Institute For National Security Strategy

www.inss.re.kr
eISSN 3140-4103

November 2026
No. 414

INSS 전략보고

AI 도구 활용 국가배후 해킹조직(APT) 활동의 안보적 영향 분석

김민정
mkim@inss.re.kr

Ⅰ. 문제제기

Ⅱ. 최근 관련 사례 분석: 바이낸스 탈취 사건

Ⅲ. 북한 배후 APT 조직의 AI 활용 정황

Ⅳ. 시사점 및 정책 고려 사항

국가안보전략연구원

November 2026
No. 414

CONTENTS

AI 도구 활용 국가배후 해킹조직(APT) 활동의 안보적 영향 분석

I. 문제제기

II. 최근 관련 사례 분석: 바이낸스 탈취 사건

1. 사건 개요와 트레이더트레이터
2. 트레이더트레이터의 AI 활용 수준

III. 북한 배후 APT 조직의 AI 활용 정황

1. AI 에이전트 활용 정황 포착
2. 북한 사이버전 현황 및 AI 투입으로 인한 변화

IV. 시사점 및 정책 고려 사항

1. 북한 배후 APT 조직의 AI 활용 판별 및 에이전틱 AI 대비
2. 예방 중심 역량 강화 및 글로벌 사이버안보 기여

AI 도구 활용 국가배후 해킹조직(APT) 활동의 안보적 영향 분석

저자 | 김민정

국문 초록

지난 2026년 10월 29일 가상자산 거래소 바이낸스에서 단일 사건 사상 최대 규모인 약 63억 달러의 가상자산이 탈취되었으며, 미국 연방수사국(FBI)은 북한 배후 해킹조직인 트레이더트레이터(TraderTraitor)를 공격 주체로 지목하였다. 이번 사건은 북한의 사이버 공격 역량 고도화에 대한 우려를 증폭시키는 동시에, 북한 지능형 지속 위협(APT) 조직의 AI 활용 수준과 잠재적 안보 위협을 검토할 필요성을 제기한다. 특히 마이크로소프트는 트레이더트레이터와 동일 조직으로 분류되어 온 제이드 슬릿(Jade Sleet)을 AI가 작전 업무 전반에 일상적 · 반복적으로 활용되는 'AI 정착 단계'로 평가한 바 있으며, 이는 북한 APT 조직의 AI 활용이 미칠 위협에 대한 질적 변화 가능성 또한 시사하고 있다. 이에 본 연구는 북한 APT 조직의 AI 활용 정황을 살펴보고 AI 투입에 따른 작전 속도와 인력 역량의 변화를 분석하였다. 지금까지 북한 조직이 유인 공작 준비와 악성코드 작성 보조에 AI를 투입한 정황은 다수 포착되었으나 AI가 침투에서 금전화에 이르는 전 단계를 자율 수행한 사례는 아직 확인되지 않았다. 그러나 중국 국가 배후 해킹조직의 AI 기반 표적화, 공격 사례와 같이 북한이 동일 방식을 도입할 경우 북한발 사이버 공격 역량은 크게 증대될 개연성이 높다. 따라서 본 연구는 다음과 같은 정책적 대응 방안을 제시하였다. 첫째, 북한 APT의 AI 관여를 조기에 판별하고 에이전틱 AI 운용 단계 진입을 고려한 대비가 필요하다. 둘째, 예방 중심의 대응 역량 강화를 위해 민관 합동 분석 협의체를 내실화하고, 공격 징후와 가상자산 탈취 · 자금세탁 관련 분석 정보의 신속한 공유체계를 강화해야 한다. 셋째, 북한 IT 인력의 활동 지역 확대에 대응하여 유럽 · 동남아 등과 위협 정보 및 대응 경험을 공유하고, 글로벌 채용 플랫폼과의 협력을 통해 국제 사이버안보에 대한 기여를 확대해야 한다.

주제어: 인공지능(AI), AI 에이전트, 지능형 지속 위협(APT), 사이버 안보, 트레이더트레이터, 복합 안보 위협

Ⅰ 문제제기

▪ 지난 10월 29일 가상자산 거래소 바이낸스에서 단일 사건 사상 최대 규모인 약 63억 달러가 트레이더 트레이터(TraderTraitor)에 의해 탈취[1]되며 북한 배후 APT* 조직의 AI 활용 여부 검토 필요성 대두

※ APT: 지능형 지속 위협(Advanced Persistent Threat)의 약자로 美 국립표준기술연구소는 "고도의 전문성과 상당한 자원을 보유하여 복수의 공격 경로로 목표 달성 기회를 만들어내는 적대 행위자"로 정의

- 북한은 김여정 노동당 중앙위원회 부장 담화(11.3)를 통해 탈취 배후 지목을 부인했으나[2] FBI와 TRM랩스 · 엘립틱 · 체이널리시스 등 제반 기술업체가 북한 트레이더트레이터를 주체로 지목
- 사건 발생 후 일주일이 경과한 시점으로 AI 관여 여부는 분석이 진행 중이나 침투 수법 고도화와 자금세탁 자동화 등을 근거로 영국 BBC와 국내 보안기업 지니언스 등이 AI 기술 활용 가능성 제기

▪ 지난 3월 마이크로소프트는 코랄 슬릿 사례를 통해 북한 연계 APT의 AI 에이전트 활용 사례 최초 확인하며[3] 안보적 함의 증대

- AI 에이전트는 소프트웨어 취약점 탐지와 악성코드 작성에서 높은 수행 능력을 보여 APT 조직 활동에 투입 시 침투 준비 시간 단축과 공격 정밀도 상승에 따른 안보 위협 가중 우려

▪ 이에 본 고는 바이낸스 사건 등 최근 APT 조직 활동 사례를 대상으로 AI 도구 활용 양상을 정리하고 AI 도입의 영향을 분석

1 Federal Bureau of Investigation, "North Korea Responsible for $6 Billion+ Binance Attack," October 31, 2026, https://www.fbi.gov/investigate/cyber/alerts/2026/north-korea-responsible-for-6-billion-binance-attack (accessed: November 2, 2026). 본문의 "약 63억 달러"는 바이낸스가 10.29~11.3에 제시한 자체 추산치로, 탈취 직후부터 블록체인 분석업체들은 약 63.6억 달러(TRM 랩스, 10.30. 10개 블록체인), 약 64억 달러(체이널리시스, 10.30), 약 64.3억 달러(엘립틱, 11.1. NEAR 포함 11개 블록체인)로 상향 추산하였으며 FBI는 10.31. 탈취 규모를 60억 달러 초과로 공지.

2 "김여정 조선로동당 중앙위원회 부장 담화 발표," 『조선중앙통신』 2026년 11월 3일.

3 Microsoft Threat Intelligence, "AI as Tradecraft: How Threat Actors Operationalize AI," March 6, 2026, https://www.microsoft.com/en-us/security/blog/2026/03/06/ai-as-tradecraft-how-threat-actors-operationalize-ai/ (accessed: November 2, 2026).

Ⅱ 최근 관련 사례 분석: 바이낸스 탈취 사건

1. 사건 개요와 트레이더트레이터

가. 탈취 규모와 행위자 귀속

▪ 지난 10월 29일 바이낸스에서 약 63억 달러 규모[4] 가상자산이 탈취되었고 美 FBI는 그 주체로 북한을 지목하면서 해당 활동을 트레이더트레이터(TraderTraitor)로 지칭[5]

- 탈취 개시 약 50시간 만에 행위자를 특정하여 바이비트 사건 당시 5일의 절반 이하로 단축[6]

  ※ 체이널리시스는 FBI 지목 약 이틀 전(한국시간 10.30. 13:30) 비트겟 탈취 자금과의 혼입을 근거로 트레이더트레이터 소행으로 귀속[7]

- 北 명목 GDP 47조 2천억 원의 19%, 작년 北 상품 수출액 4억 7천만 달러의 13배,[8] 바이비트 피해액 14억 6천만 달러의 4배 상회[9]

- 피해 자산 82%가 이더리움 · 비트코인 네트워크에서 발생하였으며 그 중 스테이블코인 위주 28억 5천만 달러(바이낸스 추산 약 30억 달러)는 테더 · 서클 등 발행사 조치로 동결

4 각주 1 참조.

