

OWASP Seoul Chapter

# **‘Security Operation in Future’: Agentic AI기반 위협탐지 자동화 PoC 구현사례 발표**

**Min Sung(Chris) Jung**  
**Security Architect**

# 00 whoami@owasp-seoul-oct-2025

- \* Min Sung (Chris) Jung
- \* Security Architect @ KT



- \* Focus areas: Use-cases implementation regarding AX in Cybersecurity

# Contents

1	생성형AI를 활용한 공격자 동향 및 대응방향	4
2	Agentic AI기반 위협탐지 자동화	8
2.1	공격 데모 시나리오 소개: 민감정보 유출	9
2.2	위협탐지/대응 자동화 구현: Sentinel Automation	12
2.3	AI 증강 자동화: Azure AI Foundry	15
2.4	Agentic AI for Security: MCP	18
3	결론	21

# 01 생성형AI를 활용한 공격자 동향

최근 생성형AI 기술 비약적 발전에 따른 공격자들의 해킹툴로 많은 활용사례가 보고되고 있음

## 정부 지원을 받는 위협 행위자의 Gemini 오용

### 정부 배후 공격자 동향

정부 지원 공격자들은 코딩 및 스크립팅 작업, 잠재적 표적에 대한 정보 수집, 공개적으로 알려진 취약점 연구, 그리고 표적 환경에서의 방어 회피와 같은 침해 후 활동을 가능하게 하기 위해 Gemini를 사용하려고 시도했습니다.

- **이란:** 이란 APT 행위자들은 Gemini를 가장 많이 사용했으며, 방어 조직 연구, 취약점 연구, 캠페인용 콘텐츠 제작 등 광범위한 목적으로 사용했습니다. APT42는 피싱 캠페인 제작, 방어 전문가 및 조직에 대한 정찰 수행, 사이버 보안 테마의 콘텐츠 생성에 집중했습니다.
- **중국:** 중국 APT 행위자들은 정찰, 스크립팅 및 개발, 코드 문제 해결, 그리고 표적 네트워크에 대한 더 심층적인 접근 권한을 얻는 방법 연구를 위해 Gemini를 사용했습니다. 그들은 측면 이동, 권한 상승, 데이터 유출 및 탐지 회피와 같은 주제에 집중했습니다.
- **북한:** 북한 APT 행위자들은 잠재적 인프라 및 무료 호스팅 제공업체 연구, 표적 조직 정찰, 페이로드 개발, 악성 스크립팅 및 회피 기술 지원을 포함하여 공격 수명 주기의 여러 단계를 지원하기 위해 Gemini를 사용했습니다. 그들은 또한 남한 군대와 암호화폐와 같이 북한 정부의 전략적 관심사에 대한 주제를 연구하기 위해 Gemini를 사용했습니다. 특히, 북한 행위자들은 서구 기업에 **비밀 IT 인력을 배치**하려는 북한의 노력을 지원할 가능성이 있는 활동인 자기소개서 작성 및 직업 연구를 위해 Gemini를 사용하기도 했습니다.
- **러시아:** 러시아 APT 행위자의 경우, 분석 기간 동안 Gemini 사용이 제한적으로 관찰되었습니다. 그들의 Gemini 사용은 공개적으로 사용 가능한 악성 코드를 다른 코딩 언어로 변환하고 기존 코드에 암호화 기능을 추가하는 것을 포함하여 코딩 작업에 집중되었습니다.

### North Korea's AI toolkit

Outside of their use of Gemini, North Korean cyber threat actors have shown a long-standing interest in AI tools. They likely use AI applications to augment malicious operations and improve efficiency and capabilities, and for producing content to support their campaigns, such as phishing emails. We assess with high confidence that North Korea has used AI to demonstrate an interest in the technology.

● This article is more than 1 year old

### North Korea and Iran using AI for hacking, Microsoft says

#### DPRK IT Workers

We have observed DPRK IT Workers using AI tools to generate phishing emails, social media posts, and other malicious content. Notably, a profile photo used by a North Korean actor to multiple different images on a social media platform to generate the threat actor's profile.

US tech giant says it has detected threats from foreign countries that used or attempted to exploit generative AI it had developed

#### APT43

Google Threat Intelligence Group (GTIG) has detected APT43 actor accessing multiple publicly available data sources. Based on the capabilities observed, it is possible these actors are using AI to generate phishing emails, lure content, and other malicious content.

- > GTIG has detected APT43 actor using AI to generate phishing emails, lure content, and other malicious content.
- > GTIG has identified APT43 actor using AI to generate phishing emails, lure content, and other malicious content.



Microsoft said in a blogpost that the techniques were 'early-stage' and 'not particularly novel or unique'. Photograph: Getty Images

1) Google: 생성형 AI의 두 얼굴: 악용 사례와 대응 전략(25.01.30)  
2) Guardian: North Korea and Iran using AI for hacking, Microsoft says(24.02.14)

# 01 보안분야 Agentic AI의 리스크 및 챌린지

Agentic AI의 발전에 따른 보안 역기능 우려 증대

- 1 **책임소재 모호성** ▶ AI 에이전트 자율성에 따른 의도치 않은 결과나 피해에 대한 책임소재 규명 어려움
- 2 **진화하는 위협환경** ▶ 공격자들 역시 Agentic AI 활용, 다단계 공격이라는 신규 공격 유형
- 3 **AI 고유 취약점** ▶ AI 에이전트는 과도한 권한, 프롬프트 인젝션 등 AI특화 신규 위협에 노출됨
- 4 **공급망 리스크** ▶ AI 에이전트 활용, 서드파티 API 및 서비스 상호작용 취약점 악용
- 5 **Human in the loop 약화** ▶ 자율 AI 시스템에 대한 과도한 의존에 따른 인간 경계심 저하, 보안팀 개입능력 감소
- 6 **데이터 프라이버시** ▶ 대량 데이터 처리에 따른 우발적 데이터 오남용 및 프라이버시 침해 우려
- 7 **윤리적 고려사항** ▶ 의사결정, 동의, 투명성, 잠재적 조작 가능성과 관련된 윤리적 딜레마 야기

# 01 사이버 보안 분야 Agentic AI 주요 Demands

생성형AI를 활용한 공격기법이 고도화 됨에 따라, 방어자(보안 담당자) 역시 AI에이전트 활용한 요구 증가

## Agentic AI 진화과정

초기 Agentic AI 시스템은 사전 정의 규칙과 레이블링된 데이터에 의존해 '적응력이 제한적'이었으나, 현대 Agentic AI는 LLM의 상식적 추론으로 새로운 상황에 유연하게 대처가 가능합니다.

