



**DARKTRACE**

# 자가 학습 AI로 사이버 신뢰를 구축하는 방법

강기호

/ Account Executive 이사

# 생성형 AI

- 생산성 향상 및 인간의 창의성 보강
- 고객의 74%가 일터에서 생성형 AI 도구를 사용 중
- IP 침해 및 데이터 유출에 대한 우려
- 공격 정교화에 따른 보안 리스크

# 생성형 AI: CISO 반응

## 불안

“정보유출 방지를  
위해 차단을 고민중”

"공격자들이 어떻게  
공격할지?"



## 기대

“어떻게 빠르게 인프라에  
통합할 수 있을까?”

“전략적 이점을 위해 어떻게  
활용할 수 있을까?”

**DARKTRACE**

© 2023 Darktrace (DARK.L). All rights reserved. | darktrace.com

# 생성형 AI : 사이버 공격에 대한 영향

↑135%

신규 공격

2023년 1월~2월 사이  
늘어난 새로운 소셜  
엔지니어링 공격

Generative AI: Impact on Email Cyber-Attacks 리포트:  
[darktrace.com/GenAIReport](https://darktrace.com/GenAIReport)



**DARKTRACE**

© 2023 Darktrace (DARK.L). All rights reserved. | [darktrace.com](https://darktrace.com)



# 보안 팀에게 생성형 AI란?

**DARKTRACE**

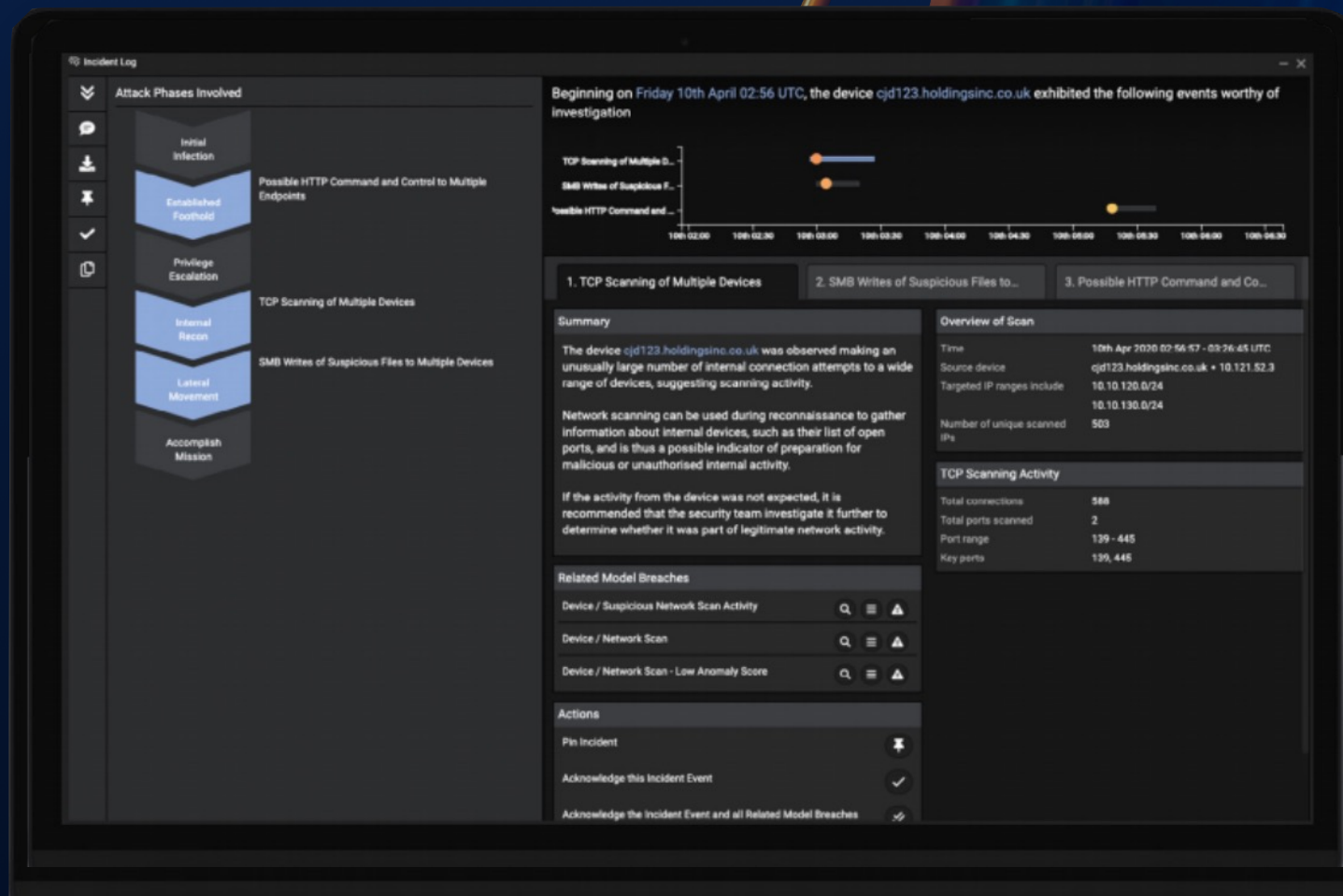
© 2023 Darktrace (DARK.L). All rights reserved. | [darktrace.com](https://darktrace.com)