5 Federal Bureau of Investigation, “North Korea Responsible for $6 Billion+ Binance Attack”.

6 소요 시간은 TRM Labs가 집계한 탈취 개시 시각(한국시간 10.30. 05:05)부터 FBI 페이스북 공지 시각(한국시간 11.1. 07:50)까지로 산정하였으나, FBI의 페이스북 게시 이전 로이터는 한국시간 11월 1일 06시 43분에 FBI 발표를 보도하였고 TheStreet는 FBI 사이버 부서의 공개 게시를 한국시간 07시 36분으로 기술. FBI Facebook, “The FBI is releasing information advising the Democratic People’s Republic of Korea (North Korea) was responsible for the theft of more than $6 billion USD in virtual assets from cryptocurrency exchange, Binance,” October 31, 2026, https://www.facebook.com/FBI/posts/the-fbi-is-releasing-information-advising-the-democratic-peoples-republic-of-kor/1550586253781602/ (accessed: November 2, 2026). TRM Labs, “North Korea Steals USD 6.36 Billion From Binance in the Largest Hack in History,” October 30, 2026, https://www.trmlabs.com/resources/blog/north-korea-steals-usd-6-36-billion-from-binance-in-the-largest-hack-in-history (accessed: November 2, 2026). Reuters, “FBI links North Korea to theft of $6 billion from Binance,” October 31, 2026, https://www.reuters.com/world/fbi-links-north-korea-theft-6-billion-from-binance-2026-10-31/ (accessed: November 2, 2026). Mehab Qureshi, “FBI names who took $6B+ from Binance,” TheStreet, October 31, 2026, https://www.thestreet.com/crypto/markets/fbi-names-who-took-6b-from-binance (accessed: November 2, 2026).

7 Chainalysis, “Binance Hack: Tracing the $6.4 Billion Theft to North Korea,” 2026, https://www.chainalysis.com/blog/binance-exploit-latest-findings-and-response/ (accessed: November 2, 2026).

8 한국은행, “2025년 북한 경제성장률 추정 결과,” 2026년 7월 31일, https://www.bok.or.kr/portal/bbs/B0000501/view.do?nttId=11063399&menuNo=201264&programType=newsData&relate=Y&depth=201264 (검색일: 2026년 11월 2일); KOTRA, “2025 북한 대외무역 동향,” 2026년 7월, https://openknowledge.kotra.or.kr/handle/2014.oak/34038 (검색일: 2026년 11월 2일).

▪ 트레이더트레이터는 2022년 美 정부 합동 경보에서 고액 채용을 빙자하여 가상자산 기업 직원에게 유포된 악성 거래 앱을 지칭하며 사용한 명칭으로 이후 해당 조직의 활동 전반에 대한 통칭으로 정착[10]

- FBI · 美 국방부 사이버범죄센터(DC3) · 일본 경찰청 등은 비트코인 사건[11] 공동 발표 시 트레이더트레이터를 제이드 슬릿(Jade Sleet), UNC4899, 슬로 파이시스(Slow Pisces)와 동일한 조직으로 병기[12]
- 마이크로소프트의 위협 행위자 명명 일람은 제이드 슬릿을 북한 조직으로 분류하면서 Storm-0954와 래버린스 천리마를 병기하는 등[13] 해당 조직에 대한 명칭은 하단과 같이 기관별 상이

· 미국 정부: 트레이더트레이터(FBI, CISA, 재무부)
· 마이크로소프트: 제이드 슬릿 (*이전 명칭: Storm-0954)
· 구글 맨디언트: UNC4899
· 팔로알토 네트웍스: 슬로 파이시스
· 크라우드스트라이크: 프레셔 천리마 (*래버린스 천리마에서 분리 지정)
· 기타: PUKCHONG (*유닛42 별칭 목록 수록)

▪ FBI는 2025년 2월 바이비트 거래소 약 15억 달러 탈취 사건에서도 동일 명칭으로 북한의 APT 조직 지목[14]

- 해당 조직의 소행으로 지목된 사건은 DMM 비트코인과 바이비트 외에 2026년의 켈프다오(KelpDAO), 비트겟(Bitget) 피해 포함[15]

9 Ravie Lakshmanan, "FBI Links North Korea's TraderTraitor to Record $6.43 Billion Binance Crypto Heist," The Hacker News, November 1, 2026, https://thehackernews.com/2026/11/fbi-links-north-koreas-tradertraitor-to.html (accessed: November 3, 2026).

10 FBI · CISA · Treasury, "TraderTraitor: North Korean State-Sponsored APT Targets Blockchain Companies," AA22-108A, April 18, 2022, https://www.cisa.gov/news-events/cybersecurity-advisories/aa22-108a (accessed: November 3, 2026).

11 북한 행위자가 악성코드를 유포한 뒤 3억 800만 달러 규모 비트코인을 탈취한 사건.

12 FBI · DC3 · NPA, "FBI, DC3, and NPA Identification of North Korean Cyber Actors, Tracked as TraderTraitor, Responsible for Theft of $308 Million USD from Bitcoin.DMM.com," December 23, 2024, https://www.fbi.gov/news/press-releases/fbi-dc3-and-npa-identification-of-north-korean-cyber-actors-tracked-as-tradertraitor-responsible-for-theft-of-308-million-from-bitcoindmmcom (accessed: November 3, 2026).

13 해당 일람에서 래버린스 천리마는 제이드 슬릿 외에 시트린 슬릿과 다이아몬드 슬릿, 문스톤 슬릿 등 북한 조직 세 곳에도 중복 병기되어 여러 조직을 아우르는 상위 명칭으로 쓰이는 것으로 판단. Microsoft, "How Microsoft names threat actors," Microsoft Learn, https://learn.microsoft.com/en-us/defender-xdr/microsoft-threat-actor-naming (accessed: November 3, 2026).

14 FBI, "North Korea Responsible for $1.5 Billion Bybit Hack," I-022625-PSA, February 26, 2025, https://www.fbi.gov/investigate/cyber/alerts/2025/north-korea-responsible-for-1-5-billion-bybit-hack (accessed: November 3, 2026).

15 FBI · DC3 · NPA, "FBI, DC3, and NPA Identification of North Korean Cyber Actors, Tracked as TraderTraitor".

나. 탈취 자금의 최종 귀속 추정

▪ FBI가 탈취 배후로 지목한 트레이더트레이터는 美 정부 문건과 UN 문건 상 정찰정보총국(舊 정찰총국) 산하 조직으로 분류

※ 작년 9월 박정천 당 중앙군사위원회 부위원장 담화로 식별된 정찰정보총국은 구 정찰총국을 확장 개편한 기관으로 추정되나 UN 제재명단과 美 재무부 문건은 여전히 舊 명칭인 정찰총국으로 등재 · 표기[16]

- 美 재무부가 라자루스 그룹을 정찰총국 산하 조직으로 적시하고[17] 美 정부 기관 공동 권고문이 라자루스 · APT38 등을 트레이더트레이터의 배후로 규정[18]
- 2016년 안보리 결의 2270호는 정찰총국을 대량살상무기 개발 지원 기관으로 제재 대상에 포함했으며 美 재무부는 후원 주체를 정찰총국 3국(해외정보국)까지 좁혀 지목[19]

※ 2023년 5월 추가 제재 시 정찰총국 통제 下 기술정찰국과 하위 사이버 부대인 110호 연구소도 제재 대상 포함[20]

▪ 한 · 미 등 11개국이 참여한 다국적제재모니터링팀(MSMT)의 2025년 보고서도 트레이더트레이터를 舊 정찰총국 산하 조직으로 언급[21]

- 트레이더트레이터는 정찰총국 3국 산하 110연구소 소속이며 MSMT는 동 연구소 산하 조직(템프 허밋, 시트린 슬릿, 트레이더트레이터, 크립토코어)을 라자루스로 통칭[22]

16 UN 안전보장이사회, "DPRK Sanctions List," KPe.031, https://scsanctions.un.org/en/?keywords=dprk (accessed: November 3, 2026); U.S. Department of the Treasury, "National Proliferation Financing Risk Assessment," March 2026, https://home.treasury.gov/system/files/246/2026-NPFRA.pdf (accessed: November 3, 2026).

17 U.S. Department of the Treasury, "Treasury Sanctions North Korean State-Sponsored Malicious Cyber Groups," September 13, 2019, https://home.treasury.gov/news/press-releases/sm774 (accessed: November 3, 2026).

18 FBI · CISA · Treasury, "TraderTraitor".

19 U.S. Department of the Treasury, "Treasury Sanctions North Korean State-Sponsored Malicious Cyber Groups".

20 U.S. Department of the Treasury, "Treasury Targets DPRK Malicious Cyber and Illicit IT Worker Activities," May 23, 2023, https://home.treasury.gov/news/press-releases/jy1498 (accessed: November 3, 2026).

21 Multilateral Sanctions Monitoring Team, "The DPRK's Violation and Evasion of UN Sanctions through Cyber and Information Technology Worker Activities," MSMT/2025/2, October 22, 2025, https://msmt.info/Publications/detail/MSMT%20Report/4221 (accessed: November 3, 2026).