이는 AI가 정해진 패턴을 넘어 '실제로 판단하고 행동할 수 있게 되었음'을



## AI에이전트 주요 보안 활용 분야<sup>1)</sup>:

- I. **자율 위협 탐지 및 대응:** 이상징후 자동식별 → 실시간 조사 → 즉각 조치 (예: 의심스러운 측면 이동 탐지 시, 자동격리)
- II. **사례 관리 자동화:** 보안사고의 분류, 추적, 해결 프로세스 자동화. 과거 사례 기반 대응 전략 제시
- III. **능동적 위협 사냥:** 시그니처 기반 탐지 넘어 숨겨진 위협 자율 탐색. 대규모 보안 데이터 분석, 신규 위협 식별
- IV. **취약점 관리 자동화:** 취약점 식별 → 패치 배포 → 설정 변경까지 전 과정 자동화
- V. **공격 보안 테스트:** 실제 사이버 공격 시뮬레이션. 자율 침투 테스트로 시스템 취약점 파악
- VI. **인력 역량 강화:** 반복 작업 처리로 보안 전문가의 전략적 업무 집중 지원
- VII. **공급망 보안:** 협력사 보안 상태 지속 모니터링. 통합 시스템 전반 보안정책 적용

## 01 [별첨] Microsoft vs. Google Security Agent Use-Cases



### Microsoft Security Copilot agents

**I. Phishing Triage:** 10 min. MTTR, 95% incidents resolved

**II.Alert Triage in DLP and IRM:** Overwhelming alert → Streamlined

**III.Conditional Access Optimization**

**IV.Vulnerability Remediation:** Prioritize and expediate critical patches

**V.Threat Intelligence Briefing:** TI mapped to you, real-time situational awareness



### Google Security SecOps agents

**I. Yara-L conversion**

**II.Investigation Summaries**

**III.Investigative Assistant** (multi-turn assistant to perform threat hunts and investigations or create rule)

**IV.Code insights** (AI-driven file analysis)

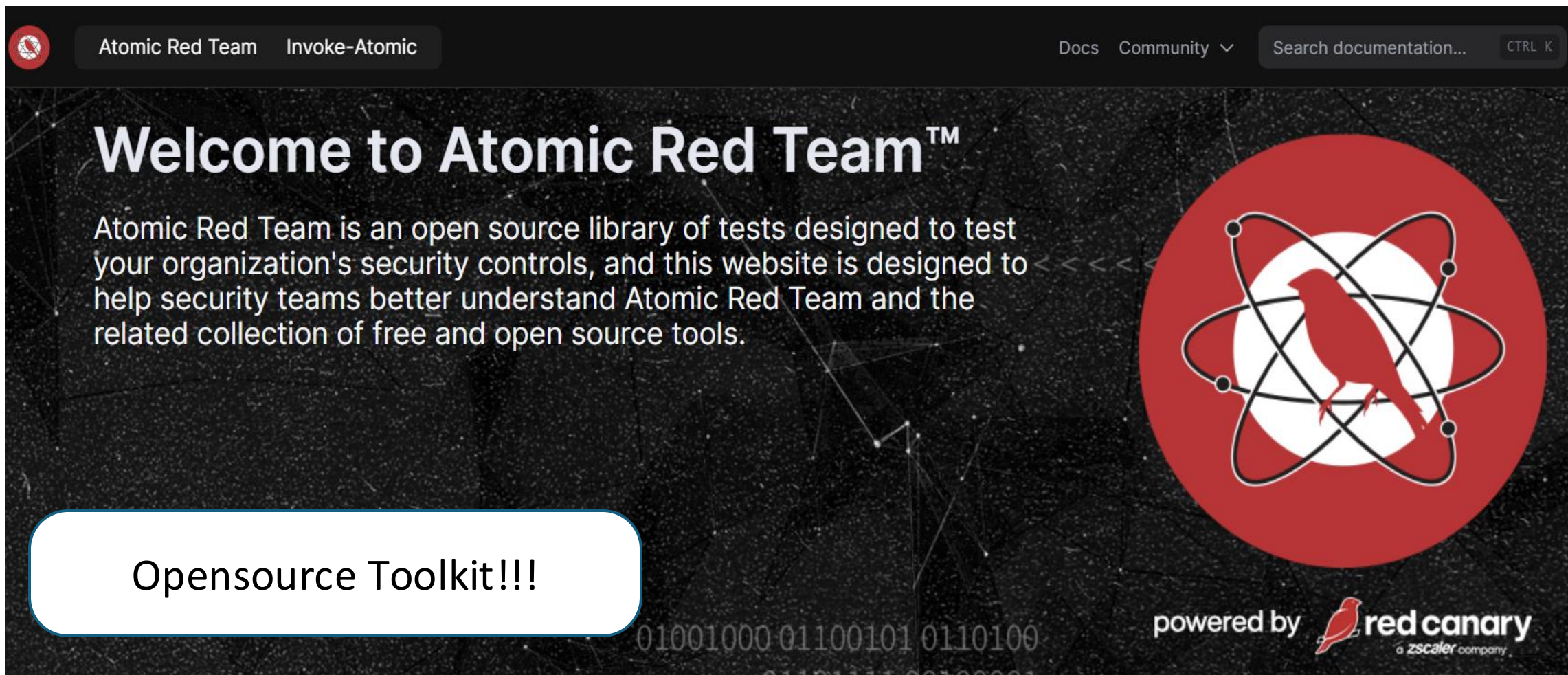
**V.Case Summaries and Playbook Generation**

1) Microsoft Security: Defense in the Age of AI with Microsoft Security Copilot agents  
2) Google Cloud: The Agentic SOC: Supercharging Security Operations with AI and Automation



## Part 1: 공격 도구 셋업 - Atomic Red Team

MITRE ATT&CK TTP기반 실행 범위가 명확하며, 효과적으로 안전한 모의침투 테스트가 가능함




The screenshot shows the Atomic Red Team website. The header includes the Atomic Red Team logo, navigation links for 'Atomic Red Team' and 'Invoke-Atomic', and a search bar with the text 'Search documentation...' and a 'CTRL K' shortcut. The main content area features the heading 'Welcome to Atomic Red Team™' and a paragraph: 'Atomic Red Team is an open source library of tests designed to test your organization's security controls, and this website is designed to help security teams better understand Atomic Red Team and the related collection of free and open source tools.' To the right of the text is a large red circular logo containing a white atomic symbol with a red cardinal bird perched on it. At the bottom left, there is a white rounded rectangle with the text 'Opensource Toolkit!!!'. The bottom right corner features the text 'powered by' followed by the 'red canary' logo and 'a zscaler company'.

Atomic Red Team Invoke-Atomic Docs Community ▾ Search documentation... CTRL K

# Welcome to Atomic Red Team™

Atomic Red Team is an open source library of tests designed to test your organization's security controls, and this website is designed to help security teams better understand Atomic Red Team and the related collection of free and open source tools.