# 사이버 보안 분야의 AI

	알려진 공격 데이터	고객사의 비즈니스 데이터	대규모 언어 모델
관찰 대상:	공격자 패턴	일상적인 비즈니스 패턴	인간 언어 사용 패턴
AI 종류:	지도 학습	자가 학습++	딥러닝
용례:	<ul style="list-style-type: none"><li>1. 알려진 공격</li><li>2. 커뮤니티 지식</li><li>3. 포렌식</li></ul>	<ul style="list-style-type: none"><li>1. 새로운 공격</li><li>2. 국가 및 정부</li><li>3. AI 공격</li><li>4. 내부자 공격/ 인적 오류</li></ul>	<ul style="list-style-type: none"><li>1. 내부 교육</li><li>2. 인간 사칭</li><li>3. 표적 공격</li></ul>

# Prompts & Outcomes Reflect Bias

내부 장치에서 흔히  
않은 외부  
엔드포인트에 HTTP  
연결을 시도 ...



# 질문과 결과는 편향을 반영

내부 장치에서 희귀한  
외부 엔드포인트에  
HTTP 연결을 시도 ...

## 보안 분석가

### > 이것은 악성인가?

> 장치가 희귀한 외부 엔드포인트에 HTTP 연결을 시도했습니다. 희귀한 엔드포인트에 HTTP 연결을 시도하는 디바이스는 여러 가지 이유로 악의적일 수 있습니다. 몇 가지 가능한 설명은 다음과 같습니다.:

- **Command and control (C2) 통신:** 장치가 명령 및 제어 서버와 통신하여 명령을 수신하거나 데이터를 유출하거나 추가 페이로드를 다운로드하는 멀웨어에 감염되었을 수 있습니다.
- **피싱 또는 사기 웹사이트:** 로그인 자격 증명이나 개인 데이터와 같은 민감한 정보를 훔치도록 설계된 피싱 또는 사기 웹사이트와 관련이 있을 수 있습니다.
- **익스플로잇 키트 또는 드라이브 바이 다운로드:** 디바이스 소프트웨어의 취약점을 악용하여 멀웨어 설치를 시도하는 익스플로잇 키트를 호스팅하는 악성 웹사이트에 연결되었을 수 있습니다.
- **데이터 유출:** 희귀 엔드포인트는 개인 또는 금융 정보, 지적 재산 또는 기밀 문서와 같은 민감한 데이터를 장치에서 유출하는 목적지가 될 수 있습니다.
- **봇넷 참여:** 장치가 봇넷의 일부일 수 있으며, 희귀 엔드포인트에 연결하여 분산 서비스 거부 공격 또는 기타 악의적인 활동을 수행하기 위한 지시를 받을 수 있습니다.

잠재적인 악성 연결을 조사하고 완화하려면 다음과 같이 할 수 있습니다:

- 디바이스와 소유자를 식별합니다.
- 네트워크에서 디바이스를 분리하고 악성코드 검사를 철저히 수행합니다.
- 로그를 검토하거나 네트워크 모니터링 도구를 사용하여 연결 및 데이터 전송을 분석합니다.
- 방화벽 또는 네트워크 보안 설정에서 드문 엔드포인트를 차단하여 추가 연결을 방지합니다.
- 네트워크 관리자 또는 IT 부서에 연락하여 상황 처리에 대한 도움을 요청하세요.
- 강력한 비밀번호를 사용하고, 방화벽을 활성화하고, 디바이스를 최신 보안 패치로 업데이트하여 안전한 네트워크를 유지하세요.



# 질문과 결과는 편향을 반영

내부 장치에서 희귀한  
외부 엔드포인트에  
HTTP 연결을 시도 ...