22 Multilateral Sanctions Monitoring Team, "The DPRK's Violation and Evasion of UN Sanctions," pp. 14-15.

▪ FBI가 해당 활동을 트레이더트레이터로 지칭할 때 탈취의 책임 주체를 '북한'으로 적시했으며[23] FBI · 재무부는 트레이더트레이터의 자금 조성 · 세탁 목적을 '정권 지원'으로 평가[24]

- 위장 취업과 자산 탈취는 서로 상이한 기관 · 인력이 수행*하나 舊 정찰총국 예하 인물의 IT 인력 사업 관여 사례[25] 및 IT 인력의 가상자산 탈취 가담 사례[26] 등 포착으로 두 활동 간 협업 가능성 무배제

  ※ 예컨대 2022년 지침은 IT 인력 파견 주체로 정찰총국과 별개 기관인 군수공업부 산하 313총국 지목[27]

- 美 당국은 두 활동의 수익이 정권에 귀속되며[28] IT 인력 사업과 가상자산 세탁으로 획득한 수익이 WMD 재원으로 활용된다고 판단[29]

2. 트레이더트레이터의 AI 활용 수준

AI 기술 활용 수준은 현재 분석이 진행 중이며 △ **탈취 주체의 과거 AI 도구 활용** 이력 △ **자금세탁 관측 결과** △ 사건 **분석소요 기간** 등을 고려할 때 AI 기술 투입 **가능성도 고려할 필요**

가. 트레이더트레이터는 AI 정착 단계

▪ FBI 등이 트레이더트레이터와 동일 조직으로 병기한 제이드 슬릿은 AI가 작전 업무 전반에 정례 편입된 정착 단계

- 사건 당일인 10월 29일 마이크로소프트는 북한 APT 조직의 AI 기술 도입 정도를 네 단계로 구분하며 제이드 · 코랄 · 재스퍼 슬릿이 "AI를 일상적인 과업 수행에 활용한다"라고 평가[30]

23 Federal Bureau of Investigation, "North Korea Responsible for $6 Billion+ Binance Attack". 단 트럼프 대통령은 "북한인지 모르지만, 북한이라면 대화하겠다"며 판단 유보(10.31).

24 FBI · CISA · Treasury, "TraderTraitor"; U.S. Department of the Treasury, "Treasury Sanctions Facilitators of DPRK IT Worker Fraud Targeting U.S. Businesses," October 28, 2026, https://home.treasury.gov/news/press-releases/sb0669 (accessed: November 3, 2026).

25 작년 재무부는 정찰총국 예하 안다리엘과 관련된 송금혁을 IT 인력 위장 취업 조력 혐의로 제재. U.S. Department of the Treasury, "Sanctions Imposed on DPRK IT Workers Generating Revenue for the Kim Regime," July 8, 2025, https://home.treasury.gov/news/press-releases/sb0190 (accessed: November 3, 2026).

26 Multilateral Sanctions Monitoring Team, "The DPRK's Violation and Evasion of UN Sanctions".

27 U.S. Department of State, U.S. Department of the Treasury, and Federal Bureau of Investigation, "Guidance on the Democratic People's Republic of Korea Information Technology Workers," May 16, 2022, https://ofac.treasury.gov/media/923126/download?inline (accessed: November 3, 2026).

28 FBI · CISA · Treasury, "TraderTraitor".

29 U.S. Department of the Treasury, "Treasury Sanctions Facilitators of DPRK IT Worker Fraud Targeting U.S. Businesses."

30 Microsoft Threat Intelligence, "Tracking AI adoption in North Korean threat actor tradecraft," October 29,

〈표 1〉 북한 조직의 AI 도입 수준[31]

단계	해당 되는 APT 조직	설명
정착 단계	제이드 · 코랄 · 재스퍼 슬릿	작전 업무 전반에 AI가 정례 절차로 편입된 단계
가속 단계	다이아몬드 · 오닉스 · 펄 · 사파이어 슬릿	복수 작전 기능에 AI를 급속 도입 중인 단계
사용 증가 단계	에메랄드 슬릿	일부 작전 활동에 AI 사용이 증가하였으나 업무 전반에는 미정착한 단계
시험 사용 단계	해당 조직 없음	제한적 · 시험적 수준에 머문 단계

- 미 · 일 수사기관이 트레이더트레이터와 제이드 슬릿을 동일 조직으로 병기하므로 제이드 슬릿 역량을 트레이더트레이터의 역량으로 판단 가능
- 다만 보안업체마다 위협 행위자 분류 단위가 상이하고 마이크로소프트 명명 일람에 트레이더트레이터가 미수록되어 두 명칭의 완전 일치 여부 미확인[32]

■ 마이크로소프트는 제이드 슬릿의 AI 활용 근거로 지난 6월 8~12일 포착된 위장 기업을 동원한 채용 위장 접근 사례 제시[33]

- AI 기반 블록체인 분석 기업을 사칭하고 소프트웨어 개발직 채용을 빙자하여 접근한 뒤 면접 일정 확정까지 진행
- 단기간 복수의 위장 기업을 동원한 활동 속도 · 규모로 볼 때 △표적 조사와 접근 확대에 생성형 AI가 투입되었으며 △커서(Cursor) 등 AI 도구가 악성코드 작성과 오류 수정을 보조하였을 가능성 다대

■ 최근에는 해당 조직이 채용 면접 과제를 가장해 개발자용 업무 프로그램의 변조 플러그인 설치를 유도*한 뒤 피해자 전산망의 정보를 탈취하고 원격으로 장악하는 수법도 공개[34]

2026, https://www.microsoft.com/en-us/security/blog/2026/10/29/tracking-ai-adoption-in-north-korean-threat-actor-tradecraft/ (accessed: November 3, 2026).

31 Microsoft Threat Intelligence, "Tracking AI adoption".

32 Microsoft, "How Microsoft names threat actors".

33 Microsoft Threat Intelligence, "Tracking AI adoption".

34 Ravie Lakshmanan, "FBI Links North Korea's TraderTraitor to Record $6.43 Billion Binance Crypto Heist"; "Albert Priego · Alex Delamotte · Matěj Havránek, "Don't Call Us, We'll Call Your APIs | TraderTraitor Backdoors Resurface on Victim With No Crypto Ties," SentinelLABS, September 18, 2026, https://www.sentinelone.com/labs/dont-call-us-well-call-your-apis-tradertraitor-backdoors-resurface-on-victim-with-no-crypto-ties/ (accessed: November 3, 2026); Zscaler ThreatLabz(Seongsu Park), "Suspected TraderTraitor Group Uses Trojanized Terraform Provider to Deliver Cross-Platform Malware," September 30, 2026, https://www.zscaler.com/blogs/security-research/suspected-tradertraitor-group-uses-trojanized-

※ 면접 과제용 코드 저장소를 통해 테라폼(Terraform)의 공급자 플러그인을 악성으로 변조하여 유포하고 피해자 운영체제에 맞는 악성코드 다운로드 유도

나. 바이낸스 탈취 과정의 AI 활용 여부

▪ 바이낸스 초기 보고서(Initial Findings and Response Update, 11.3)에 따르면 6월 공급망 공격 당시 AI 코딩 도구 설치 중 변조된 설치 프로그램이 유입되어 초기 침해 발생[35]

- 보고서는 "사내 AI 도구 침해(compromised internal AI tooling)" 등 공격자가 결합한 다양한 기법들을 열거하며 이번 사건을 "단발 침해나 피싱 사건"으로 볼 수 없는 다단계 공작으로 규정
- 첫 침해 이후 공격자는 사내 AI 도구를 내부 확산 수단으로 악용하여 배포 권한 토큰 확보

> 6월 엔지니어 A, 변조 홈브루(Homebrew) 설치본 다운로드
> →9월 엔지니어 B, 가짜 채용 면접으로 감염
> →10월 20일 사내 AI 도구로 A · C 컴퓨터 장악 후 배포 권한 토큰 확보
> →10월 29일 확보 토큰으로 탈취 실행

▪ 엘립틱은 AI 에이전트의 의사결정 정황이 현 단계에서 식별되지 않았다고 평가하였으나[36] △ 자금 세탁 과정이 명백히 자동화되었고 △ 최초 경유한 6개의 크로스 체인 브리지를 선택한 주체가 인간인지 AI인지 판별할 수 없다고 밝힌바 AI 개입 가능성은 열어 둔 것으로 사료[37]

※ 최초 경유 브리지: 탈취 자금이 10.30에 처음 거쳐 간 6개 블록체인 간 자산 전환 서비스 (크로스 체인 브리지)

- 엘립틱의 평가는 AI 에이전트(스스로 판단하여 거래 경로를 바꾸는 AI)의 의사결정에 한정되며 AI 활용 여부 전반에 관한 판단이 아님

▪ 엘립틱은 해당 판단**(AI 에이전트 無 개입)**의 근거로 반복 송금과 실패 거래의 재전송 기록을 제시했으나 AI를 활용한 공격일수록 반복 시도 · 재시도 기록이 다수 잔존할 개연성 존재

---

terraform-provider-deliver (accessed: November 3, 2026).

35 Binance, "29 October Binance Security Incident: Initial Findings and Response Update," X November 3, 2026, https://x.com/binance/status/2117626126214504729 (accessed: November 4, 2026).