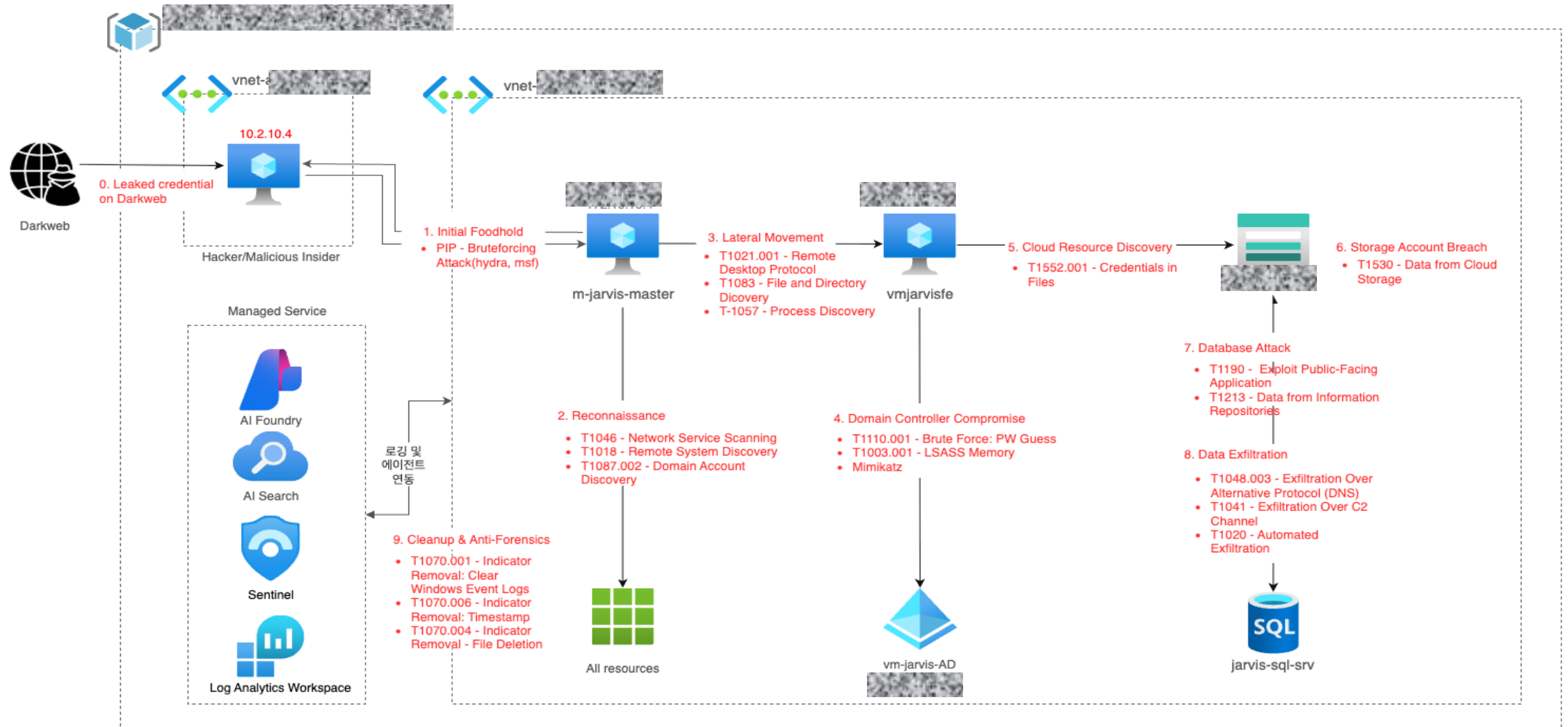
Opensource Toolkit!!!

powered by  red canary  
a zscaler company



## 02 Part 1: 공격 데모 시나리오 소개: 민감정보 유출

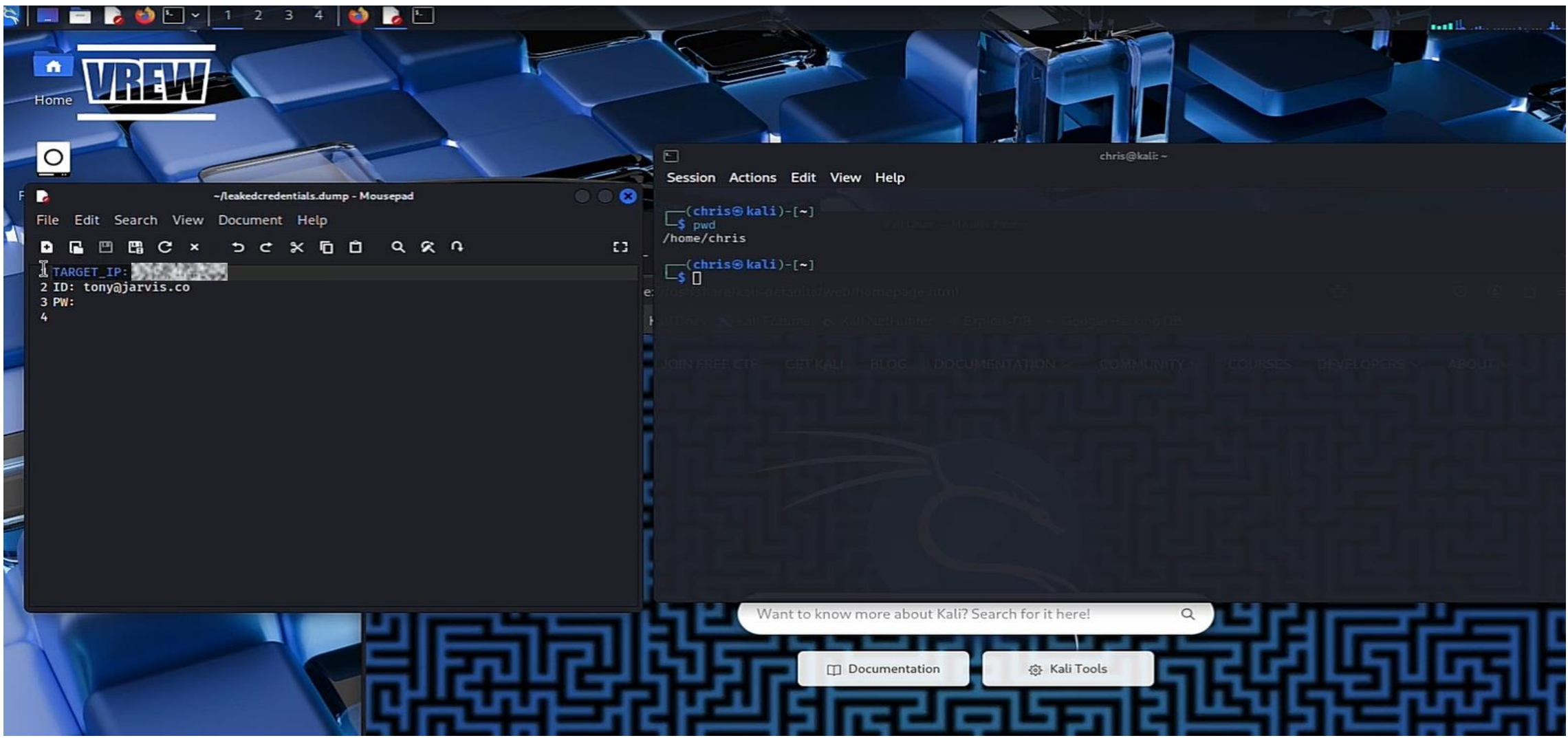
공격자는 AI S/W개발기업(Jarvis) 대상 무차별대입, 정찰, 계정탈취 공격을 통해 민감정보 유출을 시도함