## 보안 분석가

### > 이것은 악성인가?

> 장치가 흔치 않은 외부 엔드포인트에 HTTP 연결을 시도했습니다. 흔치 않은 외부 엔드포인트에 HTTP 연결을 시도하는 디바이스는 여러 가지 이유로 악의적일 수 있습니다. 몇 가지 가능한 설명은 다음과 같습니다.:

- **Command and control (C2) 통신**: 장치가 명령 및 제어 서버와 통신하여 명령을 수신하거나 데이터를 유출하거나 추가 페이로드를 다운로드하는 멀웨어에 감염되었을 수 있습니다.
- **피싱 또는 사기 웹사이트**: 로그인 자격 증명이나 개인 데이터와 같은 민감한 정보를 훔치도록 설계된 피싱 또는 사기 웹사이트와 관련이 있을 수 있습니다.
- **익스플로잇 키트 또는 드라이브 바이 다운로드**: 디바이스 소프트웨어의 취약점을 악용하여 멀웨어 설치를 시도하는 익스플로잇 키트를 호스팅하는 악성 웹사이트에 연결되었을 수 있습니다.
- **데이터 유출**: 희귀 엔드포인트는 개인 또는 금융 정보, 지적 재산 또는 기밀 문서와 같은 민감한 데이터를 장치에서 유출하는 목적지가 될 수 있습니다.
- **봇넷 참여**: 장치가 봇넷의 일부일 수 있으며, 희귀 엔드포인트에 연결하여 분산 서비스 거부 공격 또는 기타 악의적인 활동을 수행하기 위한 지시를 받을 수 있습니다.

잠재적인 악성 연결을 조사하고 완화하려면 다음과 같이 할 수 있습니다:

- 디바이스와 소유자를 식별합니다.

\* 네트워크에서 디바이스를 분리하고 철저한 악성코드 검사를 수행하세요.

- 강력한 비밀번호를 사용하고, 방화벽을 활성화하고, 디바이스를 최신 보안 패치로 업데이트하여 안전한 네트워크를 유지하세요.

# 질문과 결과는 편향을 반영

내부 장치에서 희귀한  
외부 엔드포인트에  
HTTP 연결을 시도 ...

## IT 엔지니어

### > 이것은 허용된것 인가?

장치가 희귀한 엔드포인트에 HTTP 연결을 시도했습니다. 장치가 희귀 엔드포인트에 HTTP 연결을 하는 데에는 몇 가지 합법적인 이유가 있을 수 있습니다.

몇 가지 가능한 설명은 다음과 같습니다:

- **내부 또는 사용자 지정 애플리케이션**: 디바이스가 널리 사용되지 않거나 공개적으로 액세스할 수 없는 내부 또는 사용자 지정 웹 애플리케이션에 연결하고 있을 수 있습니다. 이는 자체 소프트웨어를 개발하거나 특수 도구를 사용하는 조직에서 흔히 발생할 수 있습니다.
- **테스트 또는 개발 환경**: 프로덕션 환경만큼 널리 알려지거나 액세스되지 않는 테스트 또는 개발 환경과 관련된 연결일 수 있습니다. 틈새 서비스 또는 웹사이트: 장치가 사용자 기반이 적은 틈새 서비스 또는 웹사이트에 연결되어 엔드포인트가 덜 일반적일 수 있습니다.
- **소프트웨어 업데이트 또는 원격 측정**: 일부 소프트웨어 애플리케이션 또는 운영 체제는 업데이트 또는 원격 분석 데이터 전송을 위해 자주 사용되지 않는 엔드포인트에 연결할 수 있습니다.
- **예약된 작업 또는 자동화**: 장치에서 데이터 검색 또는 처리를 위해 특정 엔드포인트에 연결하는 예약된 작업 또는 자동화 스크립트를 실행하고 있을 수 있습니다.

연결의 적법성을 확인하려면 다음과 같이 하세요:

- 장치와 소유자를 확인합니다.
- 연결 목적을 확인하고 디바이스의 사용 목적과 일치하는지 확인합니다.
- 로그를 검토하거나 네트워크 모니터링 도구를 사용하여 연결 및 데이터 전송을 분석하세요.
- 그래도 확실하지 않은 경우 네트워크 관리자나 IT 부서에 도움을 요청하세요.
- 강력한 비밀번호를 사용하고, 방화벽을 활성화하고, 디바이스를 최신 보안 패치로 업데이트하여 안전한 네트워크를 유지하는 것을 잊지 마세요.

# 질문과 결과는 편향을 반영

내부 장치에서 희귀한  
외부 엔드포인트에  
HTTP 연결을 시도 ...

## IT 엔지니어

> Is this legitimate?