36 Elliptic, "How North Korea is laundering the $6.4 billion stolen from Binance," November 2, 2026, https://www.elliptic.co/insights/how-north-korea-is-laundering-the-6-4-billion-stolen-from-binance/ (accessed: November 2, 2026); 코인데스크, "엘립틱, 64억 달러 규모 바이낸스 횡령 자금 세탁에 AI 사용 흔적 없어," 2026년 11월 2일, https://www.coindesk.com/ko/tech/2026/11/02/no-proof-of-ai-in-north-korean-hackers-move-of-stolen-binance-funds-through-thousands-of-wallets (검색일: 2026년 11월 3일); Elliptic, "The Binance hack is laundering on autopilot," October 31, 2026, https://www.elliptic.co/insights/the-binance-hack-is-laundering-on-autopilot/ (accessed: November 3, 2026).

37 코인데스크, "엘립틱, 64억 달러 규모 바이낸스 횡령 자금 세탁에 AI 사용 흔적 없어".

- 엘립틱은 △브리지가 거부한 입금을 우회하지 않고 반복 재전송한 기록과 △송금 간격과 교환 금액이 일정하게 반복되어 인간이 작성 · 감독하는 실행 명령(스크립트)과 유사하다고 판단[38]
- 그러나 IEEE 심포지엄(2026) 실험에서 AI가 명령까지 수행하는 기존 공격 도구 중 최고 성능 도구는 명령의 47~90%를 공격과 무관한 작업에 소모했고 필요 작업의 6~41%를 잘못된 명령으로 수행[39]
- 또한 AI를 쓰지 않은 공격 도구(칼데라)도 무관한 작업을 반복 수행한 결과로 볼 때 반복 기록은 AI 활용 여부와 관계없이 발생[40]

▪ 이와 같이 △AI의 공격 계획 수립 △통상의 해킹 기법에 의한 실행이 결합할 시 피해 전산망에는 일반 해킹과 동일한 흔적만 잔존하므로, 공격 흔적으로 AI 개입 여부 판별이 불가한 것은 AI 활용 공격의 본질적 속성

- 예컨대 상기 심포지엄에서 검증된 공격 AI*인 인칼모(Incalmo)[41]는 40개의 모의 전산망 중 37개에서 목표 자산을 확보했는데 이 경우 피해 전산망에 남는 흔적은 기성 도구의 흔적과 다르지 않음

▪ 따라서 반복 재전송 기록만으로 AI 활용 여부를 판단하는 것은 불가하므로 공격 단계별 소요 시간 · 비용의 변화와 공작 서버의 AI 도구 보유 정황을 대조 검증하여 평가 필요

38 Elliptic, “The Binance hack is laundering on autopilot.”

39 Brian Singer · Keane Lucas · Lakshmi Adiga · Meghna Jain · Lujo Bauer · Vyas Sekar, “Incalmo: An Autonomous LLM-assisted System for Red Teaming Multi-Host Networks,” paper presented at the 2026 IEEE Symposium on Security and Privacy (San Francisco, CA, May 18-21, 2026), https://ieeexplore.ieee.org/document/11573490/ (accessed: November 4, 2026).

40 Singer et al., “Incalmo.”

41 인칼모의 경우 AI는 공격 계획만 수립하고 실행은 기성 해킹 도구에 위임.

Ⅲ 북한 배후 APT 조직의 AI 활용 정황

1. AI 에이전트 활용 정황 포착

가. 코랄 슬릿 등 북한 배후 APT 조직의 AI 에이전트 투입

▪ 지난 3월 마이크로소프트가 공개한 코랄 슬릿 사례[42]는 북한 배후 조직의 AI 에이전트 투입을 확인한 첫 공개 사례

- 국가 배후 조직 전체로는 2025년 11월 앤트로픽이 공개한 중국 배후 GTG-1002 사례[43]가 최초

▪ 마이크로소프트는 코랄 슬릿(Coral Sleet, 舊 Storm-1877)이 허위 웹사이트 제작에서 원격 인프라 준비를 거쳐 공격 코드 시험 · 배포에 이르는 유인 공작 과정을 AI 에이전트 도구로 수행하였다고 적시[44]

- AI 코딩 도구로 악성코드 구성요소를 생성 · 재구현하고 대형언어모델의 안전장치를 탈옥으로 우회하여 신종 악성코드를 제작
- 오터쿠키(OtterCookie) 악성코드에서 코드 사이의 이모지 표식과 대화체 주석 등 AI 생성 흔적이 식별됨
- 2026년 4월부터 비공개 깃허브 저장소를 악성코드 유포 수단으로 변조하는 기능을 가동하였으며 마이크로소프트는 이 기능을 AI 개발 도구로 제작하였을 개연성이 높다고 평가[45]

▪ 보안 업계는 코랄 슬릿 사례를 인간 주도의 개별 침투가 AI 기반 자동 침투로 전환되는 초기 징후로 평가

- 해당 사례를 재인용한 후속 보고서들은 고도화된 행위자들이 최소한의 인간 개입으로 공격 단계를 잇는 에이전틱 AI 작업 과정을 도입하기 시작하였으며[46] 위협 행위자 전반으로의 확산 우려[47]

---

42 Microsoft Threat Intelligence, “AI as Tradecraft”.

43 Anthropic, “Disrupting the First Reported AI-Orchestrated Cyber Espionage Campaign,” November 13, 2025, pp. 3-4, https://assets.anthropic.com/m/ec212e6566a0d47/original/Disrupting-the-first-reported-AI-orchestrated-cyber-espionage-campaign.pdf (accessed: November 4, 2026).

44 Microsoft Threat Intelligence, “AI as Tradecraft”.

45 Microsoft Threat Intelligence, “Tracking AI adoption”.

46 ThreatDown, “2026 State of Malware,” 2026, https://www.threatdown.com/wp-content/uploads/2026/01/2026-TD-State-of-Malware-1.pdf (accessed: April 17, 2026); Cloud Security Alliance AI Safety Initiative, “Microsoft’s AI Threat Intelligence: Documenting the Full AI-Accelerated Attack Lifecycle: Unofficial AI-assisted Research,” March 8, 2026, p. 3, “the 2026 picture shows...with the most advanced actors beginning to deploy agentic AI workflows that chain these stages together with minimal human intervention,” https://

▪ 북한 해킹 조직인 오닉스 슬릿(안다리엘)과 펄 슬릿도 지난 10월 하순부터 오픈소스 모의 해킹용 AI 에이전트 도구(ARTEX)를 공격에 활용 중[48]

나. AI 에이전트 활용 수준

▪ OECD의 구분 방법에[49] 비추어 볼 때 북한 APT 조직은 AI 에이전트를 공격 준비 작업에 투입하는 단계에 있으며 에이전틱 AI 운용 단계에 진입하였다고 볼 근거는 아직 확인되지 않음

> OECD는
> · △AI 에이전트를 일정한 자율성을 갖고 도구를 활용하여 특정 목표를 달성하는 시스템으로
> · △에이전틱 AI를 복수의 AI 에이전트가 작업을 분담하며 최소한의 인간 감독 下 장기간 복합 목표를 추구하는 시스템으로 구분

▪ 코랄 슬릿과 오닉스 슬릿 · 펄 슬릿 사례는 AI 에이전트 도구 활용에 해당하며, 마이크로소프트가 북한의 AI 활용 작전을 인간 지휘 하의 비자율 작전으로 평가한 만큼 에이전틱 AI 요건에는 미달

- 마이크로소프트는 북한 위협 행위자가 에이전틱 AI 활용의 초기 시험 단계에 진입했으며 대규모 활용은 관측되지 않았다고 기술
- 코랄 슬릿의 경우 AI 에이전트 활용이 유인 공작 준비 단계에 국한되며 자료 반출과 금전화 단계까지 AI가 자율 수행한 근거는 미확인
- 유인물 반복 제작과 코드 보정 등 공격 준비 작업의 AI 활용은 다수의 공개 사례에서 확인되나 인간 운용자의 관여 정도와 침해 성공 여부에 관한 분석 자료는 제한적(〈표 2〉)[50]

labs.cloudsecurityalliance.org/wp-content/uploads/2026/03/CSA_research_note_microsoft_ai_attack_lifecycle_intelligence_20260308-csa-styled.pdf (accessed: April 17, 2026).

47 Cloud Security Alliance AI Safety Initiative, "Microsoft's AI Threat Intelligence: Documenting the Full AI-Accelerated Attack Lifecycle", p. 6.

48 Microsoft Threat Intelligence, "Tracking AI adoption".

49 OECD, "The agentic AI landscape and its conceptual foundations," OECD Artificial Intelligence Papers No. 56, February 2026, pp. 6, 12, 15, https://www.oecd.org/content/dam/oecd/en/publications/reports/2026/02/the-agentic-ai-landscape-and-its-conceptual-foundations_a9d4b451/396cf758-en.pdf (accessed: November 4, 2026). OECD는 이 정리가 문헌 개관이며 AI 에이전트에 요구되는 자율성 수준에 관한 합의는 없다고 밝힘(p. 15).