1) Github Repo: <https://github.com/jmstar85/atomic-redteam-attack-scripts>

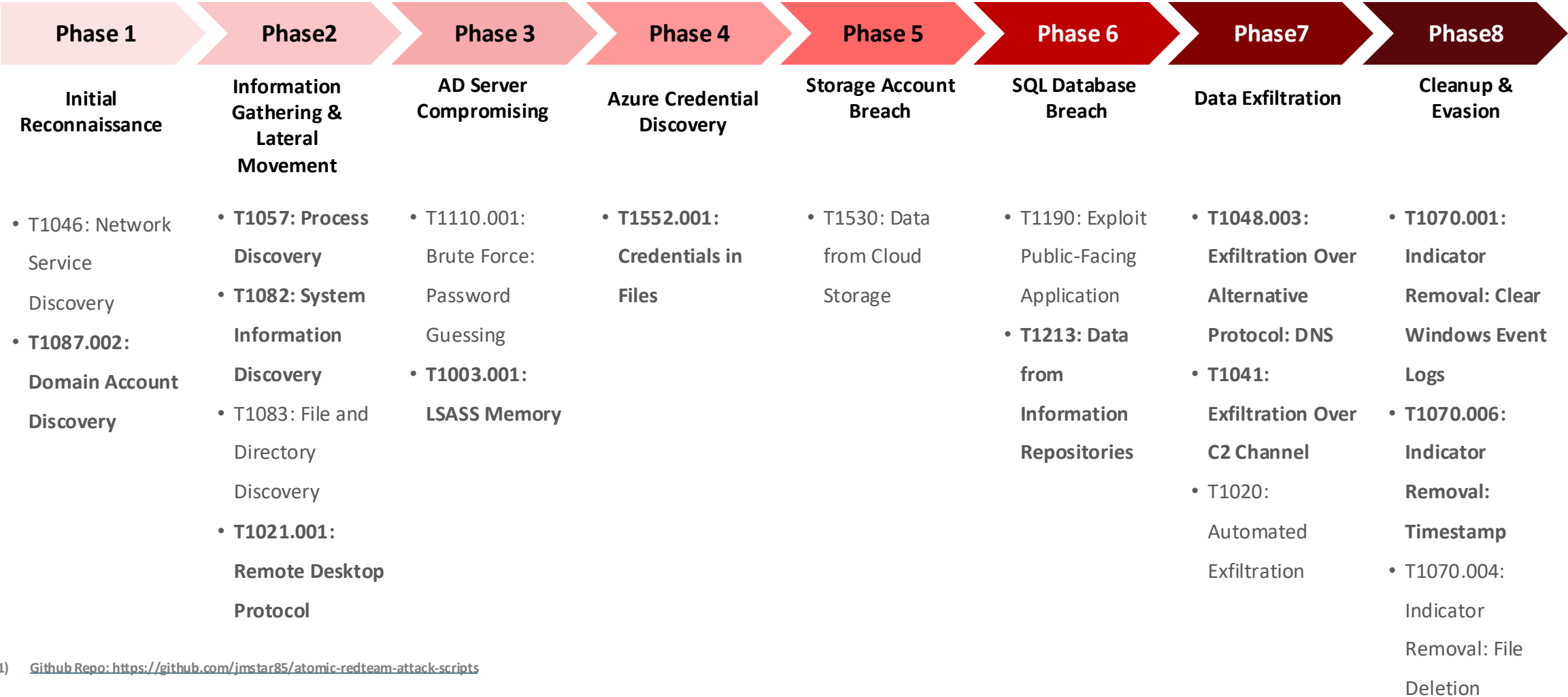
2) [공격 상세정보: 별첨]

# 02.01 Part 1: 공격 시나리오 소개: 민감정보 유출(Demo)



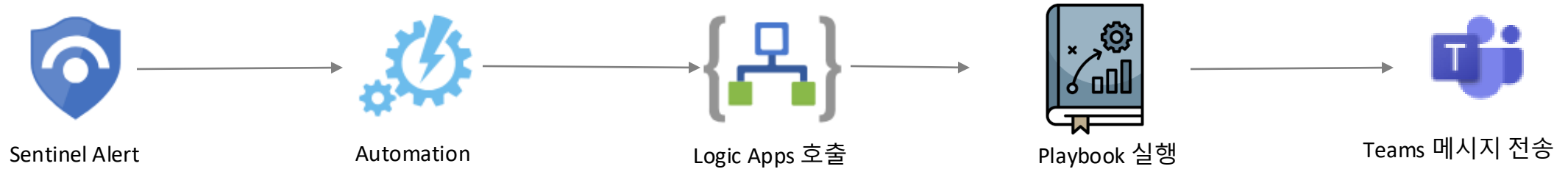
# 별첨: 공격 킬체인 및 MITRE ATT&CK 매핑

Open source인 Atomic Red Team 기반 정보유출 공격 시나리오 재현 및 Cyber Kill chain 기반 공격 구현



## 02.02 Part 2: 위협탐지/대응 자동화 구현 - Sentinel Automation

Sentinel Analytics Rule에 의해 트리거 된 incident는 사전 정의된 Automation에 따라 Playbook Callback하여 침해호스트 격리함