장치가 희귀한 엔드포인트에 HTTP 연결을 시도했습니다. 장치가 희귀 엔드포인트에 HTTP 연결을 하는 데에는 몇 가지 합법적인 이유가 있을 수 있습니다.

몇 가지 가능한 설명은 다음과 같습니다:

- **내부 또는 사용자 지정 애플리케이션**: 디바이스가 널리 사용되지 않거나 공개적으로 액세스할 수 없는 내부 또는 사용자 지정 웹 애플리케이션에 연결하고 있을 수 있습니다. 이는 자체 소프트웨어를 개발하거나 특수 도구를 사용하는 조직에서 흔히 발생할 수 있습니다.
- **테스트 또는 개발 환경**: 프로덕션 환경만큼 널리 알려지거나 액세스되지 않는 테스트 또는 개발 환경과 관련된 연결일 수 있습니다. 틈새 서비스 또는 웹사이트: 장치가 사용자 기반이 적은 틈새 서비스 또는 웹사이트에 연결되어 엔드포인트가 덜 일반적일 수 있습니다.
- **소프트웨어 업데이트 또는 원격 측정**: 일부 소프트웨어 애플리케이션 또는 운영 체제는 업데이트 또는 원격 분석 데이터 전송을 위해 자주 사용되지 않는 엔드포인트에 연결할 수 있습니다.
- **예약된 작업 또는 자동화**: 장치에서 데이터 검색 또는 처리를 위해 특정 엔드포인트에 연결하는 예약된 작업 또는 자동화 스크립트를 실행하고 있을 수 있습니다.

연결의 적법성을 확인하려면 다음과 같이 하세요:

\*연결 목적을 확인하고 검증...

관련한 비밀번호를 사용하여 비밀번호를 확인하거나 디바이스를 최신 보안 패치로 업데이트하여 알려진 네트워크를 연결하는 것을

\*로그를 검토하거나 네트워크 모니터링 도구를 사용하여 분석...

# 방어 방법

**DARKTRACE**

© 2023 Darktrace (DARK.L). All rights reserved. | [darktrace.com](https://darktrace.com)



# Cyber AI Loop™

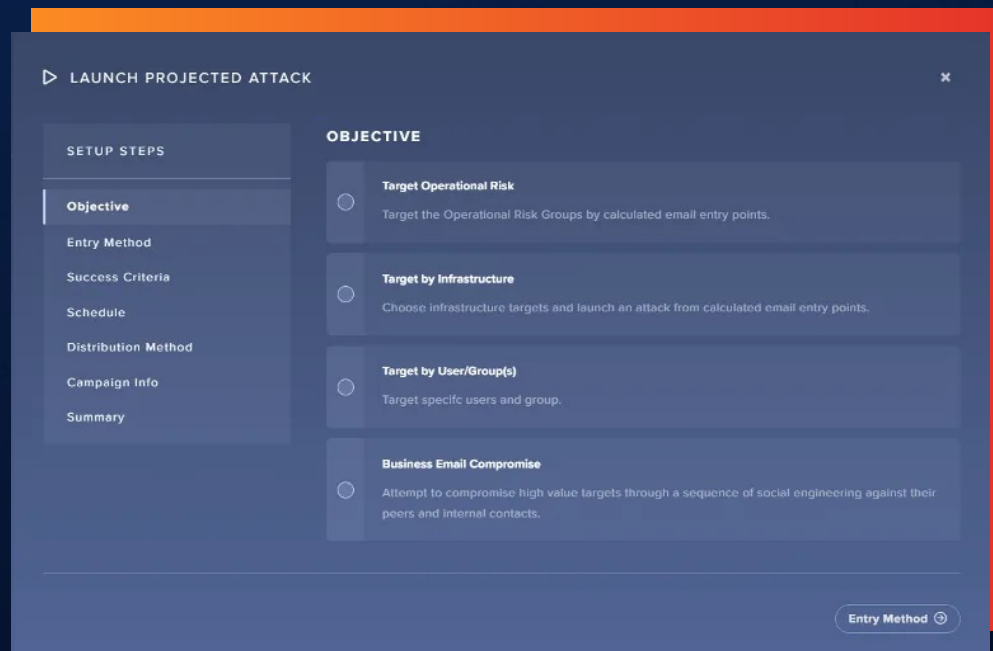
## PRE-ATTACK

## DURING AN ATTACK



# 보안 강화를 위한 생성형 AI

- 맬웨어, 피싱, 스푸핑 및 기타 새로운 위협을 모방하는 무해한 "공격"을 배포
- Darktrace DETECT 및 Darktrace RESPOND 알고리즘에 피드백



**DARKTRACE**