50 Microsoft Threat Intelligence, "Tracking AI adoption".

〈표 2〉 북한 조직의 AI 활용 분야별 관측 내용과 분석상 제약[51]

활용 분야	관측 내용	분석상 제약
표적 조사	표적 기업과 직원의 직무 정보 수집	조사 기록만으로 침투 성공 입증 불가
기만 자료 제작	채용 문안과 허위 프로필 및 유인 문서 생성	화상 통화 내 딥페이크 사용 주장의 경우 제3자 검증 전 확정 곤란
코드 작성 보조	코드 오류 수정 · 시험 및 언어 변환	코드 품질 향상만으로 공격 전 과정의 자율 수행 입증 불가
반복 업무	문서 생성과 수집 자료 요약	문서 생성량 증가를 피해액 증가의 원인으로 단정 곤란
자금 처리	탈취 자금 세탁 과정의 자동 거래 관측 (바이낸스 사건)	자동화 정황만으로 AI의 의사결정 입증 불가 (Ⅱ장 2.나 참조)

▪ 그러나 인간 개입 없는 AI 에이전트의 사이버 공격 능력이 국제 학술지에서 거듭 입증*되는 등 기술 여건이 旣 조성된바 AI 정착 단계에 이른 북한 APT 조직도 에이전틱 AI 운용 단계에 조기 진입할 개연성 존재

※ 코드 실행 권한을 부여받은 AI 에이전트가 인간 개입 없이 사이버 공격을 수행하거나[52] 복수의 AI가 역할을 분담하여 침투부터 공격 종료까지 인간 개입 없이 수행하는[53] 등 다수의 연구 · 실험 결과 존재

- 기 언급한 GTG-1002가 에이전틱 AI에 근접한 선례로, 인간 운용자는 공격당 4~6회의 판단 시점에만 개입하고 작전의 80~90%를 AI가 수행[54]

2. 북한 사이버전 현황 및 AI 투입으로 인한 변화

가. 북한 사이버전 인력 현황

▪ 북한 사이버전 인력은 2024년 8월 기준 약 8천4백 명으로 추산되며[55] 이는 참고로 「2022 국방 백서」에 수록된 6천8백여 명보다 약 1천6백 명이 늘어난 규모

51 Ross Inman · Adrian Hernandez, "UNC1069 Targets Cryptocurrency Sector with New Tooling and AI-Enabled Social Engineering," Google Cloud Blog February 9, 2026, https://cloud.google.com/blog/topics/threat-intelligence/unc1069-targets-cryptocurrency-ai-social-engineering (accessed: November 3, 2026); Elliptic, "How North Korea is laundering the $6.4 billion stolen from Binance".

52 Anshuman Chhabra 외, "Agentic AI Security: Threats, Defenses, Evaluation, and Open Challenges," IEEE Access, 2026, https://ieeexplore.ieee.org/iel8/6287639/6514899/11447227.pdf (accessed: November 3, 2026).

53 Pengchao Yao 외, "Multi-Agents Assisted Autonomous Breach and Attack Simulation Framework in Power Information System," IEEE Access, Vol. 14, 2026, https://ieeexplore.ieee.org/document/11664062/ (accessed: November 3, 2026).

54 Anthropic, "Disrupting the First Reported AI-Orchestrated Cyber Espionage Campaign".

55 국가정보원 3차장 기자간담회 발언으로 신규진, "'北 해킹 인력 8400여명'…국정원-합참 공식 평가," 『동아일보』 2024년 8월 9일, https://www.donga.com/news/Politics/article/all/20240809/126440040/1 (검색일: 2026년 11월 4일)에 수록 되었으며, 국군 사이버 작전사령부도 2024.10.30. 국회 국정감사에서 동일 수치를 보고함. 국정원은 해당 수치를 세부 분류 작업이 진행 중인 잠정치로 제시하였음.

- 2022년 6천8백 명 추산 당시 전담 해커 1천7백 명과 지원 인력 5천1백 명으로 내역이 공개되었으나 8천4백 명은 해킹 수행 인원과 지원 인원의 구성비 미공개

▪ IT 인력도 악성 사이버 작전을 수행하며 사이버 인력과의 협업 및 수법 공유가 확대되고 있어[56] 두 집단의 엄격한 분류는 곤란

- 북한 사이버 전문 역량의 대부분은 정찰정보총국에 있으나 사이버 조직과 IT 인력은 다수 기관에 분산 소속

> · 유엔 제재 대상: 정찰총국, 국방성, 원자력공업성, 군수공업부, 노동당(관련 기관 자산동결 적용)
> · 제재 비지정 기관: 국가보위성, 사회안전성, 노동당 중앙위원회 과학교육부

- 정찰정보총국 3국 예하에는 110연구소, 63연구소, 970소 등 최소 7개 APT 조직이 소속되는 반면 정찰정보총국 722 연락소*는 IT 프리랜서 업무와 해킹을 겸함
- IT 인력 회사를 여전히 가장 많이 보유하는 기관은 군수공업부 예하 313총국이며 옌볜 실버스타 등이 대표적 위장회사[57]

▪ 북한 IT 인력의 경우 MSMT 보고서 上 국가별 IT 인력 수치 합산 시 작년 기준 약 1천6백~3천1백 명 추산[58]

- MSMT는 중국 1천~1천5백 명과 북한 국내 450~1천2백 명을 주요 활동 인원으로 집계하였으며 여기에 러시아 150~300명과 라오스 20~40명 등 소수 파견국 인원 합산
- FBI는 올해 10월 "수천 명"의 원격 IT 인력이 정권 통제하에 활동한다고 언급[59]

▪ 다만 위장 취업을 통한 임금 수취 여부나 수행 행위, 혹은 침투 · 탈취 여부로 두 집단을 구분함이 대체로 타당할 것으로 사료

- 일 · 미 · 호 · 독 합동 주의보(2026.9.18.) "△해킹 조직 워터플럼과 일부 IT 인력이 모두 군수공업부 313총국 소속이며 △일부 워터플럼 요원은 IT 인력 업무를 병행하고 △두 집단이 동일한 IP 주소 사용"

56 Multilateral Sanctions Monitoring Team, "The DPRK's Violation and Evasion of UN Sanctions".

57 Multilateral Sanctions Monitoring Team, "The DPRK's Violation and Evasion of UN Sanctions," p. 16(각주 포함).

58 Multilateral Sanctions Monitoring Team, "The DPRK's Violation and Evasion of UN Sanctions". 지난 9월 MSMT 보고서의 해외 노동자 합산 수치(3만 5천600~10만 1천280명)는 IT 인력 수치가 아니므로 구분 필요.

59 FBI, "North Korean IT Workers Expanding Use of Facilitators, Broadening Recruitment," I-101326-PSA, October 13, 2026, https://www.ic3.gov/PSA/2026/PSA261013 (accessed: November 3, 2026).

- 해당 주의보는 "IT 인력이 주로 수익 창출에 집중"한다고 적시했으나 반면 美 법무부 검사가 IT 인력을 '사이버 요원(cyber operatives)'과 혼용한 사례도 존재[60]

나. AI 투입에 따른 사이버 작전 속도 변화

■ AI 활용 공격에서 최초 침투 후 수 시간 만에 기관 전산망을 장악하거나 하루 만에 대량의 인증 정보를 탈취한 사례가 잇따라 확인되며 AI가 숙련 인력의 노동을 대체하여 비용 대비 효과가 제고된 것으로 분석[61]

- 탈취한 개발자 토큰 1개로 피해 기관 클라우드의 관리자 권한 전체를 약 3시간 만에 장악한 사례 보고[62]
- 별도의 IT업체 침해 사건의 경우 200여 개 고객사 자료가 유출되고 40여 개 기업의 마이크로소프트 계정 인증 토큰 2천1백여 세트가 약 34시간 만에 탈취되었으며 AI 에이전트가 대부분의 작업 수행

■ 앞서 언급한 GTG-1002 사례에서 AI는 수천 건의 요청을 대개 초당 여러 건씩 처리하였으며 앤트로픽은 인간 해커가 따라잡을 수 없는 속도로 평가[63]

■ 사이버 공격 전반의 속도는 향상되고 있으나 AI 도구가 기여한 부분을 별도 집계한 자료는 제한

- 작년 범죄형 해킹의 평균 확산 개시(최초 침투→타 장비 이동) 시간은 전년 대비 65% 빨라진 29분이었고 최단 기록은 27초였으나[64] 해당 수치는 AI 효과와 구분 없이 발표됨

※ 다만 바이낸스 사건의 경우 6월 최초 침해부터 10.29. 탈취까지 약 4개월이 소요된 것으로 유추되어 수사 결과를 통한 추가 검증 要

60 U.S. Department of Justice, "Justice Department Announces Coordinated, Nationwide Actions to Combat North Korean Remote Information Technology Workers' Illicit Revenue Generation Schemes," June 30, 2025, https://www.justice.gov/opa/pr/justice-department-announces-coordinated-nationwide-actions-combat-north-korean-remote (accessed: November 3, 2026).