02.02 Part 2: 위협탐지/대응 자동화 구현 - Network Isolation Playbook

네트워크 격리 플레이북은 랜섬웨어(악성코드) 전파, 측면이동 감지, 계정정보 수집 등 高심각도 위협 발생 시 즉각적인 네트워크 격리 자동화 실행




## 02.02 Part 2: 위협탐지/대응 자동화 구현 - Network Isolation Playbook Demo

Workflows 오후 4:23

 **CRITICAL SECURITY ALERT**

**Incident:** [Demo] Mimikatz detected

**Compromised Host:** vm-jarvis-AD


**Source IP:** 


**Account:** Unknown

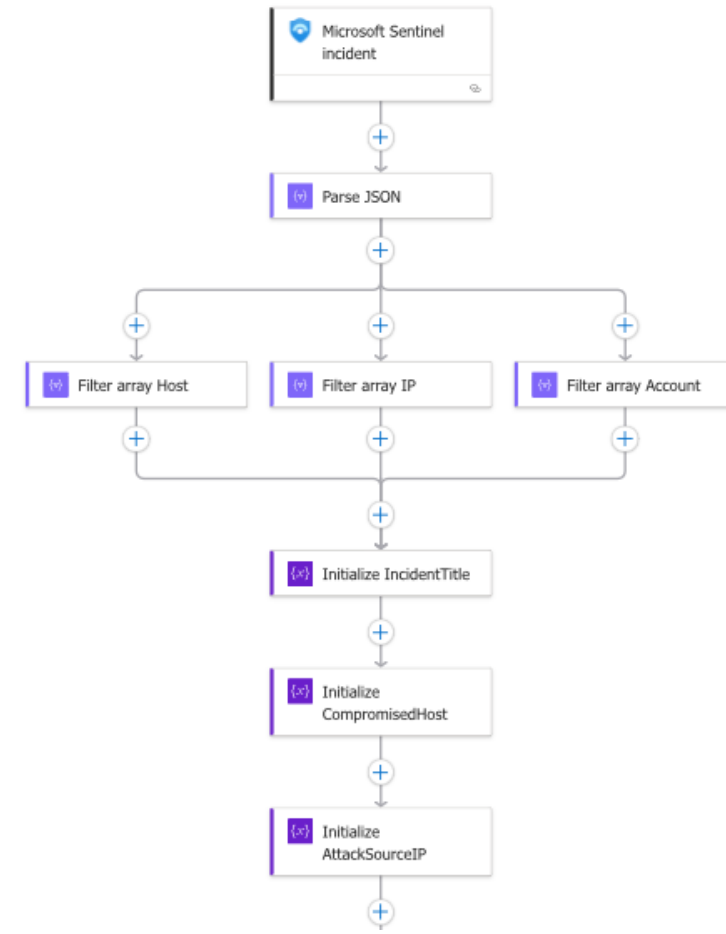
**Severity:** High

**Detection Time:** 2025-10-16 07:23:56

**Recommended Action:** Immediately isolate the compromised host to prevent lateral movement.

 **Approve Network Isolation**

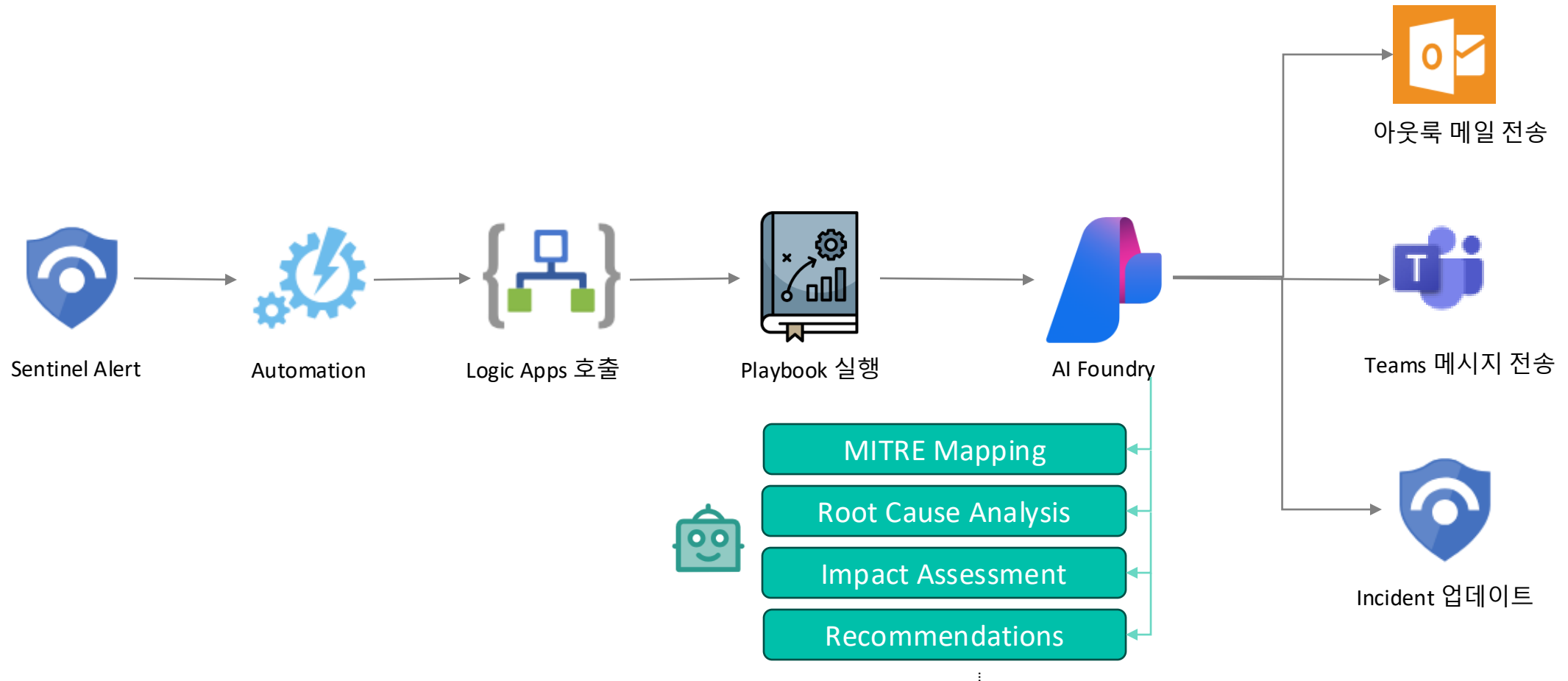
 **Reject (Manual Review)**





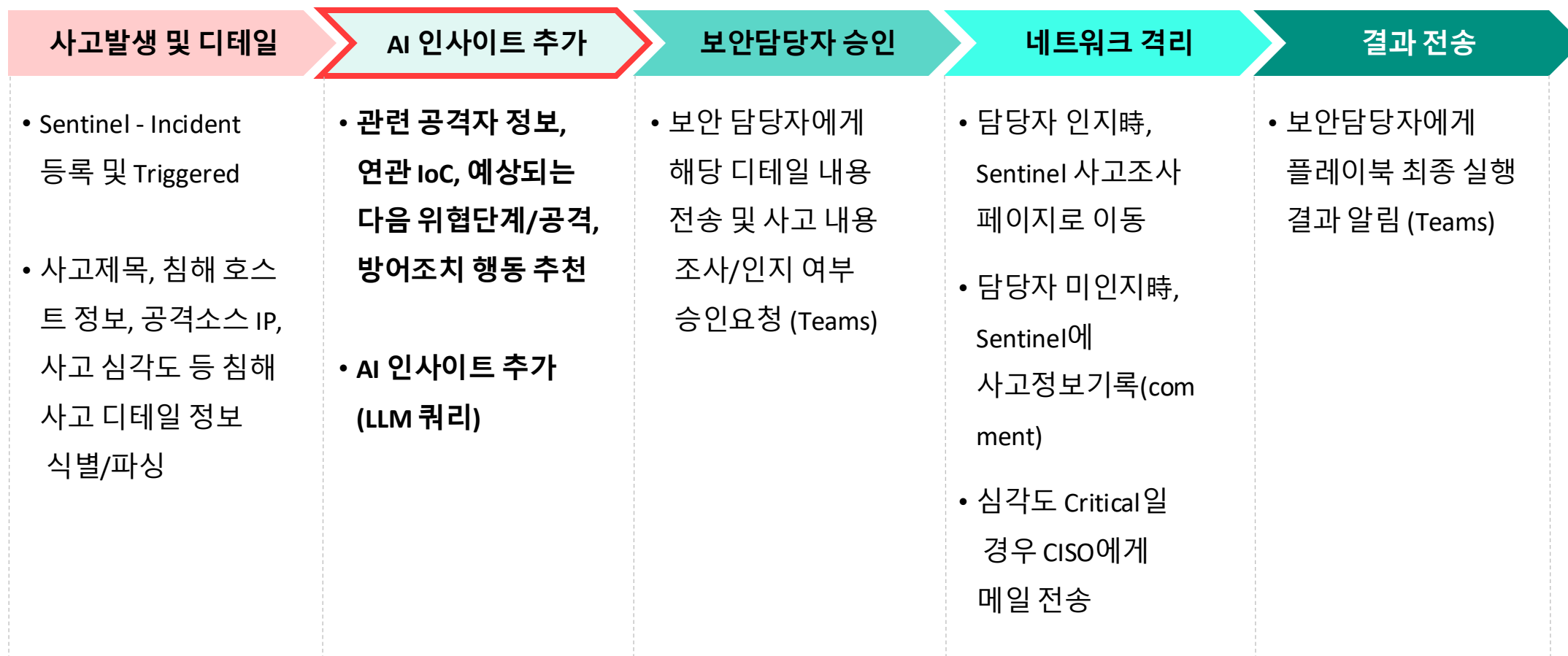
## 02.03 Part 3: AI 증강 자동화: Azure AI Foundry (SecOps Agent Playbook)

