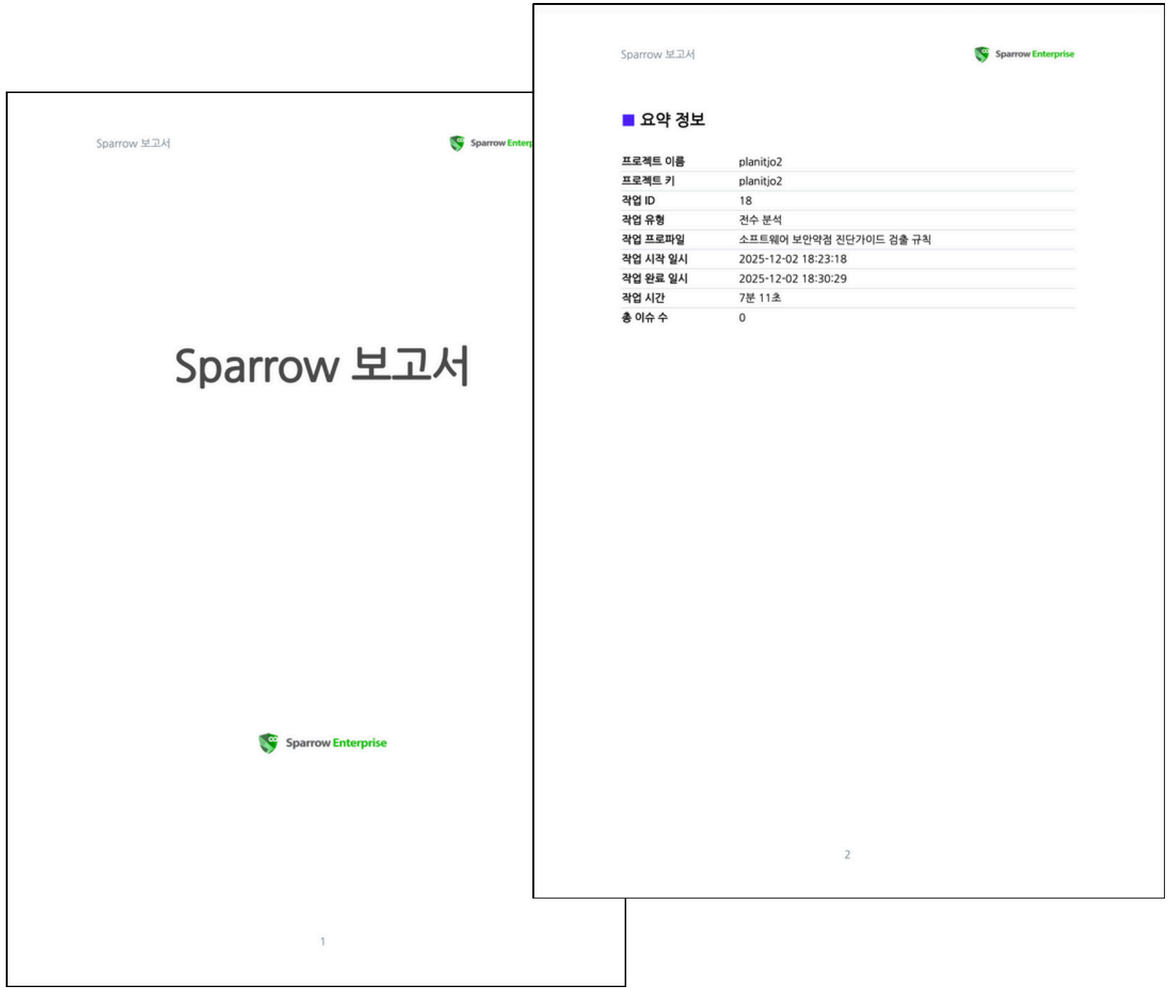
업계 표준 SW 취약점진단 가이드 12개 항목을 기준으로
자동 검사를 수행

주요 웹 취약점 규칙(12종) 무결성 확인

CRLF·SQL Injection·XSS 등 핵심 취약점 항목을 포함
전체 웹영역 점검

점검 결과 확인된 취약점 0건

분석 엔진 및 워크플로우 구현부에서 보안 취약점이
확인되지 않음

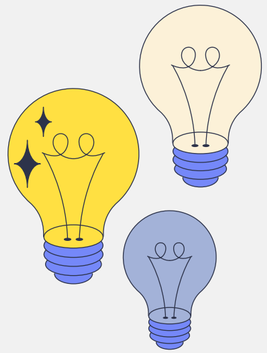


08. 차별화 포인트

지능형 티켓 그룹핑 & 유사도 기반 분석 지원

- RAG 기반으로 이벤트를 자동 클러스터링하여 분석 단위를 “로그”가 아닌 “인시던트” 중심으로 전환

- 벡터 임베딩을 활용한 유사 티켓 자동 추천으로 초보 분석가도 빠르게 대응 판단 가능



AI의 실질적 활용

완전히 새로운 모델 개발 대신,
이미 검증된 구조 + AI 접목으로 효율 극대화

현실적 가치 창출

- 실제 EDR·NAC 장비 API 연동으로 수집한 데이터 기반 분석·가공하여 현업 환경에 적용 가능한 구조 구축

- 단일 장비 대응이 아닌 다장비 연동 기반 자동화로 확장할 수 있는 실질적 가치를 제공

감사합니다