61 Anthropic, "Detecting and Countering Misuse of AI: September 2026," September 10, 2026, https://www.anthropic.com/threat-intelligence-report-september-2026 (accessed: November 4, 2026).

62 Anthropic, "Detecting and Countering Misuse of AI: September 2026".

63 Anthropic, "Disrupting the first reported AI-orchestrated cyber espionage campaign," 2025.11., https://www.anthropic.com/news/disrupting-AI-espionage (행위 조직 내부 명칭 GTG-1002).

64 Adam Meyers, "CrowdStrike 2026 Global Threat Report: The Evasive Adversary Wields AI," CrowdStrike Blog, 2026.2.24., https://www.crowdstrike.com/en-us/blog/crowdstrike-2026-global-threat-report-findings/ (accessed: November 4, 2026).

다. AI에 의한 사이버 인력 역량 증폭

- 과거 숙련 인력 다수가 필요하던 공격을 소수 운영자가 수행한 사례가 잇따라 확인되어 북한 사이버 인력의 공격 역량도 증대될 개연성
  - 앤트로픽은 단일 범죄자가 AI 코딩 도구(클로드 코드)로 정찰과 침투를 자동화하여 최소 17개 기관을 노린 사례 공개[65]
  - 해당 사건에서 유출 자료 선별과 협박 문구 작성까지 AI에 위임되었으며 앤트로픽은 AI 미활용 시 운영 인력 한 팀이 소요되었을 공격으로 평가

- 북한 인력이 AI로 허위 경력을 작성하고 기술 평가를 통과하여 美 포천 500대 기술기업의 원격 근무직을 확보한 뒤 실무까지 수행하는 등[66] AI 활용으로 북한 IT 인력 파견 규모의 제약 요인 약화
  - 앤트로픽은 수년간의 전문교육을 거쳐야 하는 북한 정권의 인력 양성 능력이 파견 규모를 제약해 왔으나 AI가 해당 제약을 제거하였다고 평가

- AI를 활용하여 공격이 정교해지면서 국가 배후 여부를 가늠해 온 귀속 판단 관행도 유효성 저하
  - 국가 배후 공작과 개인 운영자 간 인력 · 도구 격차를 AI가 해소하여 정교함의 배후 판별 단서 기능 상실[67]

65 Anthropic, “Detecting and Countering Misuse of AI: September 2026,” September 10, 2026, https://www.anthropic.com/threat-intelligence-report-september-2026 (accessed: November 4, 2026).

66 Anthropic, “Detecting and Countering Misuse of AI: August 2025”.

67 Anthropic, “Detecting and Countering Misuse of AI: September 2026”.

Ⅳ 시사점 및 정책 고려 사항

1. 북한 배후 APT 조직의 AI 활용 판별 및 에이전틱 AI 대비

가. AI 관여 판별 및 귀속 평가 역량 확보

- 북한 조직의 AI 활용 정황은 AI 업체의 자사 서비스 악용 계정 차단 혹은 위협 보고서 발간 후에야 공개되어 온바, 공개 이전 단계에서 단계별 소요 시간과 공작 서버 정황을 대조하는 판별 역량 확보 要**(Ⅱ장 2.나)**
  - 트레이더트레이터와 동일 조직으로 병기되는 UNC4899가 제미나이로 코드 개발과 취약점 공격 연구를 수행한 사실은 구글의 계정 차단 이후 작년 11월 보고서로 공개[68]
  - 코랄 슬릿이 AI 개발 도구로 제작했을 개연성이 높은 비공개 깃허브 저장소 변조 기능은 4월 가동 후 약 6개월 뒤 공개**(Ⅲ장 1.가)**

- 업체별 AI 활용 단위 · 명칭이 상이하여 명칭 대조 시 오독 우려가 있는바 업체별 평가를 동일 척도로 비교할 공통 평가 지표 수립 제안
  - 마이크로소프트는 조직 단위로 북한 조직의 작전이 인간 지휘하에 있다고 평가하였고 TRM 랩스는 범죄 유형 단위로 대형 탈취의 공격 코드를 AI가 생성한 증거는 없다고 평가[69]
  - 동일 단어(emerging)가 마이크로소프트에서는 최하위 단계, TRM 랩스에서는 중간 단계를 지칭하여 인용 시 혼동 소지

나. 에이전틱 AI 운용 단계 진입 이전 선제 대비

- 북한 조직이 우리 국민 대상 유인물을 AI로 고속 대량 생산했을 개연성이 높은 바 AI 생성 변형 유인물의 탐지 기법 고도화 要
  - 오닉스 슬릿이 8~9월 제작한 표적 맞춤형 유인용 PDF의 경우 수 분 내 수십 개가 생성되어 AI 산출물일 개연성이 높다고 평가[70]

68 Google Threat Intelligence Group, "GTIG AI Threat Tracker: Advances in Threat Actor Usage of AI Tools," November 2025, https://cloud.google.com/blog/topics/threat-intelligence/threat-actor-usage-of-ai-tools (accessed: November 6, 2026).

69 Microsoft Threat Intelligence, "Tracking AI adoption"; TRM Labs, "The 2026 AI-in-Crime Adoption Index," August 2026, https://www.trmlabs.com/reports-and-whitepapers/the-2026-ai-in-crime-adoption-index (accessed: November 6, 2026). TRM 랩스의 해킹 · 국가배후 탈취 항목은 북한 탈취와 일반 코드 해킹을 포괄.

70 Microsoft Threat Intelligence, "Tracking AI adoption".

- 해당 유인물은 우리 국민 대상 대량 피싱에 쓰였으며 마이크로소프트는 신속한 재생성·변형으로 차단·탐지가 제한될 수 있다고 평가

▪ 사내 AI 도구가 내부 확산 수단으로 악용된 만큼**(II장 2.나)** AI 개발 도구의 설치와 권한 부여 과정을 상시 점검 항목에 편입 要

- AI 코딩 도구 설치본의 변조 여부 검증과 사내 AI 도구의 권한 범위 제한을 점검 항목에 포함하여 가상자산 사업자 등 금융 부문 기관에 우선 적용하고 원격 접속 기록 점검 병행

▪ 북한 조직의 AI 작전은 아직 인간 지휘하에 있으나 AI의 자율 공격 능력이 급속히 향상되고 있어 에이전틱 AI 운용 진입 징후의 조기 포착 要

- 영국 AI 안보연구소의 32단계 기업망 모의 공격 실험에서 동일 연산량 투입 시 AI의 평균 완료 단계가 18개월 만에 1.7단계에서 9.8단계로 상승[71]
- 오닉스·펄 슬릿이 공개 배포된 에이전트형 모의 해킹 도구(ARTEX)를 10월 하순부터 공격에 투입하여**(III장 1.가)** 신규 AI 도구의 신속 도입 확인

2. 예방 중심 역량 강화 및 글로벌 사이버안보 기여

가. 예방 중심 대응 역량 강화 및 이를 위한 민간 협력 강화

▪ AI 도구 활용으로 해킹 수법이 정교해지면서 사후 대응만으로는 피해 차단에 한계가 있는바 공격 징후를 조기에 포착하고 신속히 공유하는 예방 중심 대응 강화 요망

- 금융·에너지 등 국가 기간시설의 취약점을 점검하고 신종 공격 수법에 대비한 보안 대책 보완

▪ 조기 포착과 신속 공유의 실효성은 민관 공조 수준에 좌우되는바 현재 운영 중인 민관 협력을 내실화하는 방안 검토

- 민관 합동 분석 협의체에서 수행 중인 가상자산 탈취 공동 분석을 내실화하고 공격 징후와 탈취 자금 흐름에 관한 분석 결과를 관계기관과 신속히 공유

71 Linus Folkerts 외, "Measuring AI Agents' Progress on Multi-Step Cyber Attack Scenarios," arXiv:2603.11214, March 2026, https://arxiv.org/html/2603.11214v1 (accessed: November 6, 2026). 능동 방어가 없는 모의 환경 결과임.

▪ 그간 국가 배후 APT 조직의 AI 도구 활용 평가는 민간 업체가 주도해 왔으나 공격 작업을 자동 수행하는 AI 에이전트까지 운용되기 시작한 만큼 공통 평가 지표**(Ⅳ장 1.가)**를 적용한 민간 분석 결과의 취합 · 발간 검토 要

- 美 행정부는 국무부 · FBI 다국적 경보와 재무부 제재 발표를 통해 IT 인력의 AI 도구 활용 현황 지속 평가 및 발표[72]

나. 글로벌 사이버안보 기여

▪ 북한 IT 인력이 위협 인식이 낮은 유럽 · 동남아 · 아프리카로 활동을 확대 중인바[73] 이들 지역과 글로벌 채용 플랫폼을 대상으로 한 식별 정보 공유를 우리 정부가 주도하는 방안 검토 要

- 구글은 美 수사 · 기소 강화로 취업이 제한되자 북한 IT 인력이 유럽으로 활동 대상을 확대했다고 평가하였으며 단일 인력이 위장 신분 12개 이상으로 유럽 방산 · 정부 기관 취업을 시도한 사례 확인[74]
- 독일 · 포르투갈 · 영국에서의 AI 생성 프로필 활용 취업 시도, 방산업체 · 정부 기관 대상 접근, 세르비아 가상자산 탈취 등 다수의 사례가 보도됨

▪ 이에 우리의 축적된 위협 데이터와 대응 경험을 글로벌 공공재로 제공하여 실전적 보안 지침을 제시함으로써 글로벌 책임강국 도약 기반 마련 및 AI 안보 기여도 제고 기대

72 U.S. Department of State, “Alert to Countries, Companies, and Other Entities Regarding North Korean IT Workers,” Media Note, July 31, 2026, https://www.state.gov/releases/office-of-the-spokesman/2026/07/alert-to-countries-companies-and-other-entities-regarding-north-korean-it-workers (accessed: November 4, 2026); U.S. Department of the Treasury, “Treasury Sanctions Facilitators of DPRK IT Worker Fraud Targeting U.S. Businesses”.