SecOps Agent Playbook은 incident 연관정보, 공격소스 관련 연관 정보(관련 공격자, 연관 IoC 등), 추천 대응 Actions, 비즈니스 임팩트 등 보안담당자에게 추가적인 인사이트를 제공함

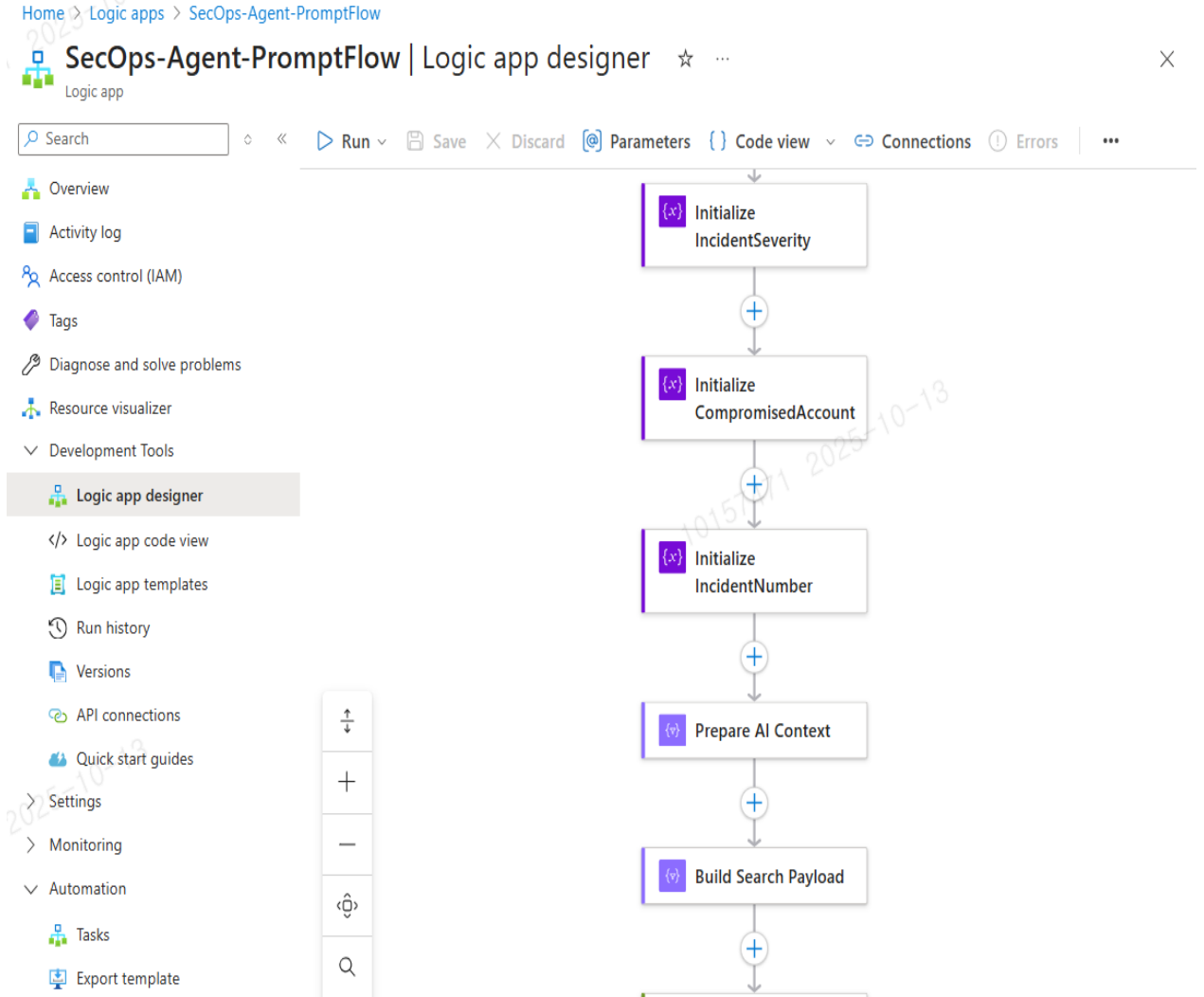
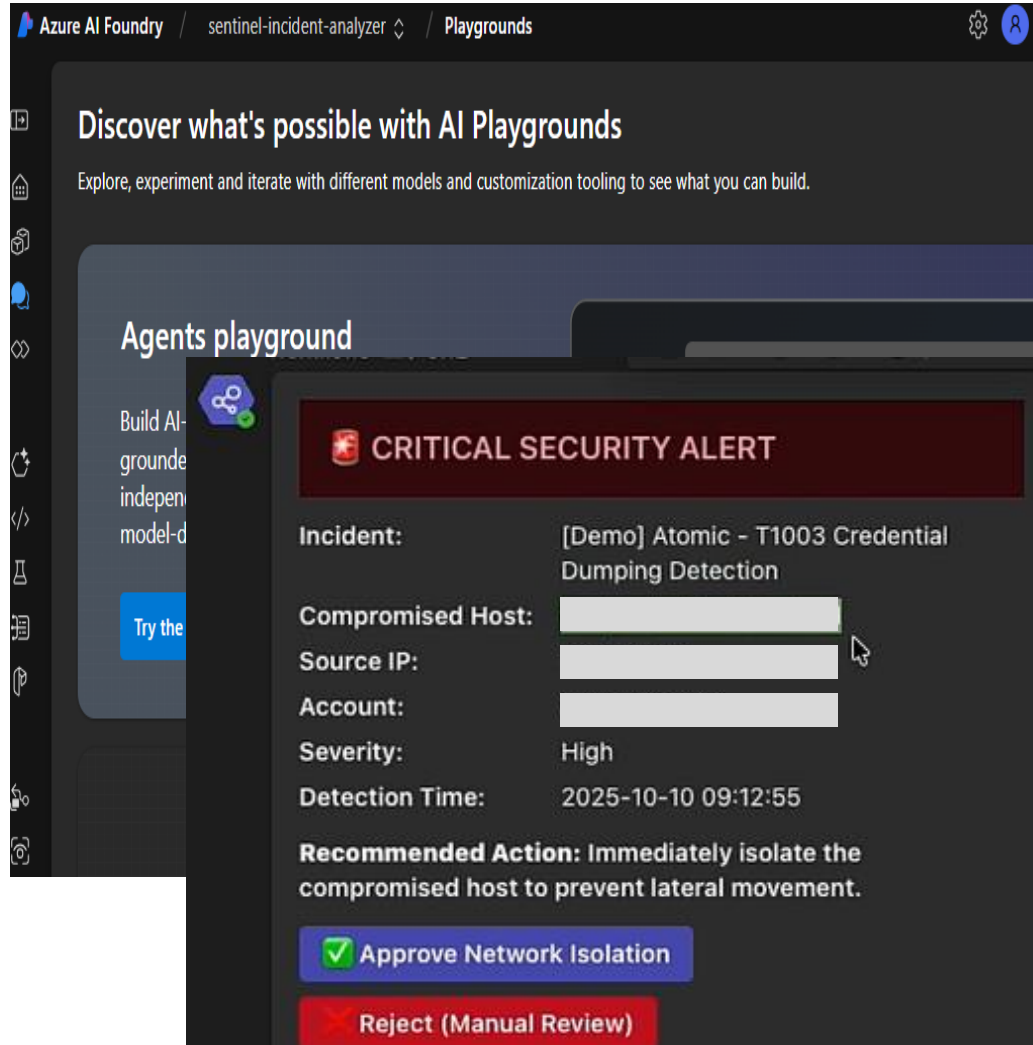