73 U.S. Department of State, “Alert to Countries, Companies, and Other Entities Regarding North Korean IT Workers”.

74 Jamie Collier, “DPRK IT Workers Expanding in Scope and Scale,” Google Threat Intelligence Group, Google Cloud Blog, April 1, 2025, https://cloud.google.com/blog/topics/threat-intelligence/dprk-it-workers-expanding-scope-scale (accessed: November 4, 2026).

참고문헌

(1) 논문

Chhabra, Anshuman, Shrestha Datta, Shahriar Kabir Nahin and Prasant Mohapatra. "Agentic AI Security: Threats, Defenses, Evaluation, and Open Challenges." IEEE Access. Vol. 14 (2026), pp.49455-49482.

Folkerts, Linus, Will Payne, Simon Inman, Philippos Giavridis, Joe Skinner, Sam Deverett, James Aung, Ekin Zorer, Michael Schmatz, Mahmoud Ghanem, John Wilkinson, Alan Steer, Vy Hong and Jessica Wang. "Measuring AI Agents' Progress on Multi-Step Cyber Attack Scenarios." arXiv:2603.11214 (March 2026).

Singer, Brian, Keane Lucas, Lakshmi Adiga, Meghna Jain, Lujo Bauer and Vyas Sekar. "Incalmo: An Autonomous LLM-assisted System for Red Teaming Multi-Host Networks." 2026 IEEE Symposium on Security and Privacy (San Francisco, CA, May 18-21, 2026).

Yao, Pengchao, Xin Sun, Changhua Sun, Bang Lv, Yilei Wang, Yifeng Wang and Yiming Mao. "Multi-Agents Assisted Autonomous Breach and Attack Simulation Framework in Power Information System." IEEE Access. Vol. 14 (2026), pp.139728-139739.

(2) 정부 · 국제기구 문서

한국은행, "2025년 북한 경제성장률 추정 결과," 2026년 7월 31일, https://www.bok.or.kr/portal/bbs/B0000501/view.do?nttId=11063399&menuNo=201264&programType=newsData&relate=Y&depth=201264 (검색일: 2026년 11월 2일).

KOTRA, "2025 북한 대외무역 동향," 2026년 7월, https://openknowledge.kotra.or.kr/handle/2014.oak/34038 (검색일: 2026년 11월 2일).

Federal Bureau of Investigation, "North Korea Responsible for $1.5 Billion Bybit Hack," I-022625-PSA, February 26, 2025, https://www.fbi.gov/investigate/cyber/alerts/2025/north-korea-responsible-for-1-5-billion-bybit-hack (accessed: November 3, 2026).

Federal Bureau of Investigation, "North Korean IT Workers Expanding Use of Facilitators, Broadening Recruitment," I-101326-PSA, October 13, 2026, https://www.ic3.gov/PSA/2026/PSA261013 (accessed: November 3, 2026).

Federal Bureau of Investigation, "North Korea Responsible for $6 Billion+ Binance Attack," October 31, 2026, https://www.fbi.gov/investigate/cyber/alerts/2026/north-korea-responsible-for-6-billion-binance-attack (accessed: November 2, 2026).

Federal Bureau of Investigation, "The FBI is releasing information advising the Democratic People's Republic of Korea (North Korea) was responsible for the theft of more than $6 billion USD in virtual assets from cryptocurrency exchange, Binance," Facebook October 31, 2026, https://www.facebook.com/FBI/posts/the-fbi-is-releasing-information-advising-the-democratic-peoples-republic-of-kor/1550586253781602/ (accessed: November 2, 2026).

Federal Bureau of Investigation, Cybersecurity and Infrastructure Security Agency and U.S. Department of the Treasury, "TraderTraitor: North Korean State-Sponsored APT Targets Blockchain Companies," AA22-108A, April 18, 2022, https://www.cisa.gov/news-events/cybersecurity-advisories/aa22-108a (accessed: November 3, 2026).

Federal Bureau of Investigation, Department of Defense Cyber Crime Center and National Police Agency of Japan, "FBI, DC3, and NPA Identification of North Korean Cyber Actors, Tracked as TraderTraitor, Responsible for Theft of $308 Million USD from Bitcoin.DMM.com," December 23, 2024, https://www.fbi.gov/news/press-releases/fbi-dc3-and-npa-identification-of-north-korean-cyber-actors-tracked-as-tradertraitor-responsible-for-theft-of-308-million-from-bitcoindmmcom (accessed: November 3, 2026).

Multilateral Sanctions Monitoring Team, "The DPRK's Violation and Evasion of UN Sanctions through Cyber and Information Technology Worker Activities," MSMT/2025/2, October 22, 2025, https://msmt.info/Publications/detail/MSMT Report/4221 (accessed: November 3, 2026).

OECD, "The agentic AI landscape and its conceptual foundations," OECD Artificial Intelligence Papers No. 56, February 2026, https://www.oecd.org/content/dam/oecd/en/publications/reports/2026/02/the-agentic-ai-landscape-and-its-conceptual-foundations_a9d4b451/396cf758-en.pdf (accessed: November 4, 2026).

United Nations Security Council, "DPRK Sanctions List," KPe.031, https://scsanctions.un.org/en/?keywords=dprk (accessed: November 3, 2026).

U.S. Department of Justice, "Justice Department Announces Coordinated, Nationwide Actions to Combat North Korean Remote Information Technology Workers' Illicit

Revenue Generation Schemes," June 30, 2025, https://www.justice.gov/opa/pr/justice-department-announces-coordinated-nationwide-actions-combat-north-korean-remote (accessed: November 3, 2026).

U.S. Department of State, "Alert to Countries, Companies, and Other Entities Regarding North Korean IT Workers," Media Note, July 31, 2026, https://www.state.gov/releases/office-of-the-spokesman/2026/07/alert-to-countries-companies-and-other-entities-regarding-north-korean-it-workers (accessed: November 4, 2026).

U.S. Department of State, U.S. Department of the Treasury and Federal Bureau of Investigation, "Guidance on the Democratic People's Republic of Korea Information Technology Workers," May 16, 2022, https://ofac.treasury.gov/media/923126/download?inline (accessed: November 3, 2026).

U.S. Department of the Treasury, "Treasury Sanctions North Korean State-Sponsored Malicious Cyber Groups," September 13, 2019, https://home.treasury.gov/news/press-releases/sm774 (accessed: November 3, 2026).

U.S. Department of the Treasury, "Treasury Targets DPRK Malicious Cyber and Illicit IT Worker Activities," May 23, 2023, https://home.treasury.gov/news/press-releases/jy1498 (accessed: November 3, 2026).

U.S. Department of the Treasury, "Sanctions Imposed on DPRK IT Workers Generating Revenue for the Kim Regime," July 8, 2025, https://home.treasury.gov/news/press-releases/sb0190 (accessed: November 3, 2026).

U.S. Department of the Treasury, "National Proliferation Financing Risk Assessment," March 2026, https://home.treasury.gov/system/files/246/2026-NPFRA.pdf (accessed: November 3, 2026).

U.S. Department of the Treasury, "Treasury Sanctions Facilitators of DPRK IT Worker Fraud Targeting U.S. Businesses," October 28, 2026, https://home.treasury.gov/news/press-releases/sb0669 (accessed: November 3, 2026).

(3) 기업 · 연구기관 보고서

Anthropic, "Detecting and Countering Misuse of AI: August 2025," August 27, 2025, https://www.anthropic.com/news/detecting-countering-misuse-aug-2025 (accessed: November 4, 2026).

Anthropic, "Disrupting the First Reported AI-Orchestrated Cyber Espionage Campaign," November 13, 2025, https://assets.anthropic.com/m/ec212e6566a0d47/original/Disrupting-the-first-reported-AI-orchestrated-cyber-espionage-campaign.pdf (accessed: November 4, 2026).

Anthropic, "Disrupting the first reported AI-orchestrated cyber espionage campaign," November 13, 2025, https://www.anthropic.com/news/disrupting-AI-espionage (accessed: November 4, 2026).

Anthropic, "Detecting and Countering Misuse of AI: September 2026," September 10, 2026, https://www.anthropic.com/threat-intelligence-report-september-2026 (accessed: November 4, 2026).

Binance, "29 October Binance Security Incident: Initial Findings and Response Update," X November 3, 2026, https://x.com/binance/status/2117626126214504729 (accessed: November 4, 2026).

Chainalysis, "Binance Hack: Tracing the $6.4 Billion Theft to North Korea," 2026, https://www.chainalysis.com/blog/binance-exploit-latest-findings-and-response/ (accessed: November 2, 2026).

Cloud Security Alliance AI Safety Initiative, "Microsoft's AI Threat Intelligence: Documenting the Full AI-Accelerated Attack Lifecycle: Unofficial AI-assisted Research," March 8, 2026, https://labs.cloudsecurityalliance.org/wp-content/uploads/2026/03/CSA_research_note_microsoft_ai_attack_lifecycle_intelligence_20260308-csa-styled.pdf (accessed: April 17, 2026).

Jamie Collier, "DPRK IT Workers Expanding in Scope and Scale," Google Threat Intelligence Group, Google Cloud Blog April 1, 2025, https://cloud.google.com/blog/topics/threat-intelligence/dprk-it-workers-expanding-scope-scale (accessed: November 4, 2026).

Elliptic, "The Binance hack is laundering on autopilot," October 31, 2026, https://www.elliptic.co/insights/the-binance-hack-is-laundering-on-autopilot/ (accessed: November 3, 2026).

Elliptic, "How North Korea is laundering the $6.4 billion stolen from Binance," November 2, 2026, https://www.elliptic.co/insights/how-north-korea-is-laundering-the-6-4-billion-stolen-from-binance/ (accessed: November 2, 2026).

Google Threat Intelligence Group, "GTIG AI Threat Tracker: Advances in Threat Actor Usage of AI Tools," Google Cloud Blog November 2025, https://cloud.google.com/blog/topics/threat-intelligence/threat-actor-usage-of-ai-tools (accessed: November 6, 2026).

Ross Inman and Adrian Hernandez, "UNC1069 Targets Cryptocurrency Sector with New Tooling and AI-Enabled Social Engineering," Google Cloud Blog February 9, 2026, https://cloud.google.com/blog/topics/threat-intelligence/unc1069-targets-cryptocurrency-ai-social-engineering (accessed: November 3, 2026).

Adam Meyers, "CrowdStrike 2026 Global Threat Report: The Evasive Adversary Wields AI," CrowdStrike Blog February 24, 2026, https://www.crowdstrike.com/en-us/blog/crowdstrike-2026-global-threat-report-findings/ (accessed: November 4, 2026).

Microsoft, "How Microsoft names threat actors," Microsoft Learn, https://learn.microsoft.com/en-us/defender-xdr/microsoft-threat-actor-naming (accessed: November 3, 2026).

Microsoft Threat Intelligence, "AI as Tradecraft: How Threat Actors Operationalize AI," March 6, 2026, https://www.microsoft.com/en-us/security/blog/2026/03/06/ai-as-tradecraft-how-threat-actors-operationalize-ai/ (accessed: November 2, 2026).

Microsoft Threat Intelligence, "Tracking AI adoption in North Korean threat actor tradecraft," October 29, 2026, https://www.microsoft.com/en-us/security/blog/2026/10/29/tracking-ai-adoption-in-north-korean-threat-actor-tradecraft/ (accessed: November 3, 2026).

Seongsu Park, "Suspected TraderTraitor Group Uses Trojanized Terraform Provider to Deliver Cross-Platform Malware," Zscaler ThreatLabz September 30, 2026, https://www.zscaler.com/blogs/security-research/suspected-tradertraitor-group-uses-trojanized-terraform-provider-deliver (accessed: November 3, 2026).

Albert Priego, Alex Delamotte and Matěj Havránek, "Don't Call Us, We'll Call Your APIs | TraderTraitor Backdoors Resurface on Victim With No Crypto Ties," SentinelLABS September 18, 2026, https://www.sentinelone.com/labs/dont-call-us-well-call-your-apis-tradertraitor-backdoors-resurface-on-victim-with-no-crypto-ties/ (accessed: November 3, 2026).

ThreatDown, "2026 State of Malware," 2026, https://www.threatdown.com/wp-content/

uploads/2026/01/2026-TD-State-of-Malware-1.pdf (accessed: April 17, 2026).

TRM Labs, "The 2026 AI-in-Crime Adoption Index," August 2026, https://www.trmlabs.com/reports-and-whitepapers/the-2026-ai-in-crime-adoption-index (accessed: November 6, 2026).

TRM Labs, "North Korea Steals USD 6.36 Billion From Binance in the Largest Hack in History," October 30, 2026, https://www.trmlabs.com/resources/blog/north-korea-steals-usd-6-36-billion-from-binance-in-the-largest-hack-in-history (accessed: November 2, 2026).

(4) 언론 기사

"김여정 조선로동당 중앙위원회 부장 담화 발표," 『조선중앙통신』 2026년 11월 3일.

신규진, "'北 해킹 인력 8400여명'…국정원-합참 공식 평가," 『동아일보』 2024년 8월 9일, https://www.donga.com/news/Politics/article/all/20240809/126440040/1 (검색일: 2026년 11월 4일).

Jamie Crawley, "엘립틱, 64억 달러 규모 바이낸스 횡령 자금 세탁에 AI 사용 흔적 없어," 『코인데스크』 2026년 11월 2일, https://www.coindesk.com/ko/tech/2026/11/02/no-proof-of-ai-in-north-korean-hackers-move-of-stolen-binance-funds-through-thousands-of-wallets (검색일: 2026년 11월 3일).

"FBI links North Korea to theft of $6 billion from Binance," Reuters October 31, 2026, https://www.reuters.com/world/fbi-links-north-korea-theft-6-billion-from-binance-2026-10-31/ (accessed: November 2, 2026).

Ravie Lakshmanan, "FBI Links North Korea's TraderTraitor to Record $6.43 Billion Binance Crypto Heist," The Hacker News November 1, 2026, https://thehackernews.com/2026/11/fbi-links-north-koreas-tradertraitor-to.html (accessed: November 3, 2026).

Mehab Qureshi, "FBI names who took $6B+ from Binance," TheStreet October 31, 2026, https://www.thestreet.com/crypto/markets/fbi-names-who-took-6b-from-binance (accessed: November 2, 2026).

Abstract

AI Tool Use by State-Sponsored Hacking Groups (APTs) and Its Security Implications

Minjung Kim
(Institute for National Security Strategy)

On October 29, 2026, roughly $6.3 billion in virtual assets was stolen from the cryptocurrency exchange Binance, the largest single theft on record. The U.S. Federal Bureau of Investigation (FBI) attributed the attack to TraderTraitor, a North Korean state-sponsored hacking group. The incident has deepened concern over the growing sophistication of North Korea's cyber capabilities. It also makes it necessary to examine how far North Korean advanced persistent threat (APT) groups have adopted AI and what security threats may follow. Microsoft has assessed Jade Sleet, which has been tracked as the same group as TraderTraitor, as having reached the "operationalized" stage of AI adoption, in which AI is used routinely and repeatedly across operational workflows. This assessment suggests that the threat posed by North Korean APT groups' use of AI may be undergoing a qualitative change. This study reviews the evidence of AI use by these groups and analyzes how AI has changed the tempo of their operations and the capabilities of their personnel. There is considerable evidence that North Korean groups have used AI to prepare lures and to assist in malware development. No case has yet been confirmed, however, in which AI carried out every stage on its own, from intrusion to monetization. Should North Korea adopt the AI-driven targeting and attack methods seen in operations by a Chinese state-sponsored hacking group, North Korea's cyberattack capabilities are likely to grow substantially. The study therefore proposes three policy responses. First, AI involvement in North Korean APT operations needs to be identified early, and preparations should be made for these groups'

move into operating agentic AI. Second, to strengthen prevention-focused response capabilities, the public-private joint analysis council should be given greater substance, and mechanisms for rapidly sharing attack indicators and analysis of virtual asset theft and money laundering should be reinforced. Third, North Korean IT workers are extending their activities into new regions. In response, South Korea should share threat intelligence and response experience with Europe, Southeast Asia, and other regions, and work with global hiring platforms to expand its contribution to international cybersecurity.

Keywords: Artificial Intelligence (AI), AI Agents, Advanced Persistent Threat (APT), Cybersecurity, Coral Sleet, Complex Security Threats

Institute For National Security Strategy

www.inss.re.kr

본지에 실린 내용은 집필자 개인의 견해이며,
국가안보전략연구원의 공식입장이 아닙니다.

INSS 전략보고

November 2026
No. 414

06295 서울시 강남구 언주로 120 인스토피아 빌딩
02-6191-1000 02-6191-1111 www.inss.re.kr

국가안보전략연구원