## 02.03 Part 3: AI 증강 자동화: Azure AI Foundry (SecOps Agent Playbook)

AI Foundry 내재된 LLM모델(gpt-5, gpt-4.1, grok-4, o3-pro, DeekSeek-v3 등)을 활용하여 분석가에게 인사이트 제공하여 최초 컨텍스트 이해용이

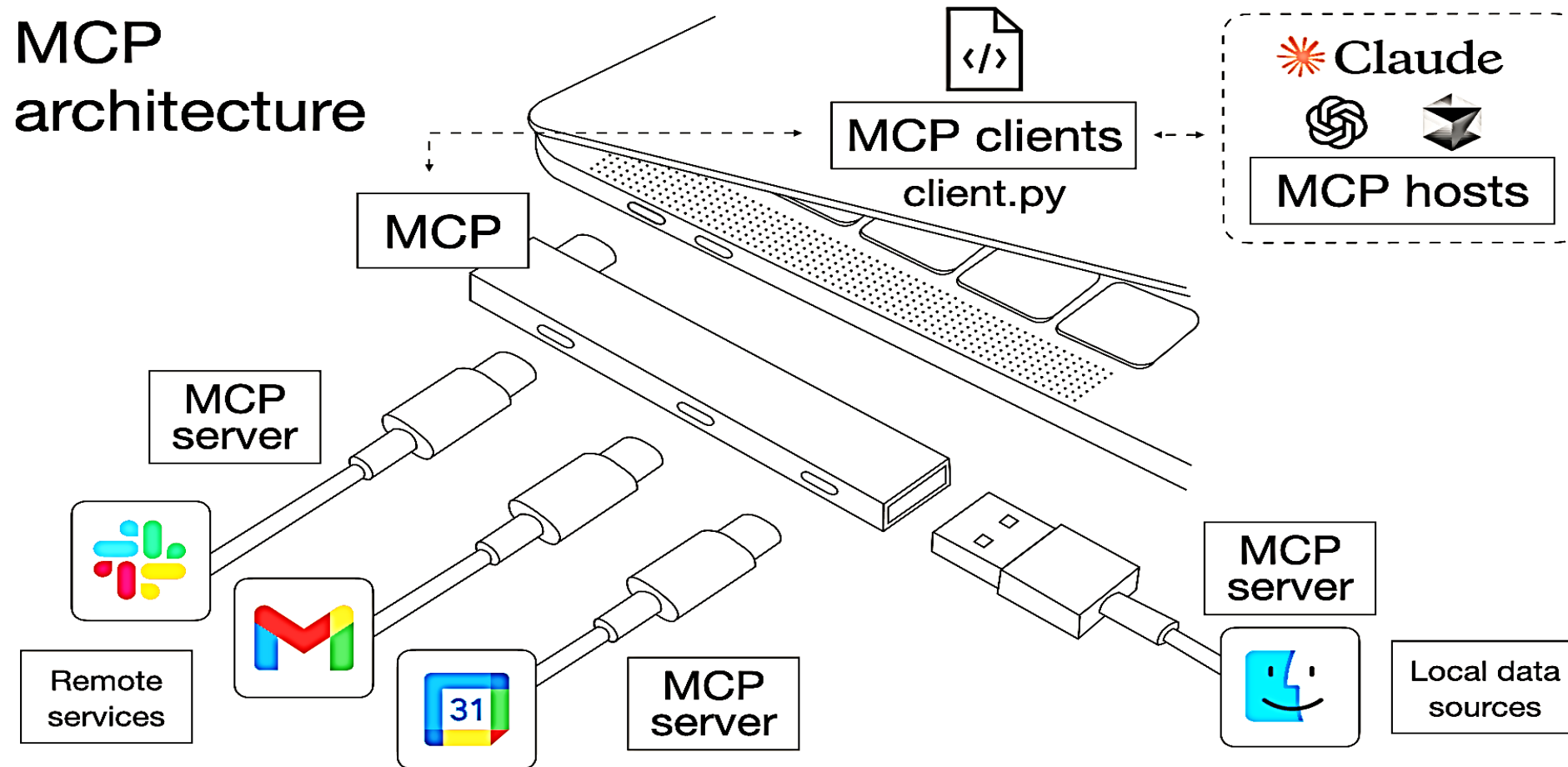


## 02.03 Part 3: AI 증강 자동화: Azure AI Foundry (SecOps Agent Playbook) Demo



## 02.04 Part 4: MCP (Model Context Protocol)

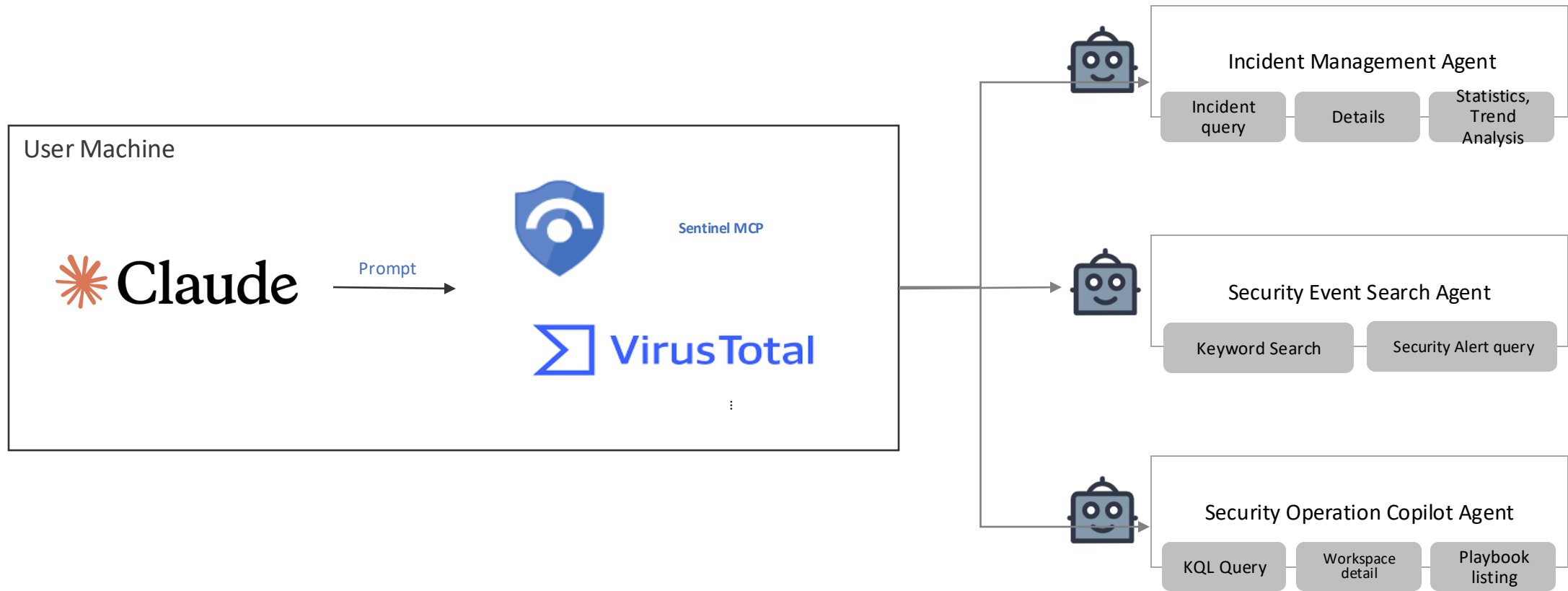
Anthropic社에서 개발, 다양한 외부 도구 연결된 AI 에이전트 ↔ LLM을 연결하는 표준 프로토콜



1) [What is MCP? How does it compare to API?](#)

## 02.04 Part 4: Agentic AI for Security: MCP (Model Context Protocol)

Sentinel MCP 서버<sup>1)</sup>는 보안담당자의 workspace와 연동하여 고도화 된 사고 관리기능(통계분석 등), 보안 이벤트 탐색 기능(서치 연계), 보안 운영 부조종사 기능(쿼리작성, 플레이북 실행 등)을 제공함



1) [Github repo: Azure Sentinel MCP Server \(Owned by Min Sung Jung\)](#)

## 02.04 Part 4: Agentic AI for Security Demo

Chris님, 오늘 하루 어떠셨나요?

오늘 어떤 도움을 드릴까요?

+ ⚙️ ⌚

Sonnet 4.5 ↗

🔍 메뉴 검색

📌 스타일 사용 >

🕒 심층 사고 모드 ☐

🔍 연구 ☐

🌐 웹 검색 ☒

🗺️ Drive 검색 연결 [🔗](#)

📧 Gmail 검색 연결 [🔗](#)

📅 Calendar 검색 연결 [🔗](#)

📌 azure-sentinel ☒ >

📌 virustotal ☒ >

+ 커넥터 추가

⚙️ 커넥터 관리 [🔗](#)

✎ 작성하기 🖨️ 일상생활

**주요 인시던트 (우선순위 높음)**

**1. 인시던트 #116 - 다수 엔드포인트 의심 활동 (HIGH)**

- 상태: New / 심각도: High
- 타임라인: 2025-10-09 11:35 ~ 현재 진행 중
- Alert 개수: 138개 ⚠️
- 영향 호스트: [REDACTED]
- 특징:
  - AI 기반 탐지 (Credential Access, Unknown patterns)
  - 네트워크 격리 거부 기록 3회 (정민성님이 수동 승인 거부)
  - 네트워크 격리 1회 성공 (2025-10-10 03:00 UTC)
  - NSG 규칙: IsolationRule-Emergency-Inbound/Outbound (Priority 100)

**3. 악성코드 & 파일 IoC**

**Mimikatz 위장 실행파일**

파일명: svchost.exe  
경로: C:\Windows\Temp\svchost.exe  
SHA256: 92804faaab2175dc501d73e814663058c78c0a042675a8937266357bcfb96c50  
설명: 정상 시스템 프로세스명을 위장한 Mimikatz 변종

**DCSync 공격 명령 파일**

파일명: dcsync\_commands.txt  
경로: C:\Windows\Temp\dcsync\_commands.txt  
설명: DCSync 공격을 위한 명령어 스크립트

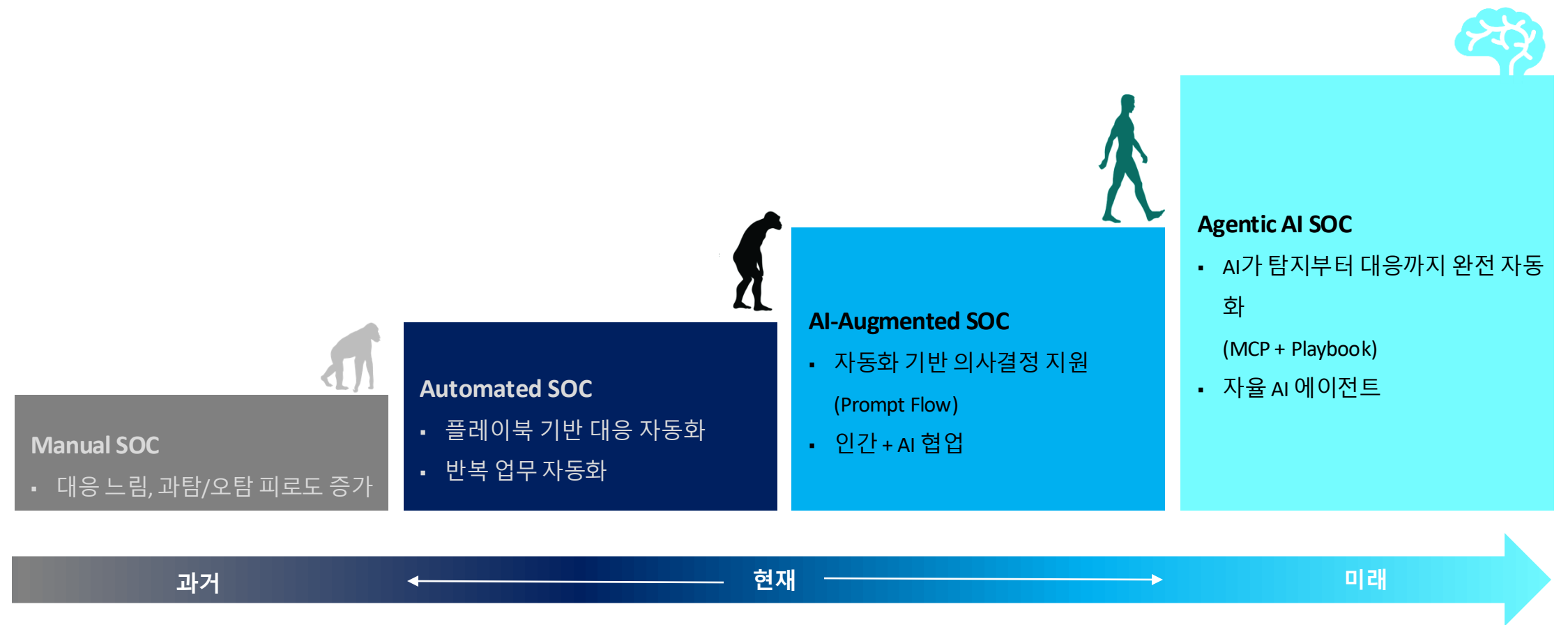
**의심스러운 실행 경로**

C:\Windows\Temp\svchost.exe  
[REDACTED]\Windows\Temp\svchost.exe (원격 네트워크 경로)



# 결론

Agentic AI는 보안분야 새로운 자동화 가능성을 열고 있지만, 동시에 본질적인 위험 요인도 내포하고 있음<sup>1)</sup>  
편의성과 보안의 tradeoff 관계에서 적절한, 최적의 밸런스를 지향하는 ‘**보안의 효율적 완화**’ 필요



1) Gartner: Emerging Tech: The Future of Agentic AI in Enterprise Applications

Thank you!



OWASP<sup>TM</sup>

Seoul Chapter

AttackScripts Repo:



Sentinel MCP Repo:

