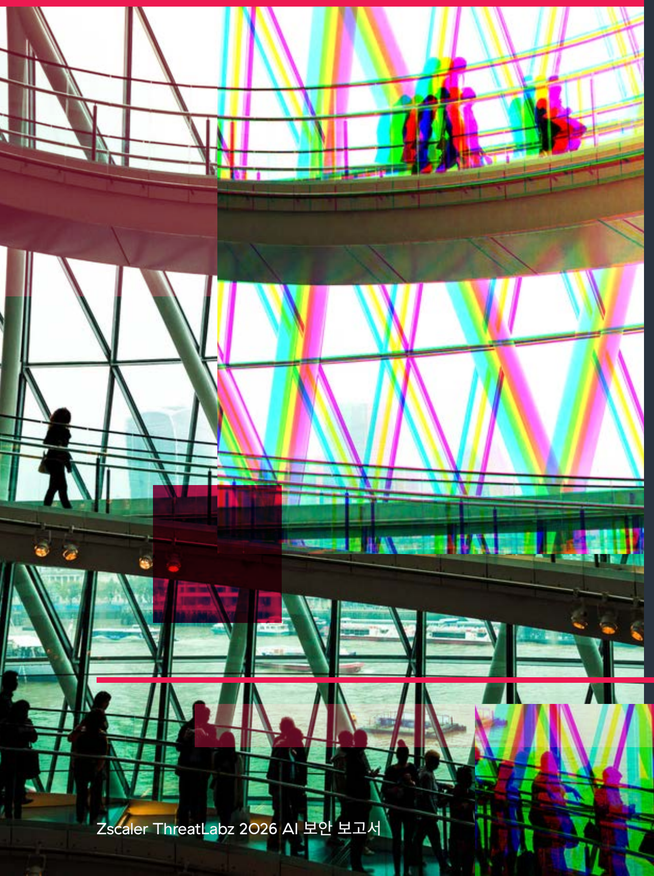
zscaler™

ThreatLabz 2026
AI 보안 보고서



목차



핵심 요약 03

주요 결과 05

AI/ML 사용 동향 07

- 전 세계 AL/ML 트랜잭션 증가 08
- 주요 LLM 벤더, 애플리케이션 및 부서 10
- 차단된 트랜잭션 13
- AI 애플리케이션으로 전송된 데이터 14
- AI 애플리케이션으로 인한 데이터 손실 15
- 임베디드 AI의 부상 17
- 산업별 AI/ML 사용 현황 18
- 국가별 AI/ML 사용 현황 22

엔터프라이즈 AI의 위험 및 위협 환경 26

- 사례 연구: GenAI로 강화된 악성 소프트웨어 및 북한 연계 캠페인에서의 사회 공학 28
- 사례 연구: 남아시아 지역을 대상으로 한 캠페인에서 나타난 새로운 AI 지표 33
- 사례 연구: 엔터프라이즈 AI 시스템에서 실제로 문제가 되는 것은 무엇인가 34

AI 거버넌스의 최신 국면 38

2026년 AI 보안 전망 40

모범 사례: 안전한 엔터프라이즈 AI 도입 42

Zscaler의 포괄적인 AI 보호 방식 45

연구 방법론 48

- ThreatLabz 소개 48

©2026 Zscaler, Inc. 모든 권리 보유.

핵심 요약

2025년 AI의 일상적인 현실은 속도, 규모, 그리고 끊임없는 움직임으로 정의되었습니다.

이제 기업들은 더 빠른 속도, 자동화된 의사 결정, 생산성 향상을 위해 비즈니스 전반에 걸쳐 인공지능 및 머신러닝(AI/ML)에 의존하고 있습니다. AI는 불과 몇 년 전만 해도 비현실적으로 여겨졌을 속도로 개발, 소통, 연구 및 운영을 지원하고 있습니다. 하지만 이러한 가속화에는 점점 더 많은 단점이 따르고 있습니다. 더욱 민감한 데이터가 더욱 많은 AI/ML 애플리케이션을 거쳐 이동하면서, 가시성은 떨어지고 가드레일도 부족한 경우가 많습니다.

AI의 영향력이 확대되면서 기업 공격 표면이 넓어졌고, 지난 한 해 동안 위협 행위자들은 이를 재빨리 포착했습니다. 진입 장벽이 낮아지고 현실감이 높아짐에 따라 공격은 더욱 빠르고 설득력 있게 이루어지고 있으며, 에이전트형 및 반자율형 AI의 오용에 대한 조기 조짐은 위협이 진화하는 방식의 변화를 시사합니다. 동시에 조직들은 섀도 AI와 임베디드 AI부터 환각 및 보안이 취약한 프라이빗 데이터 모델에 이르기까지 점점 더 다양해지는 위험에 직면하고 있습니다.

AI가 모든 것에 영향을 미치는 환경에서 기업은 어떻게 보안을 유지하고, AI 기반 혁신을 가능하게 하며, AI 기반 위협으로부터 방어할 수 있을까요? (물론, 비즈니스에 지장을 주지 않으면서 말이죠.)

Zscaler ThreatLabz의 2026 AI 보안 보고서는 기업들이 이러한 균형을 어떻게 유지하고 있는지 살펴봅니다. 이 보고서는 2025년 1월부터 2025년 12월까지 Zscaler Zero Trust Exchange™에서 관찰된 9,893억 건의 AI/ML 트랜잭션 분석을 바탕으로 작성되었으며, 전 세계 환경에서 AI가 실제로 어떻게 사용되고 (그리고 제한되는지)에 대한 실질적인 관점을 제공합니다.

데이터는 가속화가 계속되고 있음을 보여줍니다. 기업의 AI/ML 활동은 전년 대비 83% 증가했으며, 데이터 전송량은 92.6% 증가하여 18,000테라바이트(TB)를 넘어섰습니다. 이러한 규모에서 AI는 개별 도구들의 집합이라기보다는 기업 데이터를 지속적으로 이동시키고 변환하는 상시 가동 인프라처럼 작동합니다. 하지만 액세스에는 여전히 제약이 따릅니다. 여러 기관들이 AI/ML 관련 트랜잭션의 39%를 차단했으며, 이는 데이터 노출, 개인정보 보호 및 정책 시행에 대한 지속적인 우려를 반영합니다.

사용 패턴을 보면 가치와 위험이 교차하는 지점도 드러납니다. Codeium, Grammarly, ChatGPT와 같이 직원들이 가장 많이 의존하는 AI 애플리케이션은 업무 수행 방식의 중심에 자리 잡고 있으며, 가장 높은 수준의 활동을 주도하는 동시에 위험 요소 분석에서도 가장 중요한 부분으로 나타납니다.

2026년에는 AI 보안이 AI/ML 애플리케이션 제어 그 이상을 의미합니다. 기업 전반에 걸쳐 AI가 발견되고, 구축되고, 사용되고, 관리되는 방식을 보호하는 것이 핵심입니다. 기업은 AI 사용 현황과 위험에 대한 가시성, AI 시스템과 데이터를 실시간으로 강화하는 보호 조치, 그리고 혁신을 지속하면서 접근을 안전하게 제어하는 일관된 제어 체계를 필요로 합니다. 이 보고서는 AI 보안을 형성하는 트렌드와 현실을 심층 분석하고, 위험을 줄이며 AI를 안전하게 도입하고자 하는 기업에 필요한 지침을 제공합니다.

©2026 Zscaler, Inc. 모든 권리 보유.

기업 리더를 위한 시사점

- **AI는 이제 기업 인프라의 필수 요소가 되었습니다.**
  거의 1조 건에 달하는 AI 트랜잭션은 지속적이고 상시 가동되는 운영을 의미합니다. 안전하고 확장 가능한 AI 도입을 위해서는 클라우드, 신원 관리, 데이터 관리와 동일한 수준의 엄격한 관리가 필요합니다.

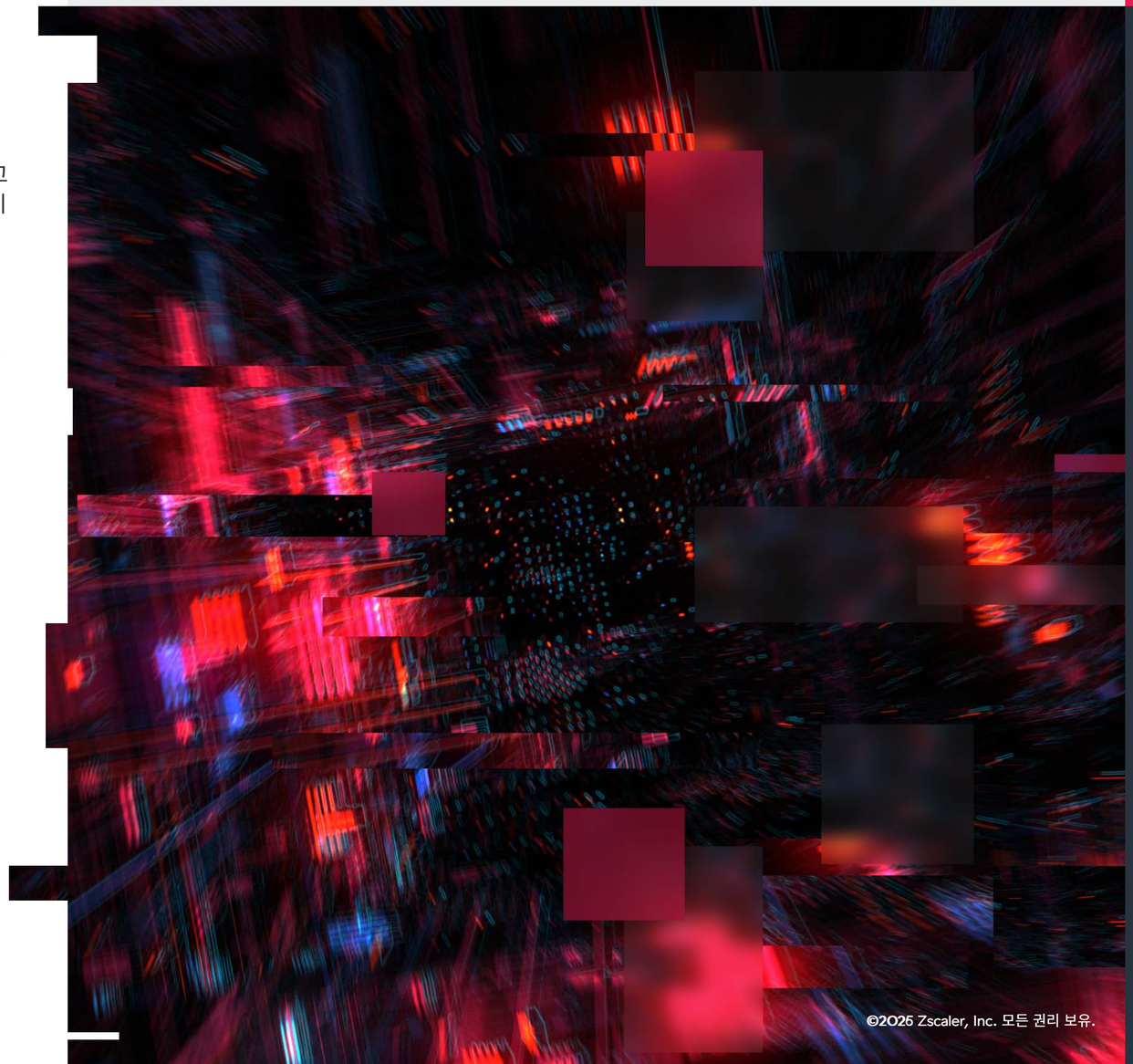
- **데이터 노출 위험은 이제 의도가 아닌 규모에 비례합니다.**
  페타바이트급의 데이터가 AI 워크플로를 통해 이동하면서, 사용이 승인되고 비즈니스 의도와 부합하는 경우에도 반복성과 속도로 인해 노출 위험이 증가합니다.

- **승인된 AI가 주요 위험 요소입니다.**
  주류로 자리 잡은 공인 AI 도구는 기업 AI 활동 및 데이터 상호 작용의 대부분을 차지합니다. 섀도 AI는 여전히 중요한 문제이지만, 미승인 도구만 해결한다고 해서 AI 관련 위험과 노출을 완전히 완화할 수는 없습니다.

- **보안은 AI 도입을 저해하는 요소입니다.**
  AI 관련 트랜잭션의 39%가 차단되는 상황에서, 정책 집행은 AI 활용 방식을 적극적으로 좌우하고 있습니다. 이는 AI에 대한 저항이 아니라, 리더들이 혁신 속도와 위험 감수 수준 사이의 균형을 맞추려는 과정에서 나타나는 거버넌스를 실행에 옮기고 있음을 보여줍니다.

- **기존 보안 모델은 AI 워크플로와 맞지 않습니다.**
  인간의 속도에 맞춘 활동과 정적인 데이터를 기반으로 설계된 제어 방식으로는 기계가 주도하는 빈번한 AI 상호 작용을 따라잡을 수 없습니다.

- **대규모 AI 관리 역량을 갖춘 조직이 경쟁 우위를 확보할 것입니다.**
  강력한 인라인 제어 기능을 통해 AI를 광범위하게 활용할 수 있도록 지원하는 기업은 관리되지 않은 위험으로 인해 사용을 전면 제한해야 하는 기업보다 더 빠르게 변화할 것입니다.



©2025 Zscaler, Inc. 모든 권리 보유.

주요 결과

ThreatLabz는 2025년 1월부터 2025년 12월까지 Zscaler 클라우드에서 발생한 **9,893억 건의 AI 및 ML 트랜잭션**을 분석했습니다. 아래 주요 결과는 비교 분석을 위해 다양한 기간*에 걸친 데이터를 기반으로 합니다.

엔터프라이즈 AI 사용량은 꾸준히 증가세를 보이고 있습니다. AI/ML 활동은 전년 대비 83% 증가하여 3,400개 이상의애플리케이션으로 구성된 생태계 전반에서 거의 1조 건의 트랜잭션을 기록했습니다.

기업들은 AI 도구로 점점 더 많은 양의 데이터를 전송하고 있습니다. 총 18,033TB의 데이터가 AI/ML 애플리케이션으로 전송되었으며, 이는 전년 대비 93% 증가한 수치입니다.

높은 차단율은 지속적인 위험 관리 노력을 시사합니다. 기업들은 전체 AI/ML 트랜잭션의 39%를 차단했는데, 이는 AI 사용이 확대됨에 따라 데이터 노출, 개인정보 보호 및 정책 조율에 대한 우려가 지속되고 있음을 보여줍니다.

엔터프라이즈 AI 시스템은 보안에 매우 취약합니다. Zscaler의 레드티밍 전문가들은 대부분의 엔터프라이즈 AI 시스템이 단 16분 만에 침해될 수 있다는 사실을 발견했으며, 테스트 대상 시스템의 100%에서 치명적인 취약점을 찾아냈습니다.

* 데이터 수집 기간:

- 연간 및 전년 대비 분석: 2025년 1월~12월, 2024년 동기간과 비교.
- DLP 위반 데이터 및 국가별 데이터: 2025년 6월~2025년 12월.

©2026 Zscaler, Inc. 모든 권리 보유.

OpenAI는 LLM 분야에서 최고의 벤더로 자리매김했습니다. OpenAI는 LLM 기반 엔터프라이즈 트랜잭션의 대부분(Codeium보다 3배 많음)을 차지하며 현재 사실상 LLM 솔루션으로 자리 잡았습니다.

ChatGPT는 DLP 위반 사례의 압도적인 대부분을 차지합니다. 분석된 모든 AI/ML 애플리케이션에서 ChatGPT는 4억 1천만 건의 DLP 정책 위반을 발생시켰으며, 이는 다양한 맥락을 이해하는 AI 어시스턴트와 관련된 기업 위험을 입증합니다.

통합 생산성 앱은 기업 AI 활용의 핵심입니다. Grammarly는 트랜잭션 규모 기준으로 1위 애플리케이션이 되었는데, 이는 커뮤니케이션 및 비즈니스 프로세스 내에서 직접 작동하는 AI에 대한 의존도를 반영합니다.

금융·보험 및 제조 부문이 기업 AI 활용을 다시 한번 주도했습니다. 이 두 부문은 3년 연속으로 현대화 노력과 방대한 문서화 워크플로에 힘입어 AI/ML 트래픽에서 가장 큰 비중(각각 23%와 20%)을 차지했습니다.

미국은 여전히 AI/ML 트랜잭션의 주요 공급원이었습니다. 트랜잭션 활동은 미국에 집중되었으며, 미국이 전체 트랜잭션의 38%를 차지했고, 그 뒤를 인도(14%)와 캐나다(5%)가 이었습니다.

AI 도입이 확대됨에 따라 기업의 공격 표면이 점점 더 넓어지고 있습니다. 기업 워크플로 전반에 걸쳐 AI 사용이 증가하면서 데이터 및 접근 권한 노출 경로가 늘어나 데이터 유출, 즉각적인 오용, AI 기반 공격 가능성이 높아지고 있습니다. 이는 Zero Trust 아키텍처와 AI 기반 보안 제어의 필요성을 더욱 강조합니다.

©2026 Zscaler, Inc. 모든 권리 보유.

AI/ML 사용 동향

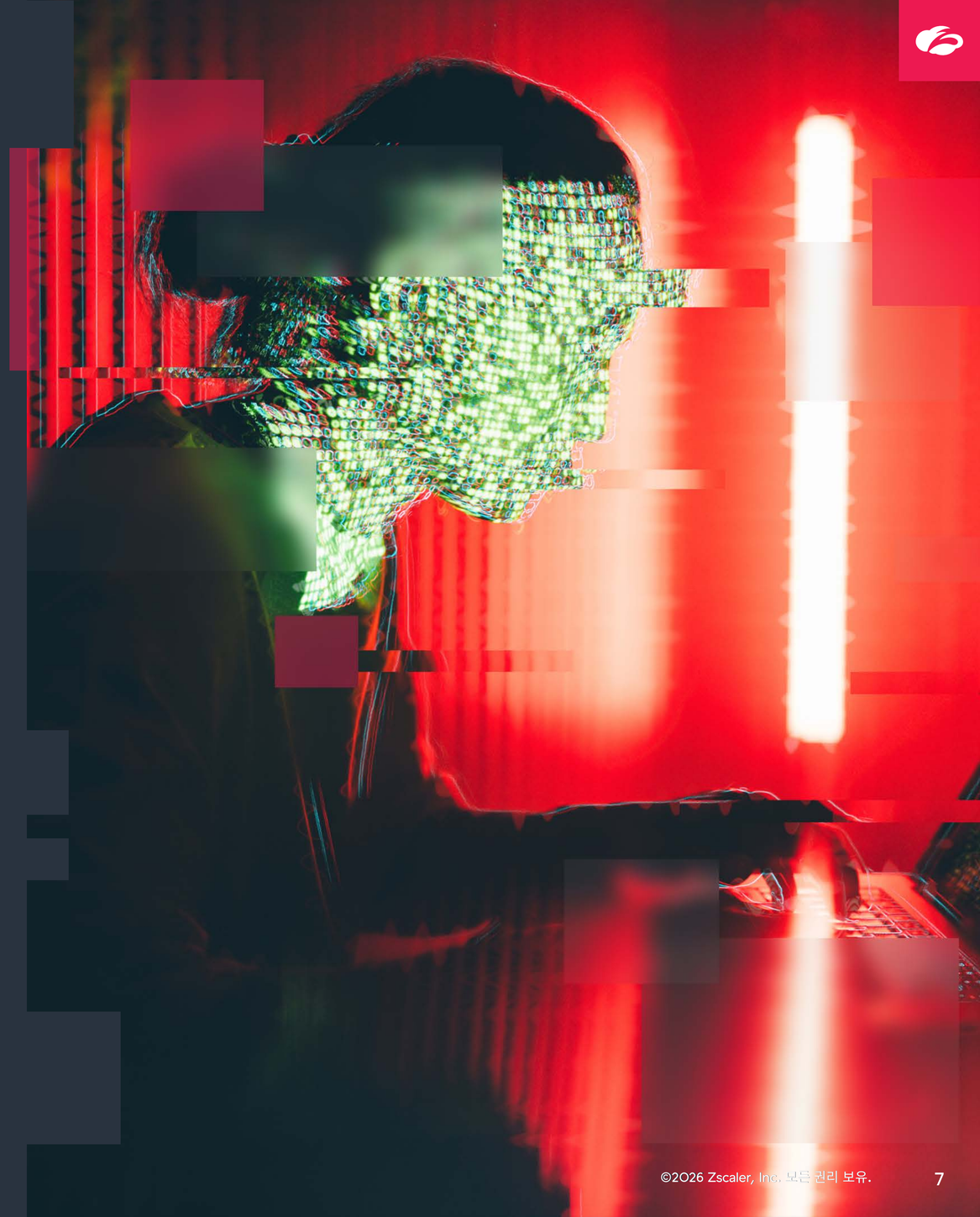
기업의 AI 활용은 2025년에도 가파르고 꾸준한 증가세를 이어갔습니다.

ThreatLabz의 AI 사용 동향 분석에 따르면 현재 AI/ML 트랜잭션을 구동하는 애플리케이션이 3,400개 이상이며, 이는 전년 대비 4배 증가한 수치입니다. 이러한 앱들 중 상당수는 트래픽이 제한적이지만, 애플리케이션 생태계 자체의 엄청난 성장세만으로도 의미 있는 지표입니다. 이는 AI 기능이 벤더, 사용 사례 및 비즈니스 기능 전반에 걸쳐 얼마나 빠르게 확산되고 있는지를 보여주며, 기회와 노출을 모두 확대하고 있습니다.

ThreatLabz는 이러한 성장이 실제 기업 환경에서 어떻게 활용되는지 이해하기 위해 여러 계층에 걸쳐 AI/ML 활동을 분석했습니다.

- **전체 AI/ML 트랜잭션** : URL 카테고리를 기준으로 허용된 활동과 차단된 활동을 모두 포함합니다.
- **LLM 벤더 순위**: 어떤 모델 제공업체가 가장 많은 AI/ML 트래픽을 생성하고 기업 AI 워크플로를 지원하는지 파악합니다.
- **주요 AI/ML 애플리케이션**: 엔터프라이즈 AI 활동 및 트래픽 규모를 주도하는 특정 애플리케이션을 중점적으로 보여줍니다.
- **부서별 AI 사용 현황**: AI 애플리케이션 사용량이 많은 기업 부서들을 매핑하여 일상 업무에서 AI가 어떻게 적용되고 있는지 파악합니다.

이러한 관점을 통해 우리는 기업 전반에 걸쳐 AI가 실제로 어떻게 도입되고 있는지, 그리고 사용량, 의존도 및 위험이 어디에서 수렴되는지에 대한 종합적인 시각을 제공하고자 합니다.



©2026 Zscaler, Inc. 모든 권리 보유.

AI/ML 트랜잭션의 글로벌 성장세

2025년 AI/ML 관련 트랜잭션 건수는 9,893억 건으로 1조 건에 육박했습니다. 이러한 성장의 상당 부분은 ChatGPT, Grammarly, Codeium과 같은 사용량이 많은 애플리케이션과 관련이 있습니다.

트랜잭션 규모별 AI/ML 사용 동향

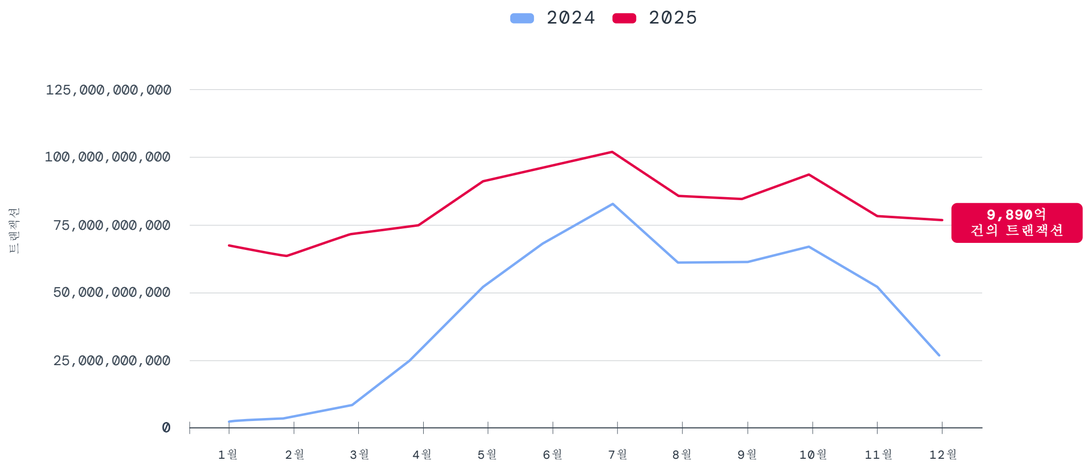


그림 1: AI/ML 트랜잭션의 전년 대비 비교(2025년 1월~12월)



주요 결과

3,400개 이상의 애플리케이션으로 구성된 생태계 전반에서 AI/ML 관련 활동이 전년 대비 91% 증가했습니다.

©2026 Zscaler, Inc. 모든 권리 보유.

예년과 마찬가지로 트래픽의 일부는 "일반 AI 애플리케이션"에 속합니다. 이는 특정 알려진 애플리케이션에 매핑되지는 않지만 Zscaler의 AI/ML 기반 URL 분류 기능을 통해 AI 관련 활동으로 식별된 AI/ML 트랜잭션을 반영합니다. 이 기능은 텍스트, 이미지 및 기타 콘텐츠 신호를 분석하여 AI 관련 활동을 인식합니다. 새로운 AI 애플리케이션은 수동으로 분류할 수 있는 속도보다 훨씬 빠르게 등장하므로, 이전에 알려지지 않았던 AI 트래픽 소스를 탐지하고 보안 정책을 시행하는 것이 필수적입니다.

달리 명시되지 않는 한, 이 보고서의 후속 분석은 기밀 애플리케이션에만 초점을 맞추었습니다. 이러한 접근 방식을 통해 기존 AI/ML 애플리케이션을 통해 AI 도입 현황에 대한 가시성을 확보할 수 있습니다.

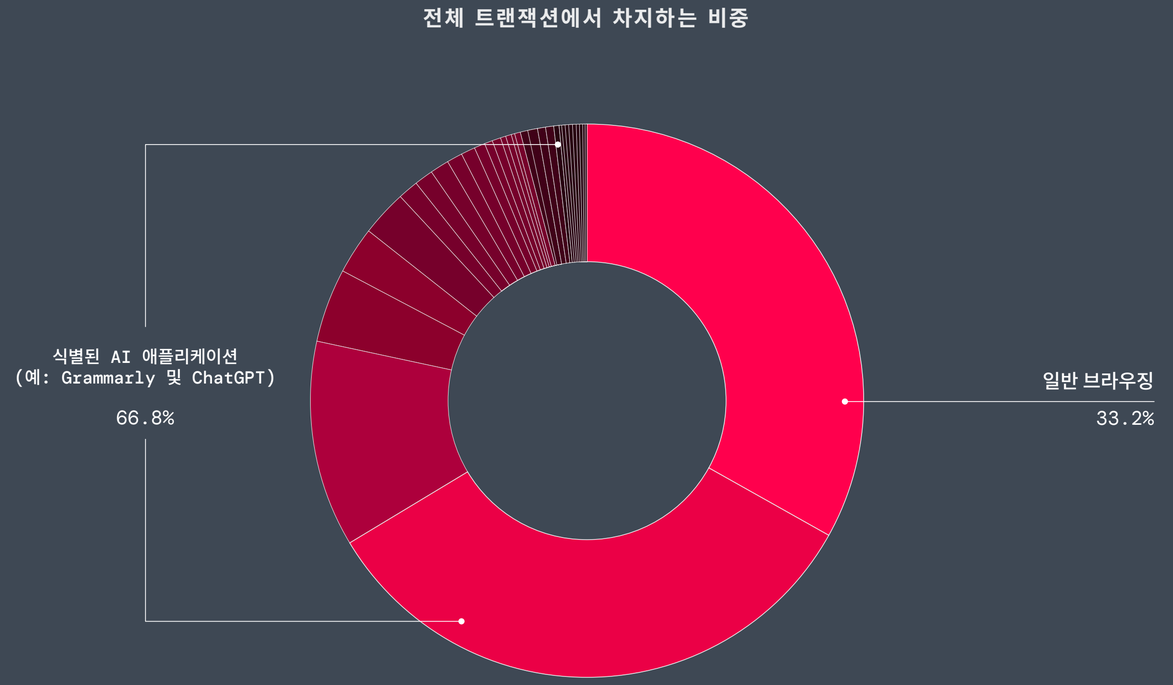


그림 2: 일반 AI 애플리케이션과 분류된 AI 애플리케이션 전반에 걸친 AI/ML 트랜잭션 분포

©2026 Zscaler, Inc. 모든 권리 보유.

주요 LLM 벤더, 애플리케이션, 부서

LLM 벤더를 통해 기업의 AI 활용 현황을 살펴보면 AI가 대규모 환경에서 어떻게 운영되는지에 대한 독특한 관점을 얻을 수 있습니다. 직원들이 매일 개별 애플리케이션 및 기능과 상호 작용하는 동안 트랜잭션 패턴을 통해 어떤 모델 제공업체가 이러한 경험을 지속적으로 뒷받침하는지 파악할 수 있습니다. 벤더 수준의 가시성은 AI 도입이 표면 아래에서 어떻게 진행되고 있는지 파악하는 데 유용한 방법입니다.

주요 LLM 벤더별 조사 결과

- **OpenAI**는 2025년에도 1,310억 건의 트랜잭션을 기록하며 LLM 벤더 중 명실상부한 선두 자리를 지켰으며, 이는 가장 근접한 경쟁업체의 3배가 넘는 수치입니다. 8월에 출시된 GPT-5는 코딩, 멀티모달 추론, 복잡한 작업 실행 등 다양한 분야에서 활용 범위를 넓혔습니다. 또한, 강화된 개인정보 보호 및 모델 격리 기능을 포함한 OpenAI의 확장된 엔터프라이즈 API 옵션은 코파일럿 및 AI 기반 SaaS 기능의 백엔드로서의 역할을 더욱 공고히 했습니다.

- **Codeium**(2025년 Windsurf로 브랜드 변경)은 기업용 LLM 트래픽(420억 건의 트랜잭션)에서 두 번째로 큰 공급업체로 부상했습니다. 이러한 성장은 소프트웨어 개발 파이프라인과 엔지니어링 환경에서 자주 사용되는 코딩 중심의 독자적인 모델 덕분인 것으로 보입니다. 이는 이어지는 부서별 분석에서 엔지니어링 부서가 AI를 가장 활발하게 사용하는 것을 보여주는 결과와 일치합니다.

- **Perplexity**는 작년 트랜잭션 규모 기준 3위(120억 건)를 차지했습니다. AI 기반 검색 외에도 자체 개발한 LLM을 활용하여 응답 엔진을 구동합니다. 따라서 기업의 Perplexity 사용은 AI 기반 연구 및 지식 통합에 대한 의존도가 높아지고 있음을 보여줍니다.

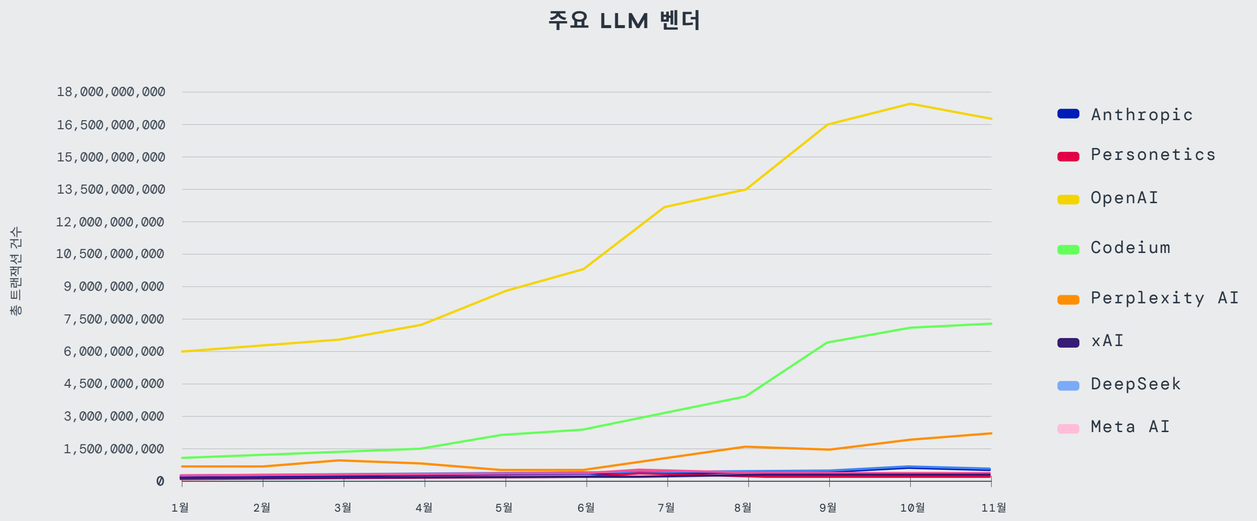


그림 3: 2025년 LLM 벤더 트랜잭션 추이

©2026 Zscaler, Inc. 모든 권리 보유.

트랜잭션 규모는 연구, 편집, 글쓰기, 코딩, 번역 및 협업과 같이 업무 흐름에 직접적으로 관여하는 가운데 널리 사용되는 애플리케이션에 여전히 크게 집중되어 있습니다.

주요 애플리케이션 조사 결과

- **Grammarly**는 기업 환경에서 가장 활발하게 사용되는 AI/ML 애플리케이션으로 부상했으며(전체 트랜잭션의 38.7%), 총 트랜잭션 규모에서 ChatGPT를 제쳤습니다. 요약부터 고급 재작성 및 어조 조정에 이르기까지 다양한 기능을 제공하는 Grammarly가 기업의 일상적인 콘텐츠 워크플로에서 중요한 위치를 차지하는 이유는 분명합니다.
- **ChatGPT**는 조사, 초안 작성 및 분석을 위해 다양한 직무에서 널리 사용되는 범용 어시스턴트(14.2%)로서 여전히 지배적인 위치를 차지하고 있으며, 기업 데이터에 대한 일반적인 접점 역할을 합니다.
- **Codeium**이 상위 5위(5%)에 진입하면서 AI가 소스 코드와 독점 로직이 일상적으로 처리되는 소프트웨어 개발 작업의 중요한 부분으로 자리 잡았음을 보여줍니다.
- **DeepL**은 글로벌 기업에서 지속적으로 높은 도입률(3.3%)을 보이며 비즈니스 핵심 콘텐츠 전반에 걸쳐 다국어 커뮤니케이션을 지원하고 있습니다.
- **Microsoft Copilot**은 Microsoft 365와의 긴밀한 통합 및 일상적인 생산성 작업 자동화에서의 역할에 힘입어 상위 5위(3%)에 들었습니다.

주요 AI 애플리케이션

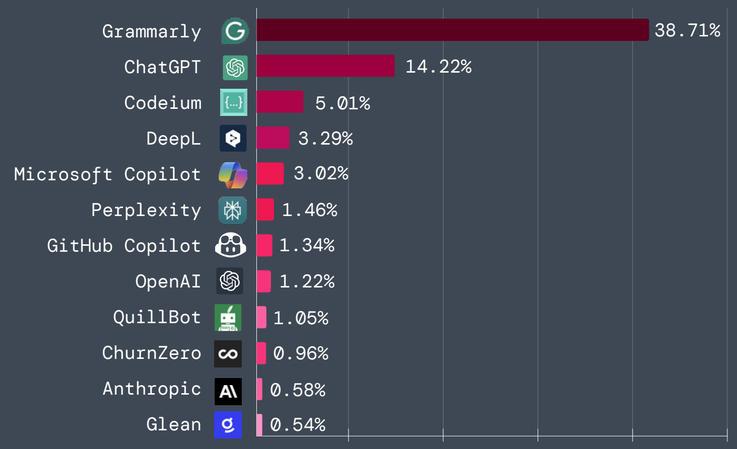


그림 4: 주요 AI 애플리케이션이 주도하는 전체 AI/ML 트랜잭션 비율

참고: Zscaler Zero Trust Exchange는 ChatGPT 트랜잭션을 OpenAI의 다른 트랜잭션과 별도로 추적합니다.

트랜잭션 처리량 기준 상위 20개 AI/ML 애플리케이션

애플리케이션	총 트랜잭션
Grammarly	327,311,080,013
ChatGPT	120,227,890,252
Codeium	42,337,652,986
DeepL	27,847,680,087
Microsoft Copilot	25,503,137,940
Perplexity	12,386,054,978
GitHub Copilot	11,348,420,722
OpenAI	10,352,420,115
QuillBot	8,913,115,535
ChurnZero	8,153,526,358
Anthropic	4,922,983,385
Glean	4,542,501,122
GliaCloud	3,249,239,347
Claude	2,850,954,278
Google Gemini	2,604,461,019
SundaySky	2,483,835,170
Yellow Messenger	1,734,555,650
Cresta	1,585,454,178
Poe	1,483,703,558

©2026 Zscaler, Inc. 모든 권리 보유.

전반적인 사용량에서 어떤 AI 애플리케이션이 우세한지를 넘어, 다음 단계의 분석에서는 도구에서 팀으로 전환합니다.

ThreatLabz는 기업의 일반 부서를 대상으로 AI/ML 트래픽을 분석하여 AI가 실제로 어떻게 활용되고 있는지 보다 자세히 파악하고자 했습니다. 이 관점은 사용량이 상당한 애플리케이션(최소 백만 건의 트랜잭션)에 초점을 맞추고, 해당 애플리케이션을 가장 자주 사용하는 부서와 연결해 보여줍니다. 표시된 비율은 전체 기업 AI 트래픽이 아닌, 특정 부서 및 애플리케이션 범위 내에서의 상대적 사용량을 반영합니다.

부서별 주요 결과

- **엔지니어링** 부서가 주도하는 엔터프라이즈 AI 활용은 이 범위 내 AI/ML 트랜잭션의 48.9%를 차지합니다. 특히 엔지니어링 팀은 AI를 일상적인 빌드 주기에 통합하여 작은 효율성 향상이라도 릴리스가 거듭될수록 빠르게 누적됩니다.
- **IT** 부서는 AI 의존도가 높은 부문으로, 전체 활동의 31.8%를 차지하며 그 뒤를 바짝 쫓았습니다. IT 분야에서 AI는 시스템 지원, 문제 해결, 내부 프로세스 자동화 등 운영 효율성 향상에 주로 활용됩니다.
- **마케팅** 부서는 이번 분석에서 기업 AI 활용률 3위(6.9%)를 차지했습니다. 마케팅 부문의 AI 도입은 콘텐츠 중심 및 디자인 중심 워크플로에 걸쳐 고르게 분포되어 있어 기술 부서에 비해 전반적인 트랜잭션 규모는 꾸준하지만 낮은 수준을 보이고 있습니다.

부서별 트랜잭션 비중

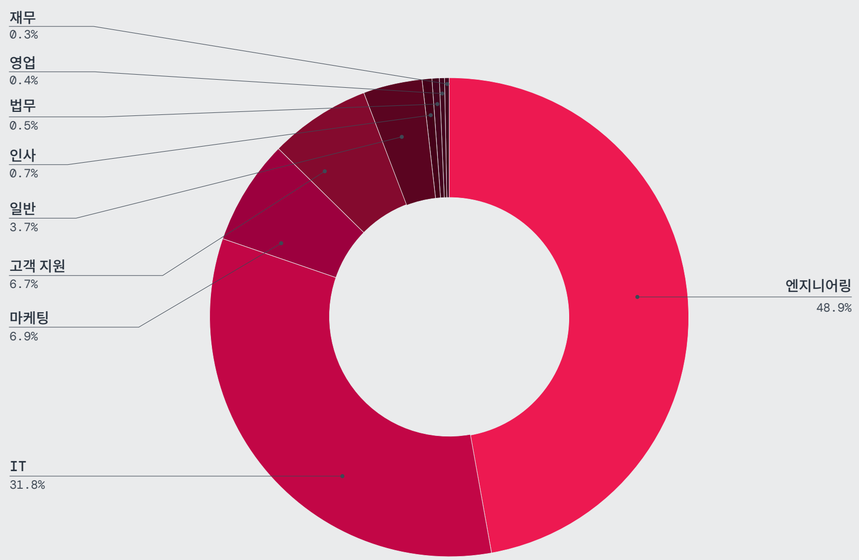


그림 5: 주요 기업 부서별 AI/ML 트랜잭션 비중

©2026 Zscaler, Inc. 모든 권리 보유.

차단된 트랜잭션

또한, 기업들은 2025년 기업용 AI에 대한 규제를 더욱 강화했습니다. 데이터 노출, 개인정보 보호 및 규정 준수에 대한 우려로 인해 전체 AI/ML 트랜잭션의 39.2%가 차단되었으며, 이는 AI 거버넌스를 일상적인 보안 운영의 표준 요소로 강화했습니다.

규제 조치로 가장 큰 영향을 받은 애플리케이션들은 기업에서 가장 널리 사용되는 AI 애플리케이션들이기도 했습니다. Grammarly는 차단된 활동에서 가장 큰 비중을 차지했는데, 1,712억 건의 트랜잭션이 차단되었으며 이는 차단된 모든 AI/ML 트랜잭션의 44.2%에 해당합니다. 광범위하게 사용되는 AI 애플리케이션 역시 감시 대상이었습니다. ChatGPT와 Microsoft Copilot은 빈번하게 차단되었으며, 각각 57억 건과 41억 건의 트랜잭션이 차단되었습니다. 이는 비정형 데이터에 대한 접근이 증가함에 따라 기업의 민감한 정보가 의도치 않게 공유될 위험이 커지고 있음을 보여줍니다.

Codeium과 Tabnine을 포함한 AI 코딩 어시스턴트 또한 독점 코드 및 개발 결과물의 노출을 제한하기 위해 흔히 차단되었습니다. QuillBot과 DeepL 같은 언어 및 콘텐츠 변환 도구들도 유사한 규제를 받았는데, 이는 외부 모델과의 콘텐츠 공유를 제한하려는 광범위한 노력의 일환이었습니다.

가장 많이 차단된 AI 애플리케이션

1	Grammarly
2	GitHub Copilot
3	ChatGPT
4	Microsoft Copilot
5	QuillBot
6	Codeium
7	DeepL
8	Tabnine
9	Poe
10	Perplexity



©2026 Zscaler, Inc. 모든 권리 보유.

AI 애플리케이션으로 전송된 데이터

트랜잭션 규모만으로는 기업들이 AI를 어떻게 활용하고 있는지 완전히 파악할 수 없습니다. 참고로 ThreatLabz는 기업 환경과 AI/ML 애플리케이션 간에 전송되는 데이터 양도 조사했습니다.

지난 한 해 동안 기업의 AI/ML 애플리케이션으로의 데이터 전송량은 지속적으로 증가하여 18,033 테라바이트(TB)에 달했으며, 이는 전년 대비 93% 증가한 수치입니다. 널리 사용되는 주요 애플리케이션 중 일부가 이러한 데이터 이동의 가장 큰 비중을 차지했습니다. Grammarly는 3,615TB의 데이터 전송량을 기록하며, 이 기준에 따르면 여전히 최고의 애플리케이션 자리를 유지했습니다. 그 뒤를 이어 ChatGPT(2021TB)가 가장 큰 규모를 차지했고, OpenAI(865TB), DeepL(625TB), Codeium(387TB)이 그 뒤를 이었습니다. 이들 애플리케이션은 일반적으로 고가치 기업 데이터를 처리하는 다양한 사용 사례를 포괄합니다.

AI가 일상 업무에 더욱 깊숙이 자리 잡으면서, 더 많은 기업 데이터가 AI를 통해 처리되고 있습니다. 트래픽과 데이터 양을 모두 분석하면 AI 사용이 증가하는 영역과 보안 및 감독이 가장 중요한 영역을 파악하는 데 도움이 됩니다.

데이터 전송 비율

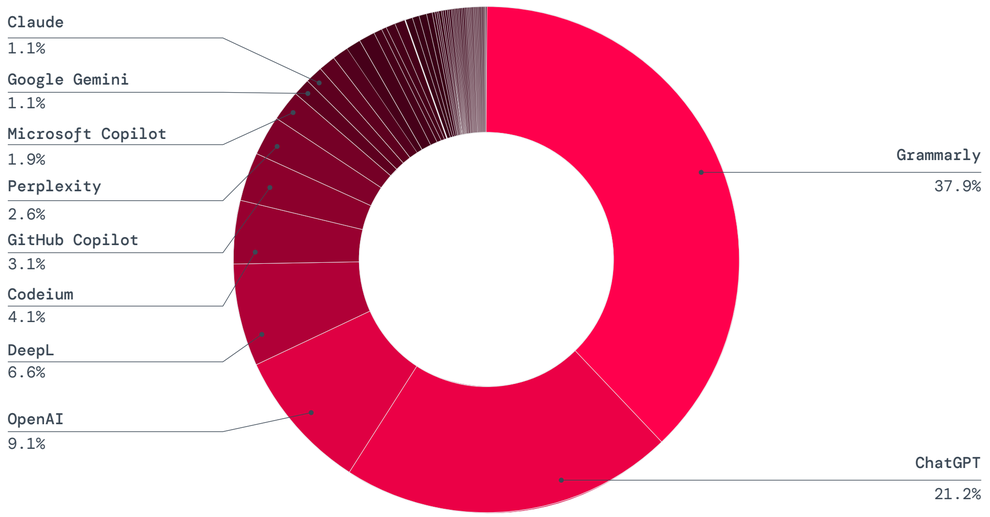
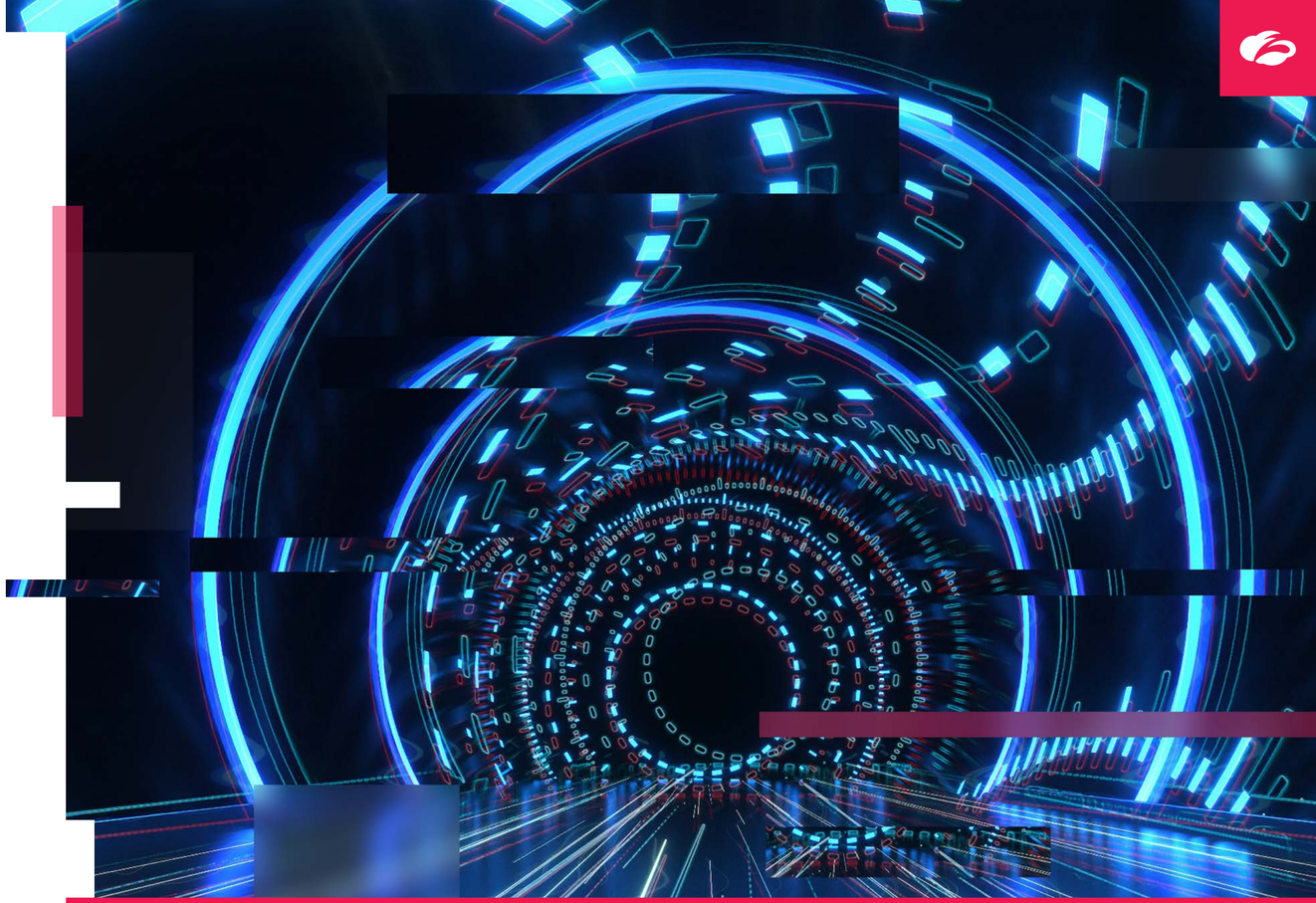


그림 6: 전체 데이터 전송량 대비 비율 기준 상위 AI/ML 애플리케이션



주요 결과

총 18,033TB의 데이터가 AI/ML 애플리케이션으로 전송되었으며,
이는 전년 대비 93% 증가한 수치입니다.

©2026 Zscaler, Inc. 모든 권리 보유.

AI 애플리케이션으로의 데이터 손실

AI는 아이디어 구상부터 결과물 도출까지 단 몇 분 만에 작업을 가속화할 수 있지만, 그만큼 위험한 단점도 있습니다. 민감한 데이터가 외부 모델과 **순식간에 공유될 수 있다는 것입니다.** 더욱이, 일반적인 SaaS 애플리케이션 및 서비스에 AI 기능이 내장되어 있는 경우 콘텐츠가 자동으로 전송되는 경우가 많아, 자신도 모르게 데이터가 노출될 가능성이 높아집니다.

외부 모델로의 데이터 손실을 방지하는 것은 올해 가장 중요한 보안 우선순위 중 하나가 되었습니다.

Zscaler 클라우드에서 AI 관련 DLP 정책 위반은 이러한 위험이 증가하고 있음을 보여주는 가장 명확한 신호 중 하나입니다. 이러한 위반은 재무 기록, 개인 식별 정보(PII), 소스 코드, 의료 데이터 및 기타 규제 대상 콘텐츠와 같은 민감한 정보가 AI 애플리케이션을 통해 조직 외부로 유출되려다 정책으로 인해 차단될 때 발생합니다. Zscaler의 AI 기반 DLP 솔루션이 없었다면 해당 데이터는 기업의 통제 범위를 벗어난 제3자 모델에 노출되었을 것입니다.

가장 위험한 AI 애플리케이션은 직원들이 생각 없이 사용하는 것들, 즉 글쓰기 어시스턴트, 코딩 어시스턴트 또는 협업 도구 모음에 포함된 AI 기능인 경우가 많습니다. 편리함이야말로 더 위험하게 만드는 요인입니다. 이러한 애플리케이션은 직원들이 보는 것과 동일한 민감한 콘텐츠를, 그것도 콘텐츠가 생성되는 바로 그 순간에 보게 됩니다.

위반 추세를 살펴보면 AI와의 상호 작용은 기업의 가장 민감한 데이터와 관련된 경우가 가장 많습니다.

DLP 정책 위반 사례가 가장 많은 AI/ML 애플리케이션

애플리케이션	DLP 위반 건수
ChatGPT	410,181,006
Codeium	242,263,311
GitHub Copilot	31,223,009
Claude	14,417,246
Wordtune	5,161,758
DeepL	2,037,613
QuillBot	1,960,391
Microsoft Copilot	1,858,952
Perplexity	1,235,129
Google Gemini	841,374

ChatGPT의 DLP 위반 사례가 전년 대비 **99.3% 증가**했습니다. ChatGPT에서 가장 흔하게 발생한 위반 사례는 이름 유출 및 국가 식별 정보 유출(고객 기록 또는 신원 정보 유출 가능성)이었습니다.

Codeium과 관련된 기업 DLP 위반 사례가 전년 대비 **100% 증가**하여 소스 코드 및 독점 로직 유출 위험이 높아졌음을 시사합니다.

©2026 Zscaler, Inc. 모든 권리 보유.



주요 AI DLP 위반 사례에서 눈에 띄는 점은 노출 범위가 전 세계적이라는 것입니다. 국가 식별 정보, 결제 데이터, 소스 코드 및 의료 정보 등 각각 엄격한 지역 규제의 적용을 받는 정보들이 AI 상호 작용에서 점점 더 많이 드러나고 있습니다.

AI DLP 정책 위반 사례 상위 10가지

1	이름 유출
2	사회보장번호(미국)
3	회사 등록 번호(일본)
4	국민건강보험 번호 (영국)
5	소스 코드
6	메디케어 번호(호주)
7	의료 제공자 식별 번호(미국)
8	사회보험 번호(캐나다)
9	의료 정보
10	신용카드 정보

이러한 DLP 추이는 실제 공격 환경에서 AI 시스템을 테스트할 때 관찰되는 실패 양상과 일치합니다. 즉, 정교한 공격보다는 일상적인 상호 작용을 통해 치명적인 오류가 발생하는 경우가 많습니다. 자세한 내용은아래의**"엔터프라이즈 AI 시스템에서 실제로 문제가 되는 것은 무엇인가"에서 확인하세요**.

GenAI 애플리케이션으로 인한 데이터 손실을 최소화하는 방법을 알아보려면 아래의 **기업들이 GenAI를 안전하게 도입하는 방법**을 읽어보세요.

©2026 Zscaler, Inc. 모든 권리 보유.

임베디드 AI의 부상

모든 기업 AI 활용 사례가 개별 생성형 AI 도구에 나타나는 것은 아닙니다. 점점 더 많은 경우, 이는 임베디드 AI를 통해 실현되고 있습니다. 요약, 추천 또는 자동화된 인사이트와 같이 특정 순간에만 AI가 작동하는, GenAI 앱으로 분류되지 않는 일상적인 애플리케이션에 내장된 기능들이 바로 그것입니다. 이러한 기능들은 사용자들이 이미 사용하고 있는 도구에 대한 자연스럽고 당연한 업그레이드처럼 느껴지는 경우가 많습니다. 그렇기 때문에 임베디드 AI가 독립형 AI 애플리케이션처럼 명확한 가시성이나 가드레일 없이 기업 데이터와 상호 작용한다는 사실을 간과하기 쉽습니다. 이는 눈에 띄지 않지만 AI 도입을 안전하게 보호하는 데 점점 더 중요해지는 요소입니다. 결과적으로 임베디드 AI는 기업 AI 위험 요소 중 가장 빠르게 성장하면서도 가장 눈에 띄지 않는 원인 중 하나입니다.

이러한 범주 변화가 중요한 이유는 임베디드 AI가 더 많은 맥락 정보를 활용하여 생산성을 향상하도록 설계되었기 때문입니다. 동일한 설계 원칙이 적용되더라도, 거버넌스와 통제가 그에 발맞춰 발전하지 못하면 위험 노출이 증가할 수 있습니다. 다음과 같은 위협 패턴은 흔히 기업 애플리케이션 전반에 걸쳐 임베디드 AI 기능과 연관되어 있습니다.

주요 관찰 사항

상속된 권한으로 인한 과도한 정보 공유

임베디드 AI는 일반적으로 기존 액세스 제어 및 콘텐츠 권한에 의존합니다. 조직에서 기본적으로 광범위한 액세스 권한을 부여했거나, 그룹 멤버십 정보가 오래되었거나, 협업 공간이 과도하게 공유되는 경우, 임베디드 AI가 의도치 않게 액세스 권한이 있지만 업무상 필요하지 않은 사용자에게 민감한 정보를 노출할 수 있습니다. 실제로 이는 오랫동안 누적된 권한 확산을 더욱 빠르고 눈에 띄는 데이터 노출로 이어지게 할 수 있습니다.

비즈니스 콘텐츠를 통한 간접적인 프롬프트 조작

임베디드 AI는 일반적인 작동 과정의 일부로 이메일, 티켓, 문서, 채팅 기록 및 첨부 파일과 같은 기업 콘텐츠를 읽는 경우가 많습니다. 이는 숨겨진 지침이나 악의적인 콘텐츠가 AI의 반응 방식, 우선순위 설정, 정보 제시 방식 등에 영향을 미칠 수 있는 위험을 초래합니다. AI 기능이 워크플로에 긴밀하게 통합되면 콘텐츠 자체가 조작을 위한 전달 경로가 될 수 있습니다.

모델 및 커넥터 공급망 노출

임베디드 AI 기능은 여러 구성 요소에 의존하는 경우가 많습니다. 여기에는 모델 제공업체, 기업 시스템에서 콘텐츠를 가져오는 검색 계층, SaaS 애플리케이션과 데이터 저장소를 통합하는 커넥터 등이 포함될 수 있습니다. 각 구성 요소는 새로운 신뢰 경계와 새로운 변화 요인을 도입할 수 있습니다. 기능이 발전함에 따라 업데이트, 구성 변경 또는 새롭게 활성화된 통합을 통해 리스크 프로필이 달라질 수 있습니다.

AI 기반 워크플로에서의 실행 및 자동화 위험

AI 기능이 요약 및 초안 작성 단계를 넘어 작업 실행 단계로 발전함에 따라 위험 요소도 확대됩니다. AI 기능이 특정 작업을 실행하거나, 변경 사항을 권장하거나, 코드를 생성하거나, 레코드를 채울 수 있는 경우, 오류나 조작된 출력물은 운영상의 문제로 이어질 수 있습니다. 직접적인 조치 실행이 없더라도 AI가 생성한 결과물은 감사가 어려운 방식으로 의사 결정 및 후속 워크플로에 영향을 미칠 수 있습니다.

실제 환경에서 발생하는 임베디드 AI 익스플로잇을 이용한 데이터 유출 용이성

Copilot 생태계에서 널리 알려진 두 가지 취약점 사례는 사용자 상호 작용이 적더라도 임베디드 AI로 인한 위험이 높아질 수 있음을 보여줍니다.

- **EchoLeak**은 Microsoft 365 Copilot에서 발견되는, 클릭 없이 프롬프트를 삽입하는 방식의 취약점으로, 일반적인 이메일 수신 패턴을 통해 데이터 유출을 가능하게 할 수 있다고 설명됩니다.
- **Reprompt**는 URL 매개변수를 통해 조작된 프롬프트를 사용하여 원치 않는 동작과 데이터 유출을 유발하는 것으로 보고된 바 있는 단일 클릭 공격입니다.

향후 더 많은 SaaS 제공업체가 AI를 기본 기능으로 제공하고 임베디드 기능을 확장함에 따라 기업은 AI가 암묵적으로 작동하는 애플리케이션 및 워크플로에 대한 AI 가시성, 거버넌스 및 데이터 보호를 확장해야 할 것입니다.

©2026 Zscaler, Inc. 모든 권리 보유.

산업별 AI/ML 사용 현황

2025년에는 모든 산업 분야에서 AI 도입이 가속화되었으며, Zscaler 클라우드에 포함된 모든 부문에서 AI/ML 활동이 전년 대비 증가세를 보인 것으로 나타났습니다. 하지만 도입 속도와 성숙도는 크게 다릅니다. 일부 분야에서는 이미 실질적인 효과를 보고 있습니다. 아직 자리를 잡아가고 있는 중인 분야도 있습니다.

금융 및 보험 기관은 2년 연속 AI/ML 트래픽에서 가장 큰 비중(23.3%)을 차지했습니다. 은행과 보험사는 데이터, 분석 및 자동화를 중심으로 운영되기 때문에 AI를 조기에 도입하는 데 유리한 위치에 있습니다. **제조 부문**은 전체 AI/ML 트랜잭션의 19.5%를 차지하며 2위를 유지했는데, 이는 AI 기반 자동화, 품질 관리, 공급망 최적화 등에 대한 투자 덕분입니다. **기술 및 커뮤니케이션**과 **교육 부문**은 전년 대비 가장 높은 증가율을 기록했으며, 자세한 내용은 아래에서 확인할 수 있습니다.

산업 분야별 AI 트랜잭션 비중

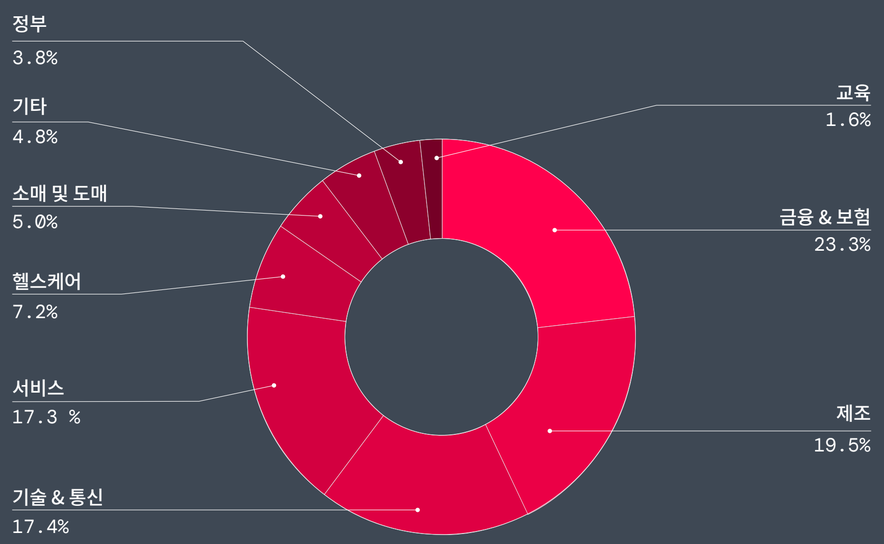


그림 7: AI 트랜잭션에서 가장 큰 비중을 차지하는 산업

업종별 차단된
AI 트랜잭션 비중

업종	AI 트랜잭션 차단 비율(%)
금융 및 보험	39.1%
제조	22.1%
서비스	13.5%
의료	8.5%
기술 및 커뮤니케이션	6.8%
정부	4.0%
기타	3.4%
도소매	2.0%
교육	0.6%

AI 활용은 고립된 환경에서 이루어지는 것이 아니라, 산업별 위험, 규정 준수 기대치, 그리고 보안 프로그램의 발전 정도에 영향을 받습니다.

차단된 AI/ML 트랜잭션 패턴은 산업계가 AI 도입과 위험 관리 사이에서 얼마나 각기 다른 균형을 유지하고 있는지를 보여줍니다. 금융 및 보험 부문은 AI 활동에서 가장 큰 비중을 차지했을 뿐만 아니라, 해당 트랜잭션의 약 40%를 차단하기도 했습니다. 높은 차단율은 단순한 신중함을 넘어, AI 사용에 대한 더욱 엄격한 통제가 요구되는 규제가 심한 환경에서 운영해야 하는 현실을 반영합니다.

AI 트랜잭션 규모 기준으로 두 번째로 활발한 산업인 제조 부문은 AI 트래픽의 약 22%를 차단했습니다. 이는 제조업체들이 AI를 광범위하게 도입하면서도 오용을 방지하고 데이터 유출을 막기 위해, 특히 IoT/OT 환경에서 상당한 수준의 감독을 유지해야 한다는 점에서 현실적인 절충점을 시사합니다.

©2026 Zscaler, Inc. 모든 권리 보유.

업계 집중 조명

금융 및 보험 부문은 2,300억 건의 트랜잭션을 처리하며 여전히 AI 도입률이 가장 높은 분야입니다.

금융 및 보험 부문은 Zscaler 클라우드의 AI/ML 분야에서 AI 활동을 주도하는 가장 큰 동력으로, 전체 기업 사용량의 거의 4분의 1을 차지했습니다. 이러한 사용량의 상당 부분은 일상적인 생산성 도구에서 비롯됩니다. Grammarly, ChatGPT, Microsoft Copilot은 2년 연속 은행 및 보험 회사에서 가장 많이 사용되는 AI 앱으로 선정되었습니다. 다양한 조직에서 이러한 도구를 사용하여 조사 요약, 규정 준수 문서 처리, 사기 탐지, 보험금 청구 처리 속도 향상, 보험 인수 지원 및 기타 필수 업무를 수행합니다. 이러한 추세는 업계 전반의 성장세에도 반영되었습니다. 모건 스탠리의 2025 AI 도입 기업 조사[1]에 따르면, 보험 업계의 AI 도입률은 상반기 기준 48%에서 71%로, 금융 서비스 기업의 경우 66%에서 73%로 급증했습니다.

이러한 가속화는 2025년의 여러 시장 요인에 의해 더욱 강화되었습니다. 은행은 비용 절감과 현대화 압력에 직면해 있으며, 이로 인해 다른 산업 분야보다 훨씬 빠르게 AI를 도입해야 하는 상황에 놓여 있습니다.

보험사들은 증가하는 보험금 청구액과 기후 변화로 인한 변동성에 직면하면서, 가격 책정 정확도를 높이고 대응 시간을 단축하기 위해 AI에 의존하고 있습니다.

동시에 해당 분야가 이러한 도구를 사용하는 방식에 있어 결코 방심하는 것은 아닙니다. 금융 및 보험 부문에서도 Zscaler 클라우드에서 AI/ML 트랜잭션의 39.1% 이상을 차단했는데, 이는 데이터 손실 위험, 규제 당국의 감시 강화, 그리고 민감한 금융 정보와의 모델 상호 작용을 엄격하게 관리해야 할 필요성에 대한 민감도가 높아졌음을 나타냅니다. 빠르게 움직이고 있지만, 언제든 제동을 걸 수 있도록 대비하고 있습니다.

금융 및 보험 산업은 2026년에도 야심찬 AI 혁신의 모습을 계속해서 정의해 나갈 것입니다.



1. 비즈니스 인사이더, **AI 열풍이 실제 수익으로 이어지는 3개 시장 분야(모건 스탠리)**, 2025년 7월 24일.

©2026 Zscaler, Inc. 모든 권리 보유.



업계 집중 조명

기술 분야에서 기업용 AI 사용이 가장 빠른 성장세를 보였습니다: 전년 대비 +202%

기술 부문은 2025년에 AI/ML 트랜잭션이 전년 대비 가장 높은 증가율(202.3%)을 기록했습니다. Zscaler 클라우드에서 다른 모든 산업 분야를 앞지르고 있습니다. 기술 분야는 생성형 AI를 일찍부터 열정적으로 도입하는 등 항상 AI를 적극적으로 활용해 왔지만, 올해의 급증세는 소프트웨어 회사, 클라우드 제공업체, 디지털 플랫폼 및 엔지니어링 팀이 제품과 내부 워크플로에 AI를 얼마나 집중적으로 통합하고 있는지를 반영합니다.

주요 생산성 지원 도구는 기술 조직 전반에서 널리 사용되며 코드 생성 및 기술 문서 작성부터 마케팅 콘텐츠 제작에 이르기까지 모든 분야에 활용됩니다.

따라서 당사의 분석 결과, Grammarly, Codeium, ChatGPT, Perplexity 등이 기술 부문 트래픽을 유발하는 주요 AI 앱으로 나타났습니다.

이처럼 빠른 성장세에도 불구하고, 많은 기술 조직에게 있어 AI는 가시성 및 정책 시행 측면에서 여러 허점을 드러내고 있습니다. 이에 대응하여, 기업들은 감독에 더 많은 투자를 하고 있으며, AI 트랜잭션의 약 7%를 차단하고 있습니다. 이는 전체적으로는 여전히 상대적으로 작은 비중이지만, 다른 많은 산업에 비해 상당히 높은 수치이며, 안전한 배포를 지원하기 위한 통제를 개선하고 있습니다.

©2026 Zscaler, Inc. 모든 권리 보유.

업계 집중 조명

교육 분야에서 AI 도입이 조용하지만 폭발적인 성장을 보이고 있습니다: 전년 대비 +184% 증가

2025년 Zscaler 클라우드에서 교육 부문이 차지하는 AI/ML 트랜잭션은 전체 트랜잭션 규모에서 작은 비중을 차지했지만, 그 성장률은 전혀 다른 양상을 보여줍니다. 교육 부문은 한 해 동안 거의 160억 건의 AI/ML 트랜잭션을 생성했으며, 전년 대비 184.4%의 증가율을 기록하며 AI/ML 활동 증가율에서 두 번째로 높은 수치를 보였습니다. 이는 교육 부문이 모든 산업 분야 중 AI 도입 속도가 가장 빠른 분야 중 하나임을 의미합니다.

이러한 증가는 학습 및 교실 수업 과정에서 생성형 AI 사용이 확대되는 추세와 밀접하게 연관되어 있습니다. ChatGPT와 Microsoft Copilot 같은 애플리케이션은 학생과 교직원이 글쓰기 지원, 콘텐츠 제작 및 수업 계획을 세우는 데 널리 사용됩니다. 관리자들은 또한 AI를 활용하여 공지사항 작성부터 학생 서비스 개선에 이르기까지 일상적인 업무를 간소화하고 있으며, 이는 트랜잭션 규모의 꾸준한 증가에 기여하는 것으로 보입니다.

특히, 이러한 급증은 거의 마찰 없이 이루어졌습니다. 교육 분야에서 AI/ML 관련 트랜잭션 중 차단된 비율은 1% 미만으로, 이는 대부분의 사용이 명시적으로 허용되었거나 관리 체계 및 가드레일이 아직 마련 중인 환경에서 이루어지고 있음을 시사합니다. 따라서 교육 부문은 다른 산업 분야에 비해 AI/ML 활용에 신중한 태도를 보이는 것은 당연합니다. 학교와 대학은 데이터 개인정보 보호 및 학문적 진실성에 대한 우려를 해결해야 합니다. 이러한 요인들 때문에 AI 도입이 빠르게 증가하고 있음에도 불구하고 전반적인 AI 사용률은 다른 산업에 비해 낮은 수준을 유지해 온 것으로 보입니다.

하지만 단 1년 만에 거의 3배에 달하는 성장은 향후 더욱 체계적이고 책임감 있는 AI 계획 및 통합을 위한 발판을 마련해 줍니다.



©2026 Zscaler, Inc. 모든 권리 보유.

국가별 AI/ML 활용 현황

2025년에도 AI/ML 활동의 지리적 분포는 전반적으로 큰 변화 없이 유지되었습니다. AI는 엔터프라이즈 AI 개발 및 배포의 중심지인 **미국**에서 확고히 자리 잡았으며, 미국은 여전히 AI/ML 트래픽 양에서 가장 큰 비중을 차지하고 있지만, 여러 해외 시장에서도 AI 사용이 크게 증가했습니다.

미국은 절대적인 사용량(2,189억 건의 AI/ML 트랜잭션, 전 세계 활동의 37.6%)에서 여전히 선두를 유지했지만, 다른 국가들의 AI 도입은 전년 대비 더 빠른 속도로 확대되었습니다. 이러한 세계적인 가속화는 **인도**에서 가장 두드러지게 나타나는데, 인도는 기업 AI 활동의 두 번째로 큰 공급처로서 823억 건의 트랜잭션을 기록하며 전년 대비 309.9% 증가했습니다. 인도의 성장은 2025년에도 지속된 디지털 전환을 위한 정부의 지속적인 노력과 AI 인프라 및 기술 개발에 대한 대규모 공공 및 민간 투자와 맥락을 같이합니다. AI를 활용하는 인력의 증가와 AI 서비스의 빠르고 확장 가능한 배포를 가능하게 하는 클라우드 우선 아키텍처가 결합되어 인도의 AI 도입이 전년 대비 크게 증가한 것으로 보입니다.

상위 두 국가 외에도 여러 선진 시장에서 기업 주도의 꾸준한 AI 확장 추세가 이어졌습니다. **캐나다**는 AI 컴퓨팅 역량에 대한 연방 투자와 특히 규제 산업 전반에 걸친 기업 도입 가속화 프로그램에 힘입어 272억 건의 트랜잭션(전년 대비 229.9% 증가)을 기록했습니다. **영국**과 **일본**은 각각 117.5%와 122.8%의 증가율을 보이며 상위 5위권에 들었습니다.

이처럼 광범위한 지리적 분포는 AI가 기업의 표준 역량으로 자리 잡았음을 반영합니다. 보안 팀은 이처럼 더욱 분산된 사용 환경을 고려하고, 여러 지역에 걸쳐 일관된 감독을 보장해야 합니다.

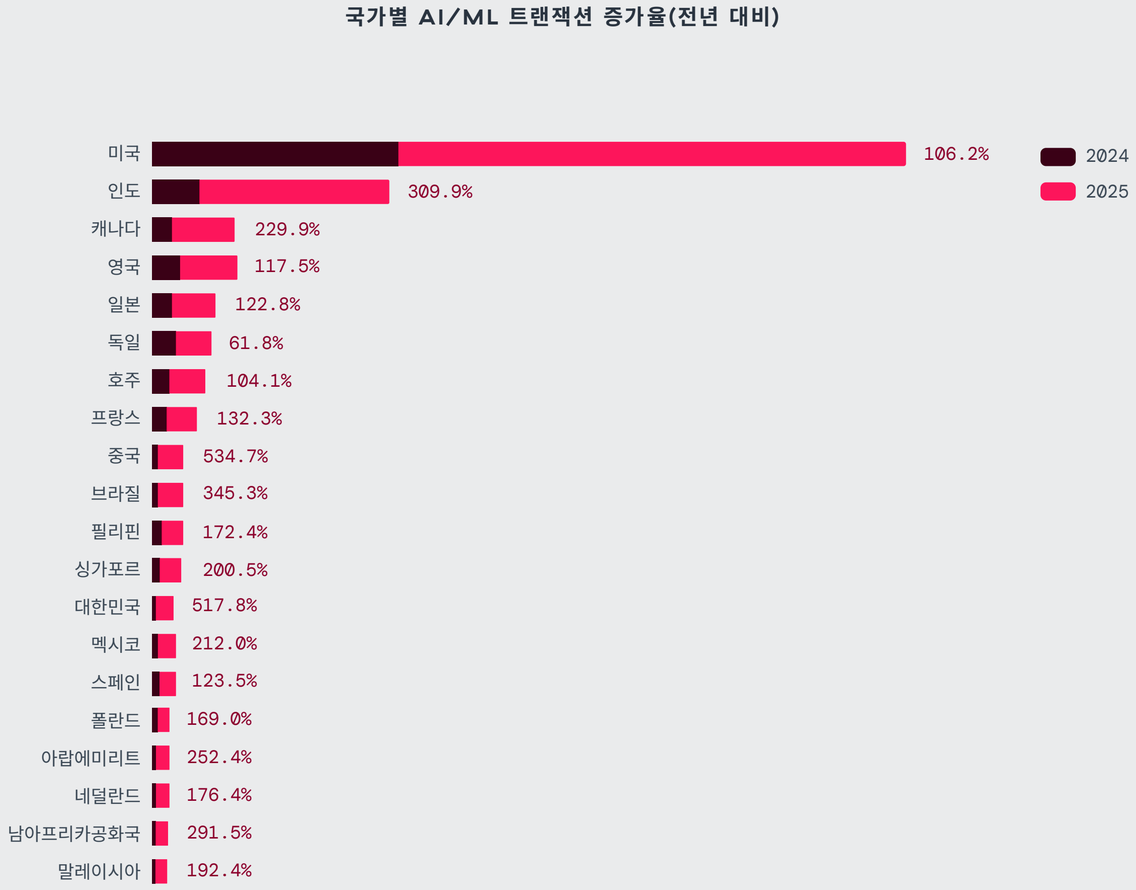


그림 8: 국가별 AI/ML 트랜잭션 전년 대비 증가율 (트랜잭션 규모 기준 상위 20개국)

©2026 Zscaler, Inc. 모든 권리 보유.

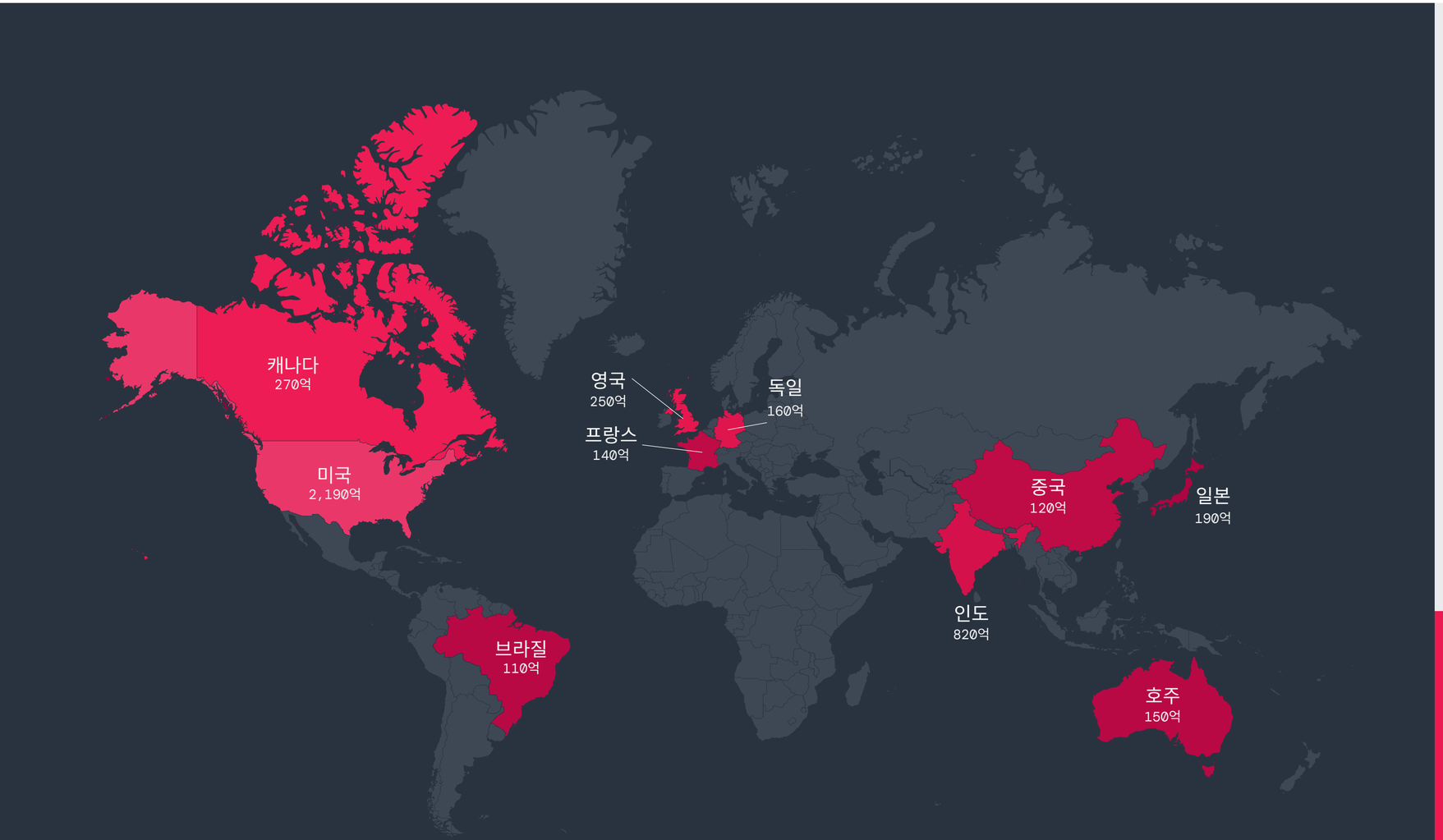


그림 9: AI/ML 트랜잭션 규모 기준 상위 10개국을 보여주는 지도
(오른쪽 표: 2025년 6월~12월 기준 비중 및 총 규모)

국가	비중(%)	AI/ML 트랜잭션
미국	37.6%	2,190억
인도	14.1%	820억
캐나다	4.7%	270억
영국	4.3%	250억
일본	3.2%	190억
독일	2.7%	160억
호주	2.6%	150억
프랑스	2.4%	140억
중국	2.0%	120억
브라질	1.8%	110억

©2026 Zscaler, Inc. 모든 권리 보유.

EMEA 인사이트

EMEA 지역의 AI/ML 관련 활동은 소수의 선진 유럽 시장에 집중되어 있는 것으로 나타났습니다. 영국, 독일, 프랑스, 스페인이 지역 트랜잭션의 거의 절반을 차지했습니다. 영국은 전 세계 AI 활동에서 차지하는 비중은 작지만, EMEA 지역에서는 2025년 6월부터 12월까지 AI/ML 트래픽의 20.3%를 차지하며 지역 내 높은 비중을 기록했습니다.

독일은 제조업 분야에서 AI 통합이 지속적으로 진행됨에 따라 55억 건 이상의 AI/ML 트랜잭션이 발생하면서 EMEA 지역 트랜잭션의 12.5%를 차지했습니다. 그 뒤를 이어 프랑스가 지역 활동의 11%를 차지했는데, 이는 AI 투자 약속을 포함한 프랑스 2030 전략과 국제 AI 액션 서밋 개최와 같은 정부 이니셔티브에 힘입은 결과입니다.

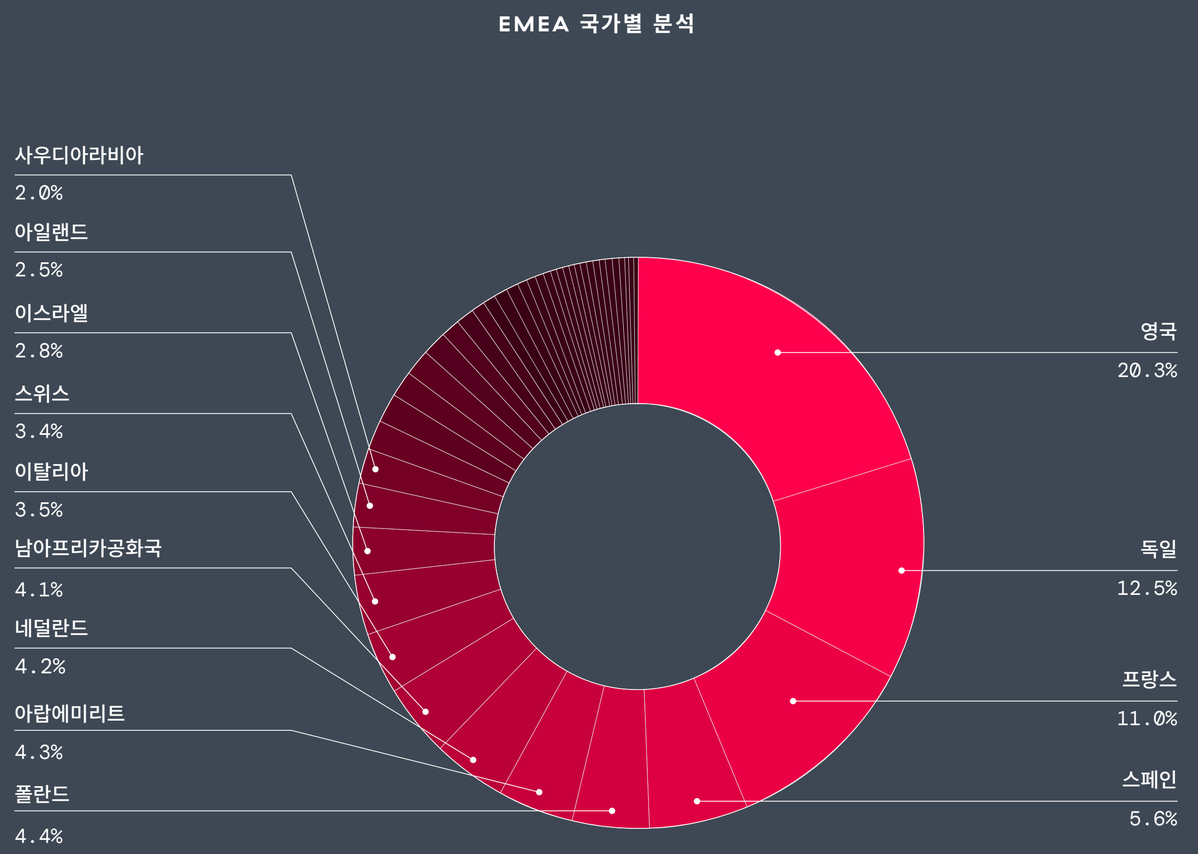


그림 10: EMEA 지역 국가별 AI 트랜잭션 비중

©2026 Zscaler, Inc. 모든 권리 보유.

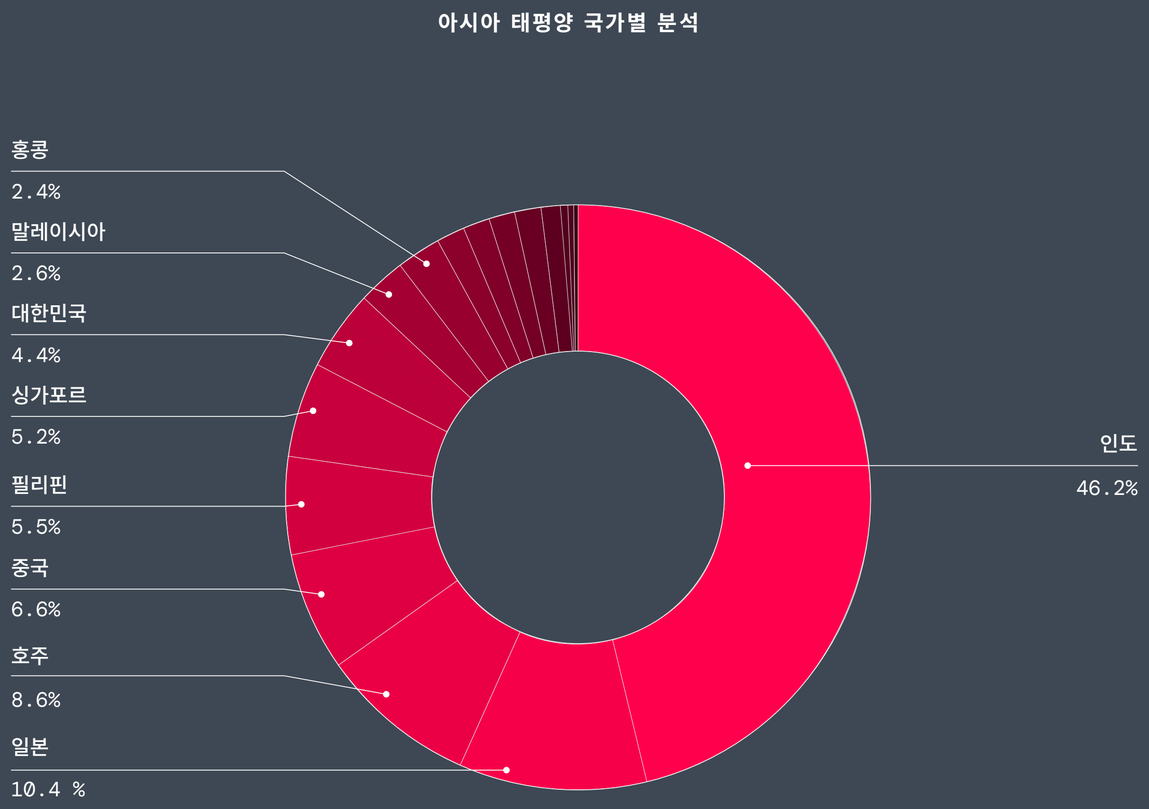


그림 11: 아시아 태평양 지역 국가별 AI 트랜잭션 비중

지역별 현황

APAC 인사이트

아시아 태평양(APAC) 지역의 AI/ML 활용은 단일 고성장 시장과 여러 선진국 간의 뚜렷한 불균형에 의해 형성되었습니다. 인도, 일본, 호주는 지역 AI/ML 트랜잭션의 대부분을 차지했으며, 특히 인도는 전체 활동의 거의 절반인 46.2%를 주도했는데, 이는 주로 기술 및 커뮤니케이션 부문(310억 건의 트랜잭션)이 견인했습니다.

일본은 진화하는 국가 AI 정책을 배경으로 아시아 태평양 지역 트랜잭션의 10.4%를 차지하며 그 뒤를 이었습니다. 일본 정부는 기업과 산업계의 AI 도입을 장려하기 위해 체계적인 지침을 제공하는 국가 AI 진흥법을 제정했습니다. 호주는 지역 활동의 8.6%를 차지했으며, 책임감 있고 안전한 AI 배포에 대한 국가적 차원의 지속적인 노력이 이어지고 있습니다.

©2026 Zscaler, Inc. 모든 권리 보유.

엔터프라이즈 AI의 위험 및 위협환경

당사의 연구 결과에서 입증된 바와 같이, AI는 퍼블릭 GenAI 도구부터 내부 LLM 및 AI 기반 SaaS 제품군에 이르기까지 기업의 모든 계층에 스며들어 있습니다. 조직은 사용량이 증가함에 따라 더욱 광범위하고 복잡해지는 공격 표면을 관리해야 합니다. 가장 중요한 위험은 다음과 같은 범주에 속합니다.

데이터 노출 및 민감 정보 유출

AI 시스템은 소스 코드, 고객 기록, 재무 정보, 법률 문서 등 기업에서 가장 민감한 데이터에 접근하지만, 명확한 가드레일이 없는 경우가 많습니다. 이러한 노출은 일반적으로 ChatGPT, Grok, DeepSeek과 같은 공개 도구에서 은밀하게 사용되는 AI 기능과 Microsoft Copilot처럼 권한이 과도하게 부여된 SaaS AI가 잘못된 구성이나 부정확한 레이블로 인해 데이터를 노출하는 경우에 발생합니다. 이와 동시에, 통제되지 않은 검색 증강 생성(RAG) 파이프라인은 규제 대상 데이터를 조용히 프라이빗 모델로 가져올 수 있습니다. 민감한 정보가 AI 시스템에 전송되면, 해당 정보는 즉각적인 조작이나 모델 동작을 통해 보존, 재사용되거나 심지어 노출될 수도 있습니다. 이는 일상적인 AI 사용을 실제 데이터 위험으로 만듭니다.

AI 사용 현황 및 사용자 프롬프트에 대한 가시성 부족

많은 조직들이 AI가 실제로 일상적으로 어떻게 활용되고 있는지에 대한 기본적인 질문에 답하는 데 여전히 어려움을 겪고 있습니다. 보안 팀은 직원들이 어떤 AI 도구를 사용하는지, 어떤 질문을 제출하는지, 그리고 민감한 데이터가 위험에 처해 있는지 여부를 명확하게 파악하지 못하는 경우가 많습니다. 또한 어떤 팀이 핵심 워크플로에 GenAI을 사용하는지 항상 명확하게 알 수 있는 것은 아닙니다. 프롬프트를 검토해 보면, 프롬프트 인젝션 시도, 조작 패턴 또는 최소한의 노력으로 안전장치를 우회하는 규정 비준수 행위가 종종 드러납니다. 하지만 대부분의 조직은 이러한 활동을 실시간으로 관찰할 수 있는 도구를 갖추고 있지 않습니다. 결과적으로 AI 거버넌스는 문제가 이미 발생한 후에야 작동하는 사후 대응적인 경향이 있습니다.

데이터 품질, 환각, 모델 조작

AI가 일상적인 비즈니스 운영에 통합됨에 따라, AI 출력물에서 발생하는 오류는 실질적인 결과로 이어집니다. 2025년에는 기업들이 AI가 생성한 지침이 권위 있어 보이지만 실제로는 잘못된 것으로 판명되는 환각을 바로잡아야 할 상황에 직면했습니다. 특히 규정 준수에 중점을 둔 팀에서 RAG 기반 시스템은 편향되거나 품질이 낮은 입력으로 인해 왜곡된 결과를 도출하는 경우가 많습니다. **레드티밍 훈련과 실제 테스트**를 통해 공격자가 AI 시스템이 입력받는 소스에 조작된 콘텐츠를 삽입하거나 미묘한 프롬프트 변형을 통해 근거 및 정확도 취약점을 악용하는 방식으로 검색 파이프라인을 오염시킬 수 있음이 밝혀졌습니다. 환각, 암묵적 변형, 근거 부족은 AI 출력물에 대한 신뢰를 지속적으로 저해합니다. 이러한 오류가 방치될 경우, 결함이 있는 출력물은 의사 결정에 직접적인 영향을 미치고 위험을 증폭시킬 수 있습니다.

매핑되지 않고 보안이 확보되지 않은 프라이빗 AI 모델

현재 기업들은 Salesforce, ServiceNow, Atlassian과 같은 플랫폼에 내장된 관리형 및 비관리형 모델과 AI 기능을 혼합하여 사용하고 있습니다.

하지만 많은 조직들이 여전히 다음과 같은 점을 놓치고 있습니다.

- 모델 및 서비스의 전체 인벤토리
- 각 모델이 어떤 데이터를 다루는지에 대한 이해
- 모델 보안, 패치 수준 또는 취약점 상태 검증
- AI 워크플로에 사용되는 소스 코드 리포지토리에 대한 거버넌스

이러한 매핑 부족은 특히 프라이빗 모델이 공개 시스템에서 관찰되는 것과 동일한 프롬프트 인젝션, RAG 포이즈닝 및 데이터 유출 취약점을 물려받을 때 매우 위험해집니다. 모델과 데이터 흐름을 알 수 없는 경우, 조직은 정책을 시행하거나 위험을 의미 있게 평가할 수 없습니다.

개인정보 보호, 규정 준수, 제공업체별 편차

AI 제공업체들은 기업 데이터 처리에 있어 각기 다른 접근 방식을 취합니다. 프롬프트는 저장되거나 교육용으로 재사용되거나 기록될 수 있는데, 이러한 방식이 항상 명확한 것은 아닙니다. 액세스 제어 방식과 모델 계보는 벤더마다 크게 다릅니다. 이러한 불일치는 GDPR, HIPAA, PCI DSS와 같은 여러 규정 체계에서 규정 준수 문제를 야기합니다. SaaS 애플리케이션이 기존 승인 절차를 우회하는 AI 기능을 기본적으로 제공하면서 위험이 가중되고, 기업 정책이 규제 당국의 기대치와 어긋나게 됩니다.

©2026 Zscaler, Inc. 모든 권리 보유.

실제 위협과 취약점

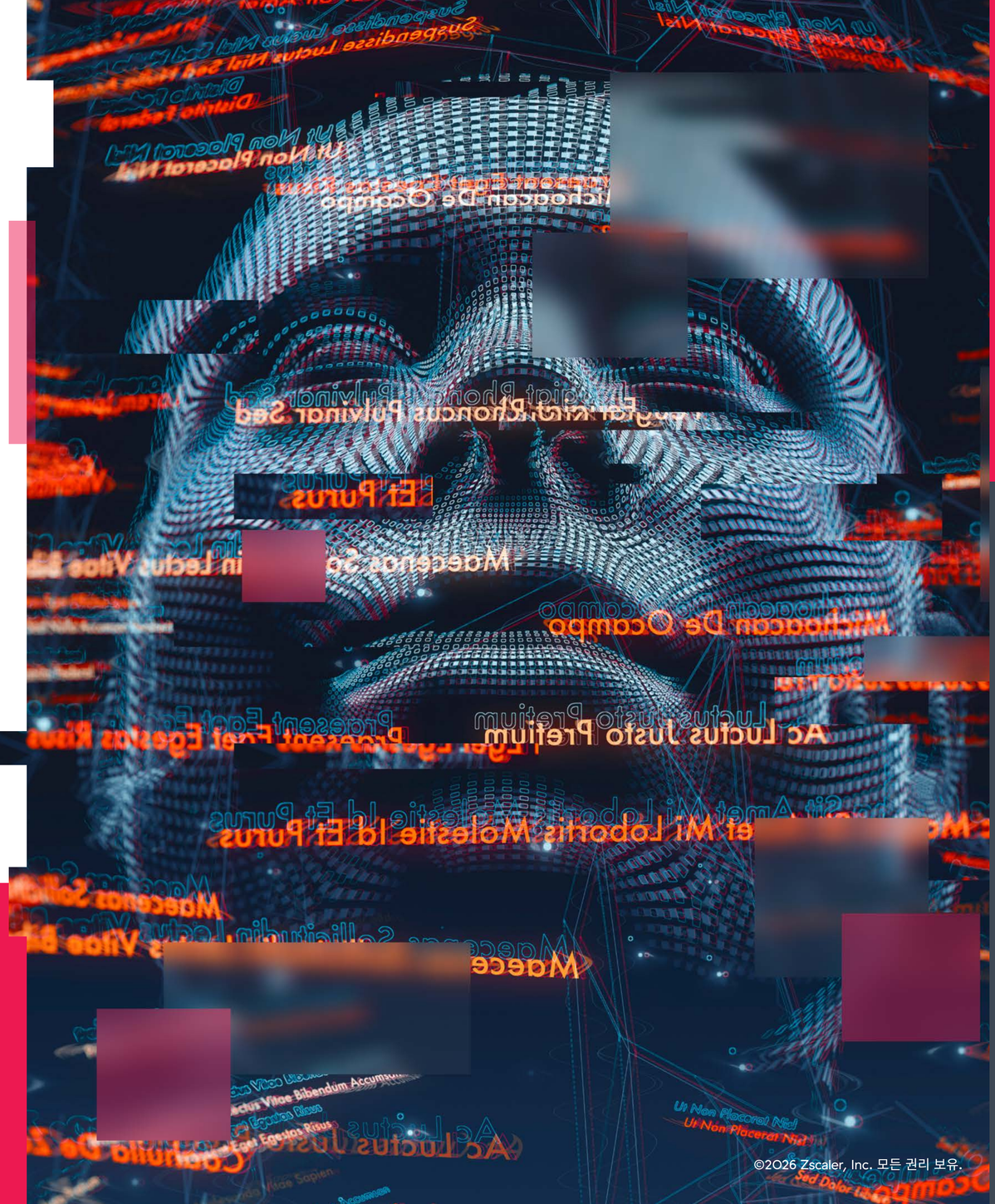
기업의 AI 도입과 관련된 핵심 위험 요소들은 2025년에도 실제 사례를 통해 계속 나타났습니다. 데이터 노출, AI 사용에 대한 가시성 부족, 환각 등과 같은 우려 사항들이 기업 환경 전반에 걸쳐 실질적인 보안 위협 및 운영 취약점으로 대두되었습니다. 실제 사례와 테스트 결과는 이러한 위험이 AI 시스템의 배포 방식, 데이터 연결 방식, 그리고 일상적인 워크플로 내에서의 신뢰도에서 비롯된다는 것을 보여주었습니다.

AI 기반 사회공학, 인공지능 애플리케이션 및 어시스턴트를 통한 데이터 유출, 그리고 에이전트형 및 반자율형 AI 시스템의 초기 오용에서 나타나는 가장 중요한 근본적인 위험 요소들이 있습니다.

AI 기반 사회공학 공격이 급증하면서 공격자들은 생성형 AI를 활용하여 더욱 그럴듯한 사칭을 시도하고 있습니다. 딥페이크 음성 및 비디오 피싱("비싱")은 2025년에 심각한 문제로 대두되었습니다. 미국 당국의 경고를 포함한 여러 권고 사항에서, 공격자들은 AI로 생성된 음성과 메시지를 통해 공무원을 사칭하는 것으로 나타났습니다.[2] 공격자들은 AI를 사용하여 특정 역할과 의사 결정 과정에 맞춘 그럴듯한 딥페이크 비디오와 음성을 제작하고 있습니다.

작년에는 **AI 에이전트를 활용한 사이버 스파이 활동**에 대한 신뢰할 만한 최초의 보고서가 나왔습니다. 중국 정부의 지원을 받는 한 그룹은 정찰, 취약점 검증, 자격 증명 수집, 측면 이동, 데이터 유출 등 침입 과정의 80~90%를 에이전트형 AI로 자동화했습니다. 인간 운영자는 상황 악화 시에만 개입했습니다. 이 사건은 자율 에이전트가 기존 공격 방식을 기계적인 속도로 실행할 수 있음을 보여주었으며, 이는 방어자가 위협을 탐지하고 대응하는 방식을 근본적으로 바꿔놓았습니다.

공격자들은 AI 시스템을 직접적으로 악용하는 것을 넘어, 자신들의 개발 워크플로에 AI를 통합하기 시작했습니다. ThreatLabz가 관찰한 여러 캠페인에서 악성 소프트웨어는 AI 기반 코드 생성과 일관된 특징을 보였으며, 이는 GenAI가 공격에 점점 더 많이 사용되고 있음을 시사합니다.



다음 사례 연구는 GenAI 기반 기만 및 공격 실행부터 기업 AI 시스템이 실제 적대적 환경에서 어떻게 작동하는지 보여주는 레드팀 테스트에 이르기까지 AI 위험을 실증적으로 보여줍니다.

[2] 사이버보안 다이브, FBI, 미국 고위 관리들이 문자 메시지와 AI 기반 음성 복제를 이용해 사칭당하고 있음을 경고, 2025년 5월 16일.

©2026 Zscaler, Inc. 모든 권리 보유.

사례 연구

GenAI로 강화된 악성 소프트웨어 및 북한 연계 캠페인에서의 사회 공학

이 사례 연구는 GenAI가 공격자의 목표나 기술을 근본적으로 바꾸지 않고도 공격 작전을 강화할 수 있도록 지원하는 방식을 보여줍니다.

북한과 연계된 활동 및 더 광범위한 북한 IT 노동자 계획과 연관된 **"전염성 인터뷰" 캠페인**에서 ThreatLabz는 위협 행위자들이 인공지능(GenAI)을 무기화하여 사회공학적 기법을 산업화하는 것을 관찰했습니다. 즉, 그럴듯한 가짜 인물을 만들고 운영하는 동시에 AI 기반 코딩을 악성코드 개발에 활용하는 것입니다. AI는 공격자가 시스템에 침입하는 방식과 그 이후에 수행하는 작업을 정상적인 활동과 구별하기 어렵게 만들어 탐지 및 대응의 난이도를 높이고 있습니다.

리소스 개발 및 사회 공학(면접 사칭)

이 캠페인은 GenAI 기술을 사용하여 디지털 신원을 구축하는 것부터 시작하여, 종합적인 학습 가이드를 제작하고, 전문적이면서도 추적 불가능한 프로필 사진을 생성하고, 딥페이크 및 음성 변조 도구를 사용하여 원격 인터뷰 중에 신원을 숨깁니다. 이러한 기만 행위는 검증 절차를 우회하고 중요한 기술 직책을 확보하기 위해 고안되었습니다.



©2026 Zscaler, Inc. 모든 권리 보유.

다음 결과는 면접 준비 단계에서 AI에 얼마나 크게 의존하는지를 분명히 보여줍니다.

면접 완벽 대비를 위한 AI 생성 학습 가이드

위협 행위자들은 GenAI를 사용하여 기술 면접에 대비하기 위한 상세한 교육용 지침서를 제작합니다.

예시: 한 권의 "학습 가이드"는 70페이지가 넘으며 백엔드 엔지니어링 및 웹3 개발과 같은 분야의 복잡한 질문들을 다룹니다.

AI의 주요 지표:

- 가이드의 답변에는 "물론이죠!"와 같은 특징적인 문구가 포함되어 있습니다(그림 12).
- 마크다운 서식의 잔여 요소가 남아 있어 AI 모델이 생성한 출력물에서 직접 복사하여 붙여넣기 작업을 수행했음을 강력하게 시사합니다(그림 13).

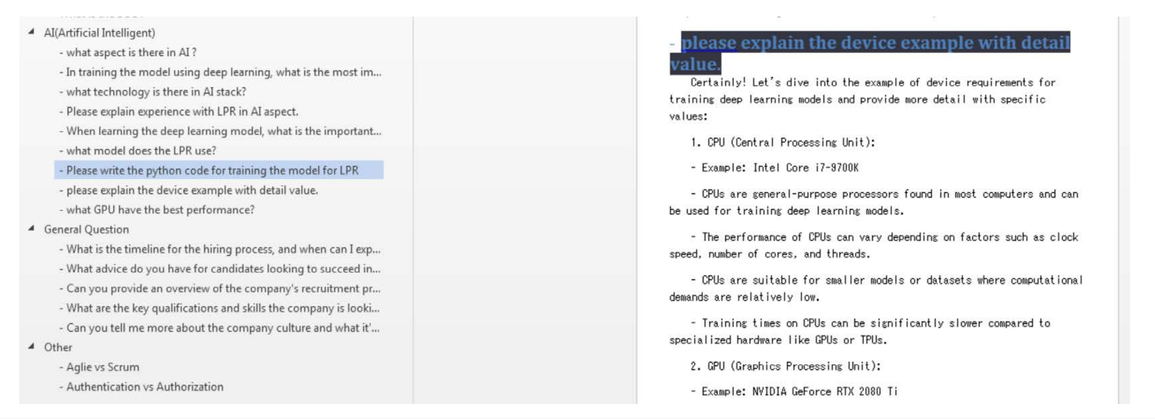


그림 12: GenAI의 전형적인 표현 방식을 보여주는 지침서 Q&A 답변

Project Requirements:

1. **Product Catalog**: Implement a product catalog where administrators can add, edit, and manage products. Users should be able to browse products with various filtering options.

2. **User Authentication and Roles**: Create a user authentication system with multiple user roles (admin, customer). Administrators should have access to the admin dashboard for managing products and orders.

3. **Shopping Cart**: Develop a shopping cart that allows users to add products, update quantities, and proceed to checkout.

4. **Order Management**: Implement order processing, allowing customers to place orders, view order history, and receive order confirmation emails.

5. **Payment Integration**: Integrate a payment gateway to handle online payments securely.

6. **Search and Filtering**: Implement search functionality to allow users to search for products based on keywords and apply filtering based on categories, price range, etc.

7. **Responsive Design**: Design the application with a responsive user interface to ensure a seamless experience across different devices.

8. **Error Handling and Validation**: Ensure proper error handling and validation throughout the application to deliver a smooth user experience.

그림 13: GenAI 출력물에서 직접 복사했을 가능성을 나타내는 마크다운 서식

©2026 Zscaler, Inc. 모든 권리 보유.

AI 기반 이미지 편집을 이용한 신원 위조

북한의 IT 종사자들은 AI 이미지 생성 및 편집 기술을 사용하여 이력서, 홍보 웹페이지, GitHub 프로필 등에 사용할 가짜 디지털 신분을 만듭니다.

예시: AI가 생성한 이미지에는 더욱 전문적으로 보이거나 서구적인 스타일을 반영한 보정된 인물 사진이 포함됩니다. 작업 환경을 숨기기 위해 배경을 제거하거나 수정하는 경우가 많습니다.

AI의 주요 지표:

- 이미지들은 지나치게 전문적으로 편집되어 부자연스럽게 보이는 특징들을 보여줍니다(그림 14).
- 이미지의 메타데이터 또는 시각적 아티팩트에서 AI를 실행해 배경을 제거한 증거가 감지되었습니다(그림 15).



그림 14: 원본 이미지(왼쪽)와 AI로 편집된 이미지(오른쪽)

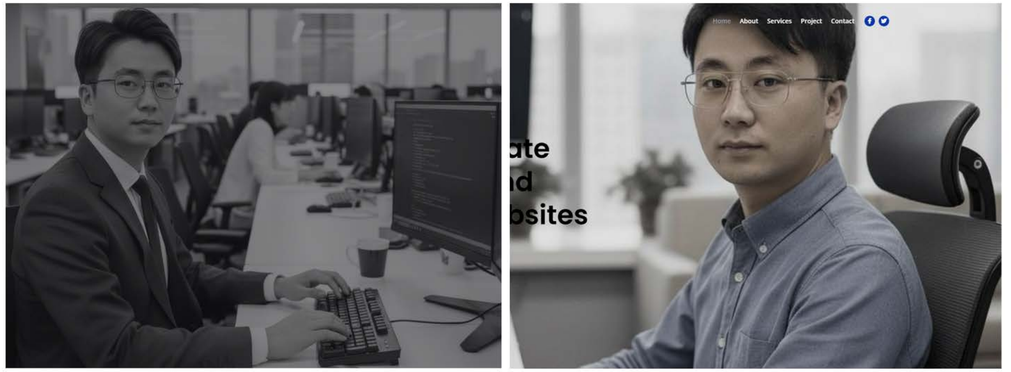


그림 15: AI로 개선된 프로필 사진

©2026 Zscaler, Inc. 모든 권리 보유.

초기 접근: 트로이목마화된 소프트웨어 유포

일단 접근 권한이 확보되면, 공격자들은 피싱 및 소셜 엔지니어링 기법을 사용하여 암호화폐 엔지니어와 같은 피해자들을 표적으로 삼습니다. 피해자들은 악성 도구가 합법적인 개발 리소스로 위장한 변조된 Node Package Manager(NPM) 패키지와 같은 트로이목마화된 소프트웨어를 다운로드하도록 유도하여 초기 침투 경로를 확보합니다.

결정적으로, 모니터링 과정에서 이러한 악성 스크립트 중 상당수가 인공지능으로 생성되었다는 뚜렷한 징후를 보였습니다. 그림 16에서 볼 수 있듯이, 해당 코드는 꼼꼼한 들여쓰기, 잘 구성된 오류 메시지, 그리고 눈에 띄게 사용된 이모티콘을 특징으로 하는데, 이는 소스 코드 생성에 사용된 특정 GenAI 엔진의 특징이라고 할 수 있습니다.

```
if [ ! -f package.json ]; then
    echo "[ERROR] package.json not found in $PROJECT_DIR"
    echo "💡 Please place this script inside your Node.js project folder."
    exit 1
fi

echo "Installing project dependencies..."
npm install

# === OPTIONAL: Auto-start on macOS login ===
PLIST=~/Library/LaunchAgents/com.local.drvierUpdate.plist
mkdir -p ~/Library/LaunchAgents

cat > "$PLIST" <<EOL
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE plist PUBLIC "-//Apple//DTD PLIST 1.0//EN"
    "http://www.apple.com/DTDs/PropertyList-1.0.dtd">
<plist version="1.0">
<dict>
    <key>Label</key>
    <string>com.local.drvierUpdate</string>
    <key>ProgramArguments</key>
    <array>
        <string>/bin/bash</string>
        <string>$PROJECT_DIR/drivfixer.sh</string>
    </array>
    <key>RunAtLoad</key>
    <true/>
</dict>
</plist>
EOL

chmod 644 "$PLIST"
launchctl load -w "$PLIST"

echo "✅ Setup complete. Your Node.js app will auto-start on login."
```

그림 16: GenAI를 이용한 개발을 제안하는 영구적인 자바스크립트 악성코드를 심는 Bash 스크립트

단계별 페이로드 실행

배포 후 악성 소프트웨어는 미리 준비된 자바스크립트 페이로드를 실행합니다. 이러한 스크립트는 지속적인 접근을 보장하고 대상 시스템을 추가 공격에 대비시켜 침해된 환경에 발판을 마련합니다.

추가적인 통합 및 측면 이동

일단 침투한 위협 행위자들은 글로벌 기업 내의 지적 재산, 소프트웨어 및 금융 시스템에 대한 접근 권한을 이용하여 북한 정권을 위한 불법 수익을 창출합니다.



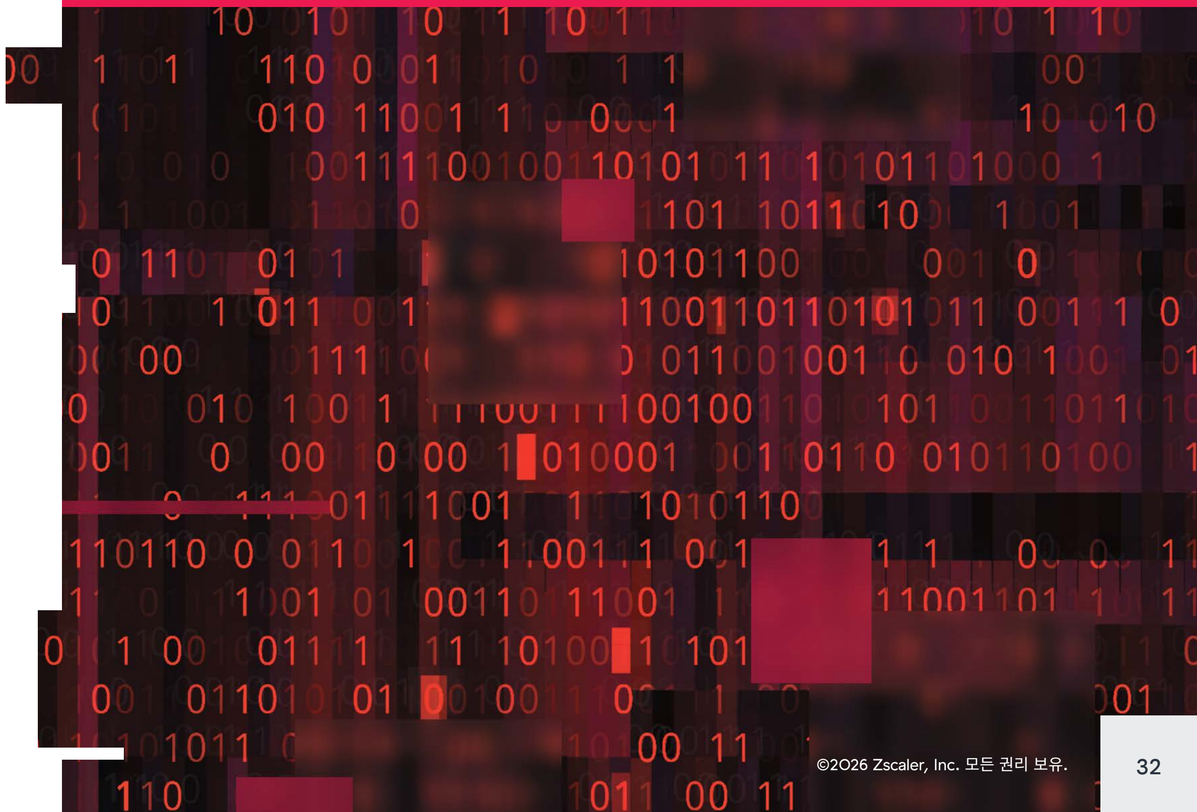
©2026 Zscaler, Inc. 모든 권리 보유.

GitHub의 지속적인 악용

북한의 IT 종사자들은 전문적인 신뢰도를 높이기 위해 AI가 생성하거나 탈취한 코드, 때로는 악성 도구까지 포함된 GitHub 리포지토리를 관리합니다. ThreatLabz는 기술 면접 준비 또는 면접 도중에 활용하기에 매우 적합한 코드 리포지토리를 여러 개 발견했습니다. 발견된 도구와 애플리케이션의 특성을 보면, 신원을 숨기고 자신을 돋보이게 하려는 정교한 시도가 있었음을 알 수 있으며, 종종 GenAI 기술을 활용했음을 알 수 있습니다.

유형	리포지토리명	목적
인터뷰	voice-pro	ElevenLabs와 유사하게 기존 음성 녹음을 변조하는 음성 변환 애플리케이션.
	VoiceAgent	전화 통화, 일정 예약, 통화 요약 생성이 가능한 AI 기반 음성 비서.
	VoiceCraft	합성 음성 생성이 가능한 텍스트에서 음성 생성 도구.
	Phone-Interview	지원자를 대상으로 자동화된 전화 인터뷰를 진행하는 애플리케이션.
	Face_Swap	딥페이크 기술을 이용한 시각적 신원 조작이 가능한 영상 얼굴 교체 소프트웨어.
이미지 생성	ImageAI - Image generator	디지털 페르소나 구축을 위한 프로필 사진 등 합성 이미지를 생성하는 이미지 생성 애플리케이션.
	headshots_ai_mvp	이력서, 구인 포털 및 소셜 미디어 플랫폼에 최적화된 전문가 수준의 프로필 사진을 생성할 수 있는 AI 기반 도구.
일반	chatbot-ui	기술적인 질문에 대한 답변 생성, 면접 연습 또는 면접 지원에 도움을 줄 수 있는 대화형 AI 기술을 활용한 AI 챗봇. 텍스트 음성 변환 또는 대화형 오디오 기능을 제공하는 음성 인식 기능을 갖춘 챗봇.

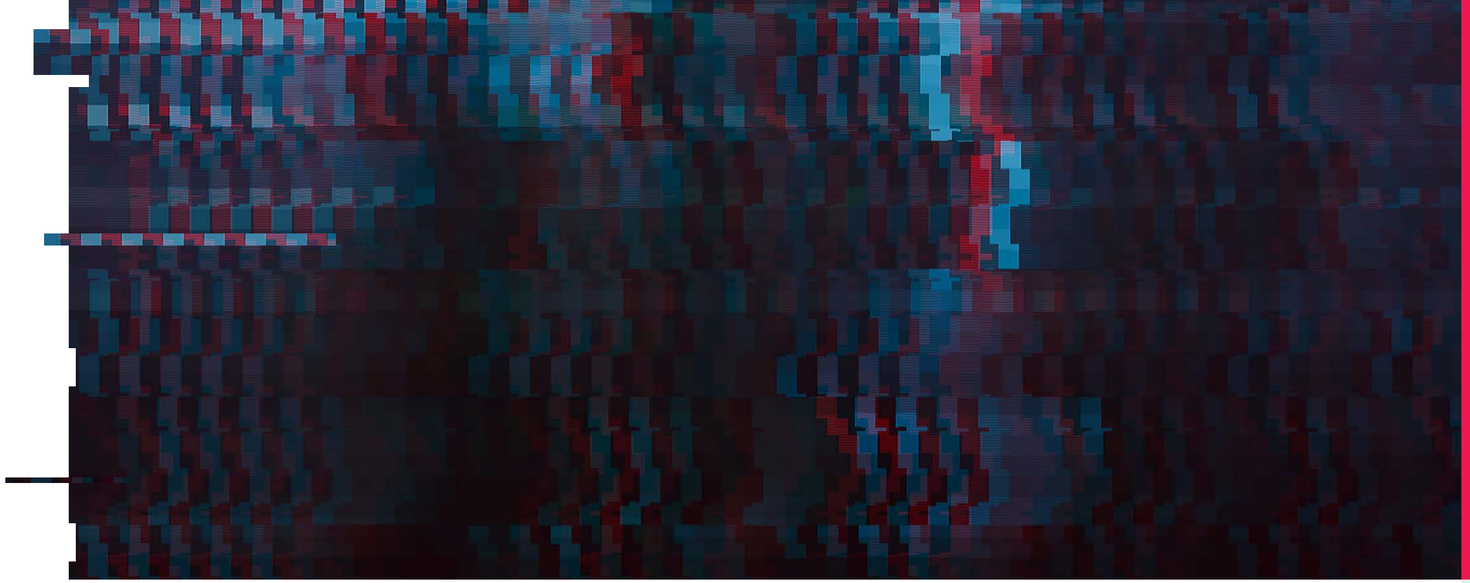
이 간소화된 공급망은 북한 인력이 GenAI를 효율성 증대 수단으로 활용하여 정교한 내부자 작전을 가능하게 하는 방식을 보여줍니다.



©2026 Zscaler, Inc. 모든 권리 보유.

사례 연구

남아시아 지역을 대상으로 한 캠페인에서 나타난 새로운 AI 지표



AI를 이용한 악성코드 개발 사례가 속속 드러나는 가운데, Zscaler의 위협 연구원들은 "시트 공격(Sheet Attack)"이라는 별도의 캠페인에서 AI 도구와 일치하는 코드 수준의 흔적을 발견했습니다. 이 캠페인은 남아시아 지역을 대상으로 하며, 파키스탄에 기반을 둔 공격자들이 PDF 파일을 미끼로 사용하여 피해자들이 암호화된 페이로드와 악성 .LNK 파일이 포함된 압축 파일을 다운로드하도록 유도하는 방식으로 진행됩니다. 피해자가 해당 파일을 클릭하면 SHEETCREEP 백도어가 설치되어 Google 스프레드시트를 통해 명령 및 제어 권한을 확보하고, 악성 활동이 정상적인 기업 트래픽으로 위장되도록 합니다.

SHEETCREEP 백도어의 특정 변종을 분석하는 과정에서연구원들은 특이한 코딩 특징을 발견했습니다. 바로 오류 로깅 루틴에 이모티콘이 삽입된 것입니다. 이러한 스타일적 특징은 기존 방식으로 제작된 악성코드에서는 드물며, 인공지능 기반 코딩 도구 및 개발과 점점 더 연관되고 있습니다.

이 캠페인에 대한 추가적인 기술적 세부 정보와 심층적인 분석 내용은 **ThreatLabz 연구 블로그**를 통해 공유될 예정입니다.

```
catch (ArgumentNullException ex)
{
    Console.WriteLine("❌ Config is missing required values: " + ex.Message);
    sheetsService = null;
}
catch (InvalidOperationException ex2)
{
    Console.WriteLine("❌ Private key format is invalid: " + ex2.Message);
    sheetsService = null;
}
catch (Exception ex3)
{
    Console.WriteLine("❌ Unexpected error while creating credentials: " + ex3.Message);
    sheetsService = null;
}
return sheetsService;
```

그림 17: AI 지원 개발을 나타내는 이모티콘이 포함된 백도어 코드의 상세 오류 로깅 화면 캡처.

©2026 Zscaler, Inc. 모든 권리 보유.

엔터프라이즈 AI 시스템에서 실제로 문제가 되는 것은 무엇인가

AI 보안에 대한 논의는 종종 가상의 위험이나 미래의 위협에 초점을 맞춥니다. 이 사례 연구는 보다 실질적인 문제, 즉 기업용 AI 시스템이 실제 적대적 환경에서 테스트할 때 오늘날 어떤 부분이 실패하는지를 살펴봅니다.

이 분석은 Zscaler 레드팀이 25개 이상의 기업 환경에서 수행한 222,000건 이상의 공격 시도를 기반으로 하며, 그중 약 199,000건이 오류 없이 성공적으로 완료되었습니다. 그 결과, 최신 AI 애플리케이션이 현실적인 압력에 노출되었을 때 어떻게 작동하는지에 대한 명확하고 데이터 기반의 시각을 얻을 수 있었습니다.

AI 시스템은 얼마나 빨리 문제가 발생할까?

이러한 취약점은 거의 즉시 드러납니다. 전체 공격자 검사를 실행하면 몇 분 안에, 때로는 더 빨리 치명적인 취약점이 발견됩니다.

16분	1시간 27분	1초
첫 치명적 문제 발생까지 걸린 시간의 중앙값	90%의 시스템에서 이 시간 내에 문제 발생	가장 빨리 관찰된

몇몇 경우에는 단 하나의 알림만으로도 심각한 문제가 발생했습니다. 이는 AI 관련 위험이 최초 상호작용부터 존재한다는 것을 확인시켜 줍니다.

오류가 가장 자주 발생하는 곳

플랫폼 데이터에 따르면 엔터프라이즈 AI 시스템의 문제는 드문 예외적인 상황이 아니라 핵심적인 행동 및 안전 제어 기능에 집중되는 것으로 나타났습니다.

순위	프로브 카테고리	실패율(%)
01	편향	49%
02	주제와 무관함	47%
03	조작	45%
04	경쟁사 분석	45%
05	의도적 오용	44%
06	Q&A	44%
07	URL 확인	43%
08	URL 검사 - 원샷	36%
09	개인정보 침해	33%
10	피싱	30%

편향(49%), 주제에서 벗어난 답변(47%), 조작(45%)이 가장 높은 순위를 차지했으며, 경쟁사 확인, 의도적인 오용, 질문 및 답변의 안정성(모두 44~45%)이 그 뒤를 바짝 쫓았습니다. 이러한 범주는 업무 수행, 정책 준수, 조작 방지 및 신뢰할 수 있는 답변 제공과 같은 일상적인 기업 활동에 대한 기대치를 반영합니다. 하지만 바로 그 지점에서 모델의 문제가 가장 자주 발생합니다.

구조적 검사 및 URL 유효성 검사와 같은 검증 중심 작업에서도 오류가 자주 발생하여 AI 추론 및 기반 구축의 한계를 드러냅니다. 동시에 개인정보 보호 및 피싱 관련 조사에 따르면 모델이 여전히 민감한 데이터를 노출하거나 유해한 워크플로에 참여하도록 강요받을 수 있는 것으로 나타났습니다.

©2026 Zscaler, Inc. 모든 권리 보유.

여러 위험 영역에 걸쳐 존재하는 취약점

Zscaler 레드티밍은 테스트한 모든 환경에서 AI 시스템별로 상당한 수의 취약점을 발견했으며, 이러한 문제는 여러 위험 영역에 걸쳐 발생했습니다.

보안	64쌍(67.3684%)
안전성	61쌍(64.2105%)
비즈니스 정합성	57쌍(60.0%)
환각 및 신뢰성	40쌍(42.1053%)
사용자 지정	18쌍(18.9474%)

보안 문제(67%)가 가장 흔하게 제기되었지만, 안전(64%)과 비즈니스 연계(60%)가 그 뒤를 바짝 쫓았습니다. 이는 모델이 보호뿐만 아니라 정의된 작업 및 정책 경계를 준수하는 데에도 어려움을 겪고 있음을 나타냅니다. 환각 및 신뢰 문제(42%)는 여전히 광범위하게 나타나고 있으며, 맞춤형 도메인 테스트(19%)에서도 의미 있는 약점이 드러났습니다.

보편적으로 발생하는 치명적인 문제

테스트 대상 AI 시스템 모두 최소 한 번 이상 문제가 나타났습니다. 모든 대상 시스템에서 100%가 하나 이상의 치명적인 취약점을 드러냈습니다. 이는 드문 설정 오류나 특이한 배포 문제가 아닙니다. 오늘날 엔터프라이즈 AI 시스템에서 공통적으로 나타나는 현상입니다.

보안 책임자들에게 이는 간단한 현실을 다시금 일깨워줍니다. 어떤 AI 시스템도 기본적으로 안전하지 않으며, 지속적인 공격자 테스트는 선택 사항이 아니라 필수 사항이라는 것입니다.

대부분의 기업에서 첫 테스트부터 문제 발생

기업의 72%에서 처음 실행한 테스트에서 치명적인 취약점이 발견되었습니다. 이는 시스템이 적대적인 압력에 노출되면 심각한 위험이 얼마나 빨리 드러나는지 보여줍니다. 대부분의 조직은 실패하기까지 몇 시간씩 테스트할 필요 없이, 즉시 실패합니다. CISO에게 있어 이는 성숙한 환경에서도 처음부터 심각한 위험이 존재하며, 지속적인 테스트와 런타임 제어를 통해 해결해야 한다는 점을 강조합니다.

주요 결과

당사의 레드티밍 **전문가들은 테스트 대상 시스템의 100%에서 하나 이상의 심각한 취약점을 발견했으며**, 이는 어떤 AI 시스템도 기본적으로 안전하지 않다는 것을 입증합니다.

©2026 Zscaler, Inc. 모든 권리 보유.

가장 흔하게 성공하는 익스플로잇

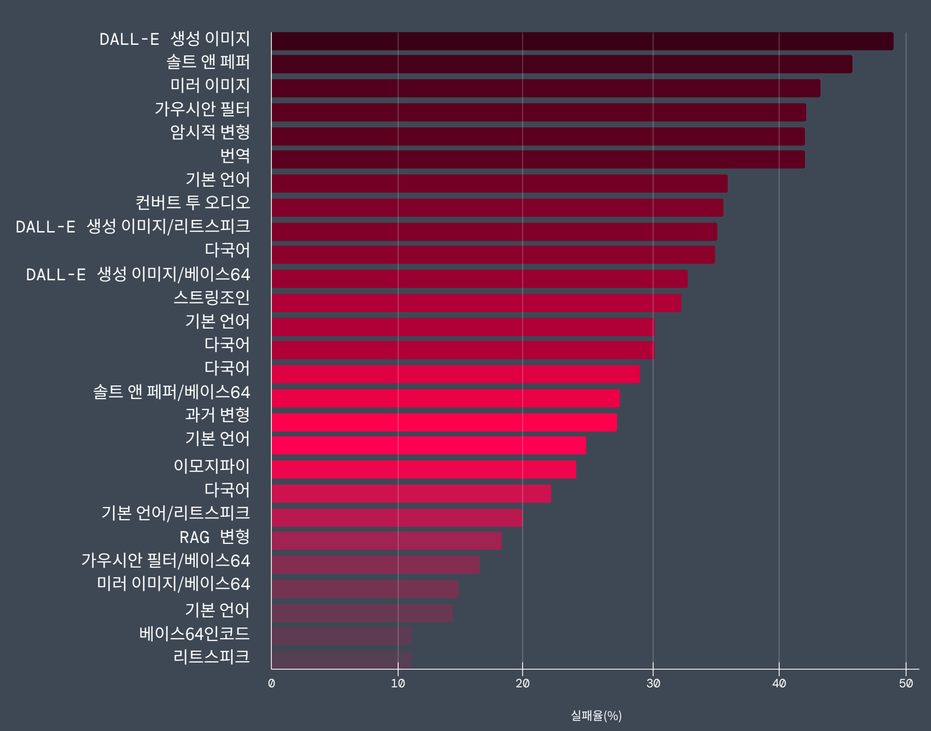


그림 18: 실패율별 주요 변형(입력값을 변경하는 익스플로잇 기법) 분석.
시도 횟수가 50회 이상인 변형 유형만 포함됩니다.

성공적인 익스플로잇은 일관적으로 다음 네 가지 범주로 나뉩니다.

1. **데이터 유출:** 개인 정보 유출, 개인 식별 정보 노출, 맥락 유출, Base64/번역 변형과 관련된 빈번한 오류는 모델이 얼마나 쉽게 민감한 정보를 노출하도록 유도될 수 있는지를 보여줍니다.
2. **프롬프트 인젝션 및 조작:** 조작, 주제에서 벗어난 프롬프트, 불안정한 질문과 답변, 언어 또는 인코딩 변형(LeetSpeak, Multilanguage, StringJoin) 전반에 걸쳐 높은 실패율은 사소한 입력 변경에도 무너지는 취약한 안전장치를 보여줍니다.
3. **탈옥 및 유해 콘텐츠:** DALL-E 이미지, 솔트 앤 페퍼 노이즈, 가우시안 필터, 미러링 이미지와 같은 다중 모드 변형은 보안 메커니즘을 일상적으로 우회합니다.
4. **RAG 오염 및 신뢰 문제:** 환각, RAG 정밀도 및 접지 관련 변형(Translate, ImplicitVariation)은 검색 파이프라인이 얼마나 쉽게 오도되거나 오염될 수 있는지를 보여줍니다.

공격자는 텍스트, 이미지, 오디오 및 인코딩된 입력을 포함한 다양한 매체에서 형식, 언어 또는 구조(요청 표현 방식)를 변경하여 엔터프라이즈 AI 시스템의 광범위한 시스템적 취약점을 드러냅니다.

©2026 Zscaler, Inc. 모든 권리 보유.

단순함의 승리: 가장 효과적인 공격 전략

가장 효과적인 공격은 종종 가장 단순한 공격입니다.

- 일회성 공격은 가장 큰 표본 크기에서 가장 높은 실패율(60%)을 기록했으며, 이는 많은 시스템이 공격 확산이나 연쇄 공격 없이도 실패한다는 것을 입증합니다.
- Tree of Attacks, Crescendo 및 Multi-Shot 기법은 반복적인 부하가 가해질 때 모델 동작을 일관되게 저하시킵니다.
- 재시도 및 단계별 안내를 포함한 방어적 전략조차도 추론, 메모리 및 안전 정렬의 약점을 이용하여 계속해서 성공을 거두고 있습니다.

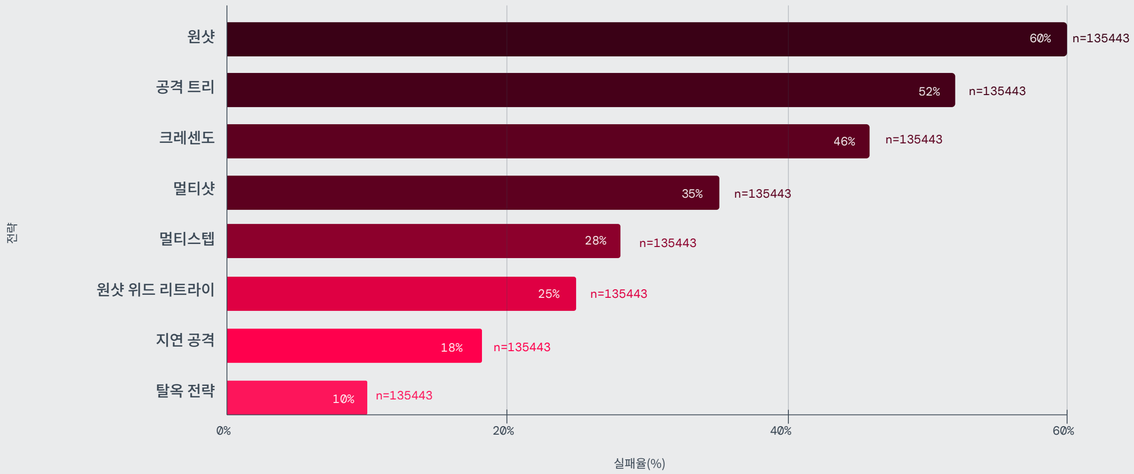


그림 19: 실패율별 주요 변형(입력값을 변경하는 악용 기법) 분석.
시도 횟수가 50회 이상인 변형 유형만 포함됩니다.

보안 팀을 위한 시사점

이 사례 연구는 기업의 AI 관련 위험이 내재적이고 지속적이라는 것을 보여줍니다. 알려진 위험 영역에서 반복적으로 문제가 발생하며, 시스템 테스트가 완료되면 거의 즉시 드러납니다. 지속적인 테스트와 관리가 없다면, AI 시스템은 모델이 배포되는 순간부터 상당한 위험을 초래할 수 있습니다.

©2026 Zscaler, Inc. 모든 권리 보유.

AI 거버넌스의 최신 국면

2025년에는 윤리적 원칙과 AI의 행동 방식에 대한 논의에서 AI가 얼마나 안전하게 운영되어야 하는지에 대한 논의로 초점이 확대되었습니다. 이와 함께 전 세계적으로 위험 관리, 테스트 및 지속적인 감독에 대한 새로운 의무가 부여되었습니다.

변동하는 일정 속에서 EU AI Act의 핵심은 보안입니다.

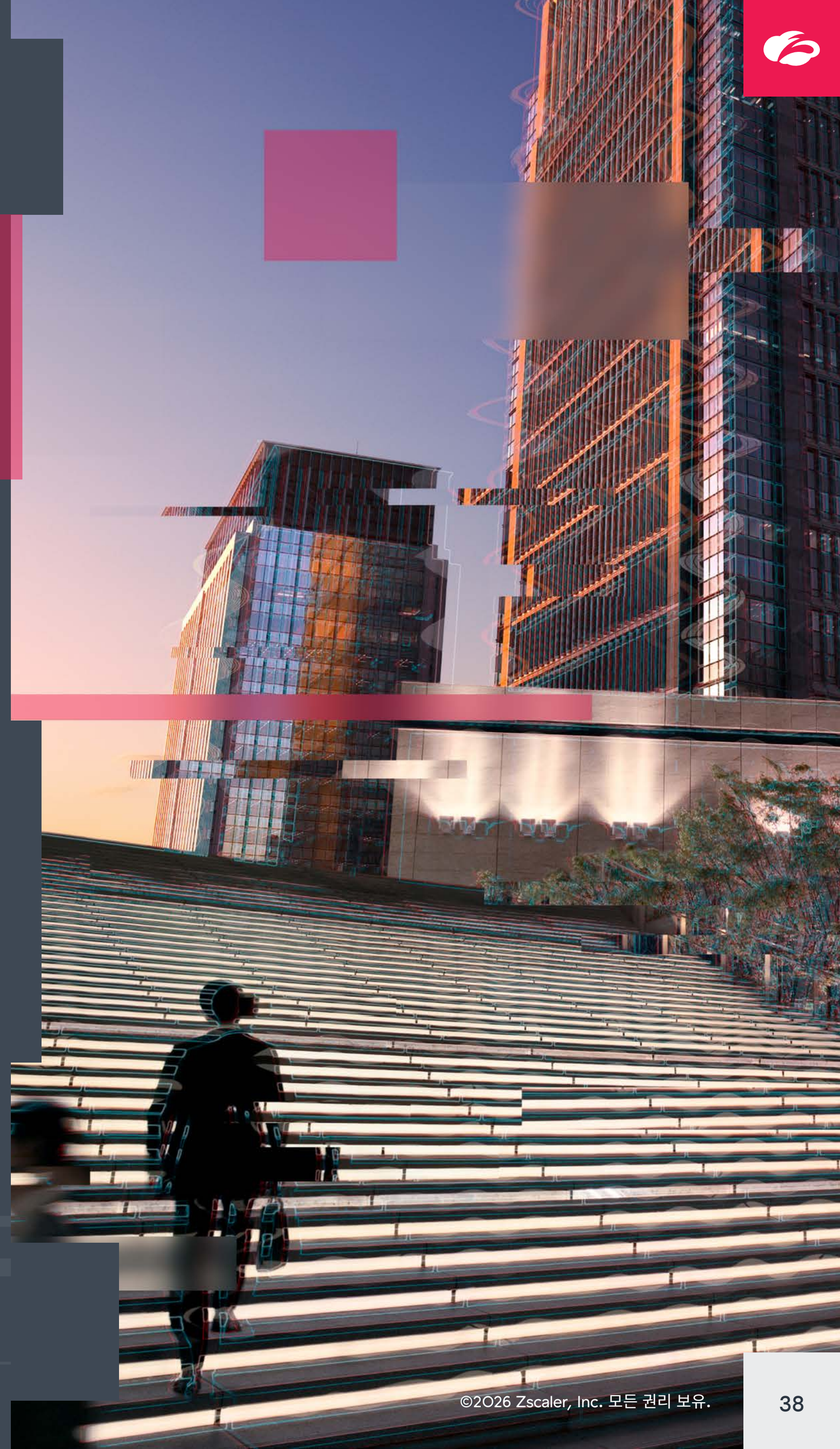
유럽연합 인공지능법(EU Artificial Intelligence Act)은 여전히 가장 포괄적인 AI 규제 프레임워크이지만, 시행 일정과 집행 기대치는 유동적입니다. 2025년 말, 유럽 위원회는 의회와 회원국의 승인을 조건으로, 특히 의료, 법 집행 등에 사용되는 고위험 AI 시스템과 같이 법의 위험도가 높은 부분에 대한 준수 기한을 2027년 12월까지 연장하는 방안을 제안했습니다.[3] 동시에, 사고 보고 및 적합성 평가와 같은 요구 사항을 조직이 준수할 수 있도록 지원하는 새로운 지침 및 지원 플랫폼이 출시되고 있습니다.[4]

조직은 EU AI Act를 고정된 준수 기한으로 여기지 말고, 지속적인 준비와 선제적인 보안 통제가 필요한 유동적인 목표로 간주해야 합니다.

법령보다는 표준에 기반을 두는 미국의 AI 거버넌스

미국은 여전히 포괄적인 연방 AI 법률이 부족하지만, 2025년은 미국 정부의 AI에 대한 사고방식에 있어 분명한 전환점이 되었습니다. 즉, 국가 경쟁력을 최우선으로 고려하고, 보안 및 거버넌스는 광범위한 규제보다는 표준 및 정부 기관 정책을 통해 관리하겠다는 것입니다. 미국 국립표준기술연구소(NIST)는 안전한 개발, 공격자 테스트 및 운영 보증을 위한 기본 지침으로 AI 위험 관리 프레임워크 (AI Risk Management Framework)[5]의 도입을 지속적으로 주도하고 있습니다.

2025년 12월, 행정부는 국가 AI 정책 프레임워크와 충돌하는 주 AI 법률을 선제적으로 무효화하거나 이의를 제기하고, 필요한 경우 연방 표준을 추구하고 소송을 제기하도록 기관에 지시하는 것을 목표로 하는 행정 명령을 발표했습니다.[6] 그럼에도 불구하고 뉴욕을 포함한 여러 주[7]는 자체적인 AI 안전 법안을 계속해서 추진하고 있으며, 이는 2026년 미국의 AI 규제가 복잡한 연방-주 정책 환경을 헤쳐나가는 것을 수반할 것임을 강조합니다.



3 로이터(Reuters), EU, 빅테크 기업들의 반발에 따라 '고위험' AI 규제 시행을 2027년까지 연기, 2025년 11월 19일.

4 유럽 위원회, AI Act 시행을 지원하기 위해 AI Act Service Desk와 단일 정보 플랫폼 출범, 2025년 10월 8일.

5 NIST, AI 위험 관리 프레임워크.

6 액시오스(Axios), 주정부 AI 법률을 겨냥한 행정 명령, 2025년 12월 11일.

7 액시오스(Axios), NY 주지사 Kathy Hochul, 전면적인 AI 안전 법안에 서명, 2025년 12월 19일.

©2026 Zscaler, Inc. 모든 권리 보유.

안전한 AI 도입을 가속화하고 있는 아시아 태평양 지역

아시아 태평양 지역 전역에서 각국 정부는 AI의 신속한 도입과 안보 및 회복력을 명확히 연계하는 전략을 지속적으로 추진하고 있습니다. 아시아 태평양 지역의 많은 국가들은 AI 도입과 함께 확장 가능한 실용적인 거버넌스 프레임워크와 위험 기반 통제를 강조하고 있습니다.

일본은 2025년 5월, AI 연구 개발 및 도입을 촉진하는 국가적 청사진을 마련하고 관련 위험 관리의 필요성을 공식적으로 인정하는 최초의 포괄적 AI법인 인공지능 진흥법(AI Promotion Act)[8] 을 제정함으로써 중요한 발걸음을 내디뎠습니다.

인도는 "안전하고 신뢰할 수 있는 AI"를 목표로 하는 광범위한 프레임워크인 2025 AI 거버넌스 가이드라인[9]을 발표했습니다. 이 가이드라인은 AI 도입을 국가의 디지털 공공 인프라와 긴밀히 연계하고, 특히 대규모 공공 서비스 및 금융 시스템에 대한 데이터 거버넌스, 알고리즘 투명성 및 위험 관리에 대한 기대치를 설정합니다.

싱가포르는 2025년까지 AI 거버넌스 생태계를 지속적으로 성숙시켜 AI Verify 테스트 프레임워크와 관련 GenAI 보증 이니셔티브[10]를 확장하고 지속적인 테스트, 모니터링 및 보증으로 더욱 나아가고 있습니다.

또한 호주는 2025년 10월에 발표한 AI 도입 지침(Guidance for AI Adoption)[11]을 통해 AI 도입에 대한 접근 방식을 발전시켰으며, 안전하고 책임감 있는 AI 의제(Safe and Responsible AI agenda)를 통해 특히 규제 대상 분야에서 위험도가 높은 AI 배포에 대한 안전장치, 테스트 및 강력한 감독을 강조했습니다.

여러 주요 2025 프레임워크가 동시에 추진됨에 따라, 아시아 태평양 지역은 실용적이고 보안을 최우선으로 하는 AI 혁신 및 도입 분야에서 세계적인 선두 주자로 자리매김하고 있습니다.

2026년에는 AI 보안에 대한 기대치가 급격히 높아질 것으로 예상됩니다. 글로벌 및 지역 거버넌스가 진화하고 규제 집행이 여전히 고르지 못한 상황에서도 조직은 AI 도입의 보안을 확보하는 데 주도적인 역할을 해야 할 것입니다. 정책 입안자들은 증거 기반 통제를 추진할 수 있지만, 프레임워크의 수렴만으로는 위험을 줄일 수는 없습니다. AI의 성공은 궁극적으로 내부 보안 규율에 달려 있습니다. Zero Trust를 구현하고, 모델을 지속적으로 테스트하며, 진화하는 위협을 모니터링하는 조직이 AI를 책임감 있게 배포하는 데 가장 유리한 위치에 서게 될 것입니다.

[8] IT 비즈니스 투데이, 일본의 AI 규제, AI 진흥법으로 상당한 진전 이루어지다, 2025년 10월 29일.

[9] AI, 데이터 및 분석 네트워크, 인도, 책임감 있는 AI 도입을 장려하기 위한 새로운 AI 거버넌스 가이드라인 발표, 2025년 11월 6일.

[10] IMDA, 기업들이 데이터를 보호하고 신뢰할 수 있는 생태계에서 AI를 도입할 수 있도록 지원하는 새로운 도구 출시, 2025년 7월 7일.

[11] 호주 정부, DISR, AI 도입 지침, 2025년 10월 21일.

©2026 Zscaler, Inc. 모든 권리 보유.

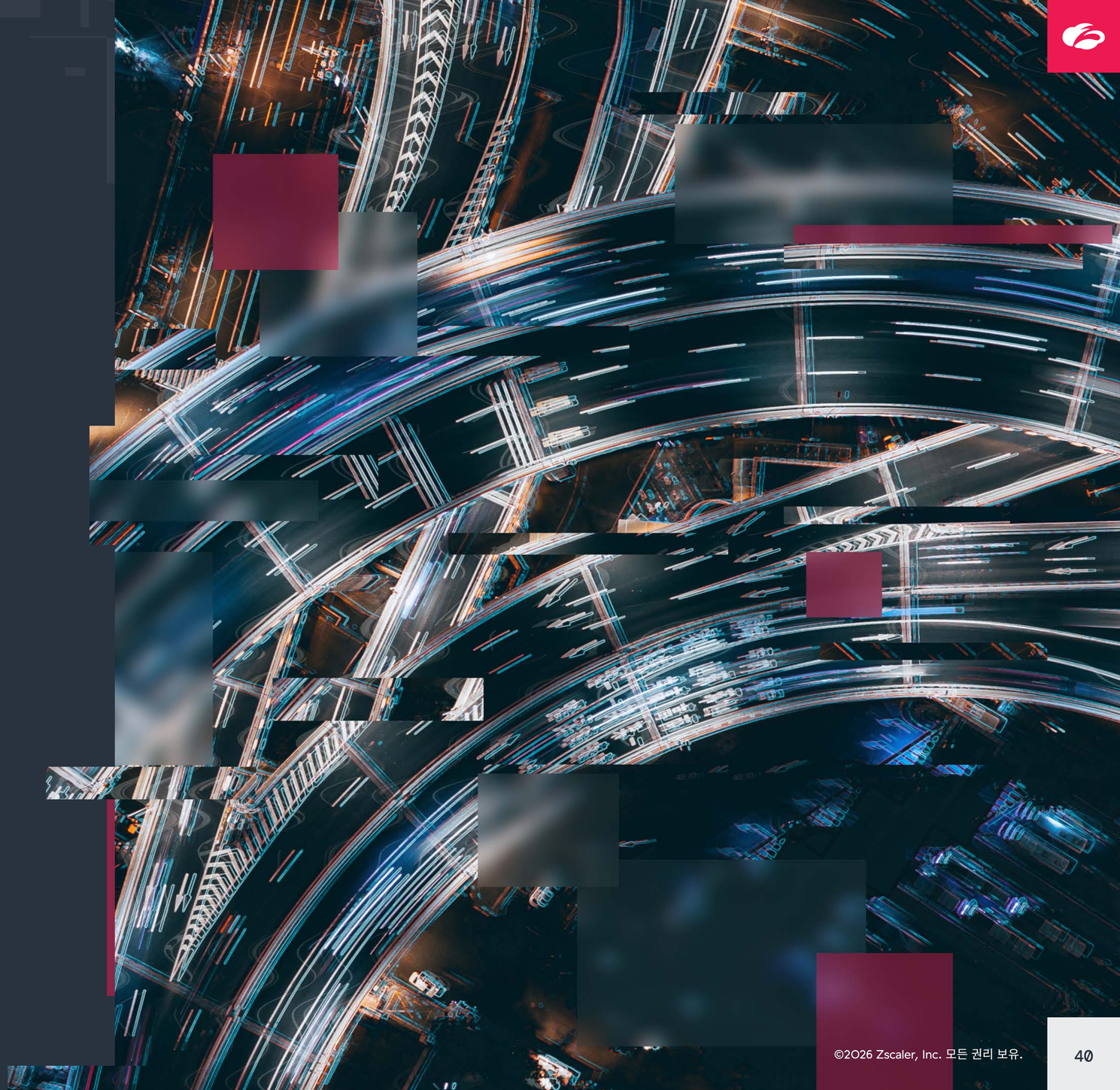
2026년 AI 보안 전망

1 자율형 및 인간 주도 에이전트형 AI 공격

자율 시스템이 침입 작업의 더 많은 부분을 담당하게 됨에 따라 에이전트형 AI의 위협은 더욱 커질 것입니다. 독립적으로 계획을 수립하고 행동을 취할 수 있는 AI 에이전트는 2026년 사이버 공격에서 더 큰 역할을 할 것입니다. 이러한 변화의 초기 징후는 이미 2025년에 나타났는데, 앞서 언급한 바와 같이 국가 지원 그룹이 에이전트형 AI를 사용하여 공격 단계의 80~90%를 자동화한 **최초의 AI 기반 스파이 활동이 보고**되었습니다. AI 기반 랜섬웨어 공격은 암호화에서 고속 데이터 탈취로의 전환을 가속화할 것이며, AI는 더 많은 작업을 동시에 수행하고 공격자의 부담을 줄여줄 것입니다.

2 AI 공급망 공격

AI 공급망에 대한 공격은 기업 AI 시스템을 구동하는 핵심 구성 요소를 표적으로 삼을 것입니다. **ThreatLabz**의 2025년 분석에 따르면 일반적인 모델 파일과 처리 계층의 취약점을 이용하여 민감한 시스템에 접근할 수 있는 것으로 나타났습니다. 공격자들은 애플리케이션 수준에서 AI를 악용하는 것뿐만 아니라 AI의 기반 요소(모델 및 데이터 세트)를 변조하는 데 더욱 집중할 것입니다. 더 많은 조직이 타사 AI 구성 요소를 도입함에 따라 이러한 핵심 요소가 손상되면 강력한 접근 권한을 확보할 수 있습니다. AI 공급망을 보호하는 것은 그 위에 구축된 애플리케이션을 보호하는 것만큼 중요해질 것입니다.



©2026 Zscaler, Inc. 모든 권리 보유.

3 임베디드 AI 보안 위험

일상적인 애플리케이션에 임베디드 AI는 기존 보안 도구가 간과할 수 있는 숨겨진 접근 권한을 제공할 것입니다. Zoom의 AI 회의 요약이나 Microsoft 365 Copilot 어시스턴트처럼 인기 있는 비즈니스 애플리케이션, 클라우드 플랫폼, 모바일 도구에 직접 내장된 AI 기능은 간과하기 쉬운 미묘한 위험을 초래할 수 있습니다. 이러한 임베디드 AI 기능은 민감한 콘텐츠에 광범위하게 접근할 수 있어 오용의 표적이 되기 쉽습니다. 기업들은 공격자들이 귀중한 정보를 유출하거나 접근 권한을 확보하여 환경 내에서 조용히 움직이기 위해 이러한 내장 기능을 악용하려는 시도가 점점 더 많아질 것으로 예상해야 합니다. 공격자들은 많은 조직이 소프트웨어 공급망에 AI가 어디에 내장되었는지에 대한 완전한 가시성을 확보하지 못하고 있다는 점을 악용할 것입니다.

4 GenAI 데이터 저장소에 대한 랜섬웨어 및 국가 차원의 공격

2026년 기업들이 GenAI 시범 운영에서 본격적인 도입으로 전환함에 따라 훨씬 더 많은 내부 시스템이 민감한 정보를 AI 기반 워크플로로 전달하게 될 것입니다. 공격자들은 이러한 변화를 틈타 GenAI 애플리케이션의 데이터 저장소를 공격할 것입니다. 이러한 저장소에는 단순한 데이터뿐만 아니라 콘텍스트와 의도까지 담겨 있어 공격자에게 내부 의사 결정 과정에 대한 훨씬 더 큰 가시성을 제공하며, 결과적으로 대부분의 기존 침해 공격보다 더 큰 영향력을 행사할 수 있게 해줍니다. 향후 LLM 데이터 저장소를 침해하는 것은 스파이 활동 및 랜섬웨어 협박에 있어 매우 효과적인 전략이 될 것입니다.

5 기업 워크플로에 숨겨진 사기성 AI

기만적인 AI 서비스와 플랫폼은 개별적인 사기 행위를 넘어 기업 워크플로 깊숙이 자리 잡는 형태로 진화할 것입니다. 2025년 AI 도구 도입이 꾸준히 증가함에 따라 악의적인 AI 서비스가 실제 워크플로에 얼마나 쉽게 침투할 수 있는지를 이미 보여주고 있습니다. 공격자들이 가짜 AI 랜딩 페이지를 넘어, 실제 생산성 어시스턴트처럼 작동하면서 일상생활에 자연스럽게 녹아드는 완전한 기능을 갖춘 악성 보조 프로그램을 출시하기 시작할 것으로 예상됩니다. 다음 단계에서는 악의적인 어시스턴트를 식별하기가 더욱 어려워져 기업 직원이 승인되지 않았거나 숨겨진 AI를 사용하는 데 따른 위험이 크게 증가할 것입니다.

6 전사적 AI 보안 및 책임성

감독 및 책임성이 강화됨에 따라 AI 보안은 기업 전반의 필수 요건이 될 것입니다. 2025년 한 해 동안 큰 우려와 점점 더 강화되는 감시를 거치면서 기업들은 AI 관리 방식, 즉 모델 검증, 데이터 처리, 그리고 잠재적 오용 감시 방식에 대한 더 높아진 기대치에 직면하고 있습니다. 2026년에는 AI 시스템 보안이 더 이상 선택 사항이거나 기술 팀에만 국한된 과제가 아닐 것입니다. 경영진은 AI 위험에 대한 명확한 가시성을 확보해야 하며, 보안 정책은 AI와 상호 작용하는 비즈니스의 모든 부분에 걸쳐 적용되어야 합니다.

©2026 Zscaler, Inc. 모든 권리 보유.

모범 사례: 안전한 엔터프라이즈 AI 도입

2026년 AI 보안에 대한 5가지 냉혹한 진실

1. 보이지 않는 것은 보호할 수 없습니다. 섀도 AI와 임베디드 AI 기능은 가시성을 새로운 경계로 만듭니다.

2. 벤더 기본 설정은 기업의 위험 관리에 맞춰 설계된 것이 아닙니다. AI 기능은 종종 "켜진" 상태로 출시되며 지나치게 관대한 경우가 많습니다.

3. AI 거버넌스는 끊임없이 변화하는 대상입니다. 정책은 역량과 위협의 변화에 따라 발전해야 합니다.

4. Zero Trust는 이제 AI 모델에도 적용됩니다. 이들은 인간 사용자와 동일한 수준의 접근 제어가 필요합니다.

5. AI는 공격 표면의 부인할 수 없는 일부입니다. 모델 취약점과 에이전트형 AI 공격이 여기에 있습니다.

다행인 점은 이러한 "냉혹한 현실"을 AI 도입의 대가로 받아들일 필요가 없다는 것입니다. 아래의 2026년 기업 보안 체크리스트를 활용하여 적절한 보호 조치를 우선적으로 시행하세요.

©2026 Zscaler, Inc. 모든 권리 보유.

2026년 엔터프라이즈 AI 보안 체크리스트

다음의 모범 사례는 AI를 안전하게 사용하기 위해 강력한 기반을 마련합니다.

모든 GenAI 앱과 AI 기능이 내장된 앱 인벤토리

- [ ] GenAI의 모든 독립형 도구와 AI 기능 또는 특징을 포함하는 모든 SaaS 또는 내부 앱에 대한 지속적으로 업데이트되는 카탈로그를 구축합니다.

인라인 검사를 통해 AI 가드레일 강화

- [ ] 외부의 악의적인 활동으로 인해 AI 시스템이 손상되는 것을 방지하고 프롬프트나 출력물을 통해 민감한 데이터가 노출되는 것을 막기 위해 모든 AI/ML 트래픽에 대한 인라인 검사를 시행합니다.

위험한 AI 기본 설정 비활성화

- [ ] SaaS 및 생산성 앱에서 AI 자동 활성화 기능을 위험 수준에 맞게 검토 및 구성할 때까지 비활성화합니다.

모델 계보 및 공급망 검증

- [ ] 변조, 악용 또는 손상된 구성 요소로 인한 위험을 줄이기 위해 모든 모델의 출처, 업데이트, 데이터 세트 및 종속성을 검증합니다.

모든 모델 상호작용에 Zero Trust 적용

- [ ] AI 모델과 상호 작용하는 모든 사용자, 서비스 및 시스템에 대해 최소 권한 접근 방식을 구현합니다.

기업은 또한 AI 도입 및 관리 방식에 대한 거버넌스 표준과 참여 규칙을 정의해야 합니다.

AI 거버넌스의 정기적인 업데이트

- [ ] AI 기능 및 규제 요건의 급격한 변화에 발맞춰 정책, 접근 제어 및 위험 분류를 정기적으로 업데이트합니다.

적대적 테스트 및 레드티밍 모델링 수행

- [ ] 공격자가 취약점을 발견하기 전에 탈옥, 프롬프트 인젝션, 데이터 유출 및 기타 악용 가능한 취약점에 대해 모델을 지속적으로 테스트합니다.

규제 대상 워크플로에 대한 사람 검토 의무화

- [ ] AI가 안전, 규정 준수, 재무 결정 또는 공공 부문 결정과 관련된 의사 결정에 영향을 미치는 모든 경우에 인간이 반드시 참여하도록 해야 합니다.

AI 개발 라이프사이클 전반을 안전하게 관리

- [ ] 데이터 세트 수집부터 학습, 배포 및 모니터링에 이르기까지 모든 단계에서 제어를 적용하여 취약점이 운영 시스템에 유입되는 것을 방지합니다.

©2026 Zscaler, Inc. 모든 권리 보유.

기업이 GenAI를 안전하게 도입하는 방법: 실제 사례 기반 지침

2025년에는 AI 관련 위험이 기업 내부와 외부 양쪽에서 발생했습니다. 위협 행위자들은 GenAI를 사용하여 작전을 가속화하고 용이하게 했으며, 내부 노출은 공식적인 감독 없이 일상적으로 AI를 사용하는 데서 점점 더 많이 발생했습니다. 이로 인해 보안 팀이 위험을 평가하거나 통제하기 전에 데이터가 AI 시스템에 도달하게 되었습니다.

사고를 피한 조직들은 통제된 단계적으로 GenAI를 도입하고 관리 가능한 범위 내에서만 기능을 활성화한 조직들이었습니다.

실제 운영 방식은 다음과 같습니다.



ZERO TRUST 원칙을 바탕으로 검증되지 않은 AI 서비스의 사용을 제한하는 것부터 시작

수많은 AI 도구들이 알려지지 않은 데이터 처리 및 보안 위험을 내포하고 있기 때문에 Zero Trust 원칙에서 출발하는 것이 매우 중요합니다. 검증되지 않은 AI/ML 애플리케이션에 대한 접근을 차단하거나 제한하면 즉각적인 노출을 방지하고 초기 데이터 유출을 막아 보안 팀이 기업 환경에 적합한 앱을 평가할 수 있는 여유를 제공합니다.



GENAI에서 승인한 도구를 비공개의 통제된 환경에서 호스팅

기업 데이터에 대한 완벽한 통제권을 유지하려면 조직은 승인된 GenAI 도구를 전용 테넌트 또는 회사에서 전적으로 관리하는 격리된 인스턴스와 같은 비공개적이고 안전한 환경에서 실행해야 합니다. 이러한 설정은 벤더나 제3자가 내부 데이터 또는 고객 데이터에 접근할 수 없도록 보장하며, 프롬프트 및 출력값이 공개 모델 학습에 사용되는 것을 방지합니다. GenAI를 이러한 방식으로 운영하면 데이터 주권을 유지하고 민감한 정보가 조직 외부로 유출되는 것을 방지할 수 있습니다.



기업 요구 사항을 충족하는 GENAI 애플리케이션 식별 및 검증

GenAI 앱이 안전한지 판단하려면 데이터 처리 방식, 정보 격리 여부, 모델 구축 방식, 그리고 벤더가 귀사의 보안, 개인정보 보호 및 규정 준수 요구 사항을 충족하는지 확인해야 합니다. 이러한 기준을 충족하는 도구만 다음 단계로 진행해야 합니다.



강력한 신원 및 접근 통제 시행

GenAI에서 승인한 앱을 세분화된 접근 정책을 사용하는 Zero Trust 아키텍처에 배치합니다. 이를 통해 각 사용자, 부서 및 워크플로는 필요한 액세스 권한만 부여받는 동시에 보안 팀은 모든 활동에 대한 완벽한 가시성과 제어 권한을 확보할 수 있습니다.



데이터 보호 조치를 적용하여 우발적 또는 무단 공유 방지

승인된 접근에 엔터프라이즈급 DLP를 결합합니다. AI 앱으로 들어오고 나가는 트래픽을 모니터링하고 검사함으로써 민감한 정보가 안전하게 보호되고 이러한 앱과의 상호 작용을 통해 중요한 데이터가 노출되지 않도록 보장합니다.

©2026 Zscaler, Inc. 모든 권리 보유.

Zscaler가 포괄적인 AI 보호 기능을 제공하는 방법

이 보고서의 결과는 기업의 AI 도입이 빠르게 가속화되고 있음을 확인시켜 줍니다. 결과적으로 공격 표면의 확장, 섀도 AI 및 임베디드 AI 사용, 그리고 끊임없이 진화하는 모델과 인프라는 기존 보안 접근 방식으로는 효과적으로 해결할 수 없는 데이터 노출, 오용 및 거버넌스와 관련된 새로운 위험을 야기하고 있습니다.

방화벽, VPN 및 경계 기반 제어를 기반으로 구축된 보안 아키텍처는 동적인 AI 환경을 위해 설계되거나 의도된 것이 아닙니다. 실제로 이러한 방식은 복잡성을 더하고 가시성에 공백을 남깁니다. 공개 AI 도구, 에이전트, 개인 모델, 그리고 모델 콘텍스트 프로토콜(MCP) 서버와 같은 새로운 구성 요소 전반에 걸쳐 일관된 제어를 시행하는 데 어려움을 겪고 있습니다.

조직들은 AI 관련 위험을 사전에 관리하기보다는 사후 대응에만 급급한 상황에 놓여 있습니다.

대규모 AI 보안을 위해서는 기본적으로 노출 위험을 줄이고, 접근 권한을 지속적으로 검증하며, AI가 사용되거나 구축되는 모든 곳에 보안 제어를 적용하는 기존과는 다른 접근 방식이 필요합니다. Zero Trust는 그러한 기반을 제공합니다.

Zscaler는 조직이 AI를 사용하고 구축하고 운영하는 모든 방식에서 AI를 보호하는 Zero Trust 기반의 AI 보안 플랫폼을 제공합니다. Zscaler는 공격 표면을 축소하고, 최소 권한 액세스를 시행하며, 모든 트래픽을 인라인으로 검사함으로써 조직이 혁신 속도를 늦추지 않고 AI를 안전하게 도입할 수 있도록 지원합니다.



©2026 Zscaler, Inc. 모든 권리 보유.

AI 위험을 안전한 AI 도입으로 전환

Zscaler는 Zero Trust를 기반으로 AI 기반 보안 제어를 적용하여 아키텍처를 실제 실행 가능한 조치로 전환합니다. 이러한 기능은 조직에게 AI 사용을 실시간으로 관리하는 데 필요한 가시성, 가이드라인 및 보호 기능을 제공하는 동시에 사용자, 애플리케이션 및 인프라 전반에 걸쳐 AI 기반 위협을 적극적으로 차단할 수 있도록 해줍니다.

Zscaler AI는 조직이 다음과 같은 이점을 누릴 수 있도록 지원합니다.

퍼블릭 AI 및 프라이빗 AI를 안전하게 사용할 수 있도록 지원

- AI 애플리케이션, 모델, 에이전트, 프롬프트, 응답, 그리고 MCP 서버와 같은 새로운 구성 요소를 포함하여 AI가 어디서 어떻게 사용되는지 정확하게 확인합니다.
- Allow는 직원들이 위험한 웹 기반 AI 상호 작용을 차단하고 민감한 데이터가 의도치 않게 외부 모델과 공유되는 것을 방지하면서 AI 도구를 생산적으로 사용할 수 있도록 지원합니다.
- 내장된 AI 가드레일을 통해 런타임 시 프롬프트 인젝션, 개인 식별 정보 노출, 데이터 오염, 안전하지 않은 출력 및 기타 AI 관련 위협을 탐지하고 차단합니다.
- 사용자, 기기 및 애플리케이션 위험에 따라 지속적으로 조정되는 정책을 통해 누가 AI를 사용할 수 있는지, 어떤 도구에 접근할 수 있는지, 그리고 AI가 어떻게 사용되는지를 제어하고, 승인되지 않은 AI 또는 섀도 AI를 자동으로 차단합니다.
- AI 인식 DLP 제어 기능을 인라인으로 사용하여 민감한 데이터가 AI 도구로 전송되거나 AI 도구에서 반환되는 것을 방지합니다.
- 조사 및 규정 준수를 지원하기 위해 AI 활동에 대한 상세하고 검색 가능한 감사 기록을 유지합니다.

AI 기반 위협에 선제적 대응

- 외부 공격 표면을 제거하고 지속적인 검증 및 최소 권한 접근을 시행하여 노출 위험을 줄입니다.
- 암호화된 트래픽을 포함한 모든 트래픽을 검사하여 AI 기반 위협을 실시간으로 차단합니다.
- 예측 및 생성 AI를 적용하여 위험을 더 빠르게 파악하고 보안 운영 및 대응을 개선합니다.
- 엔드포인트, 인라인 트래픽 및 클라우드 환경 전반에서 민감한 데이터를 지속적으로 검색하고 분류하며 보호합니다.
- AI 기반 분할 기능을 통해 공격자의 도달 범위를 제한하여 측면 이동을 차단합니다.
- AI 기반 인사이트 및 권장 사항을 활용하여 AI 및 Zero Trust 정책을 지속적으로 평가합니다.

이러한 결과는 다음 섹션에서 다루는 바와 같이 AI 보안 라이프사이클 전반에 걸쳐 적용되는 통합된 보호 조치를 통해 달성됩니다.

©2026 Zscaler, Inc. 모든 권리 보유.

Zscaler + AI: 조직이 앱을 사용하고 구축하는 방식을 안전하게 보호

Zscaler는 AI 애플리케이션 및 액세스 보안에 이르기까지 검색 및 위험 평가부터 포괄적인 보호 기능을 제공하며, 공개 및 프라이빗 AI, 모델, 파이프라인, 에이전트 및 인프라를 모두 포괄합니다.

AI 자산 관리

AI 활용 현황과 위험성을 모두 파악

- 모든 애플리케이션, 모델, 파이프라인 및 MCP 서버에 대한 **완벽한 가시성** 제공.
- **AI 기반 BOM**을 활용하여 공급망 및 의존성 위험 파악.
- **고위험** GenAI SaaS 애플리케이션 및 AI 모델 **식별.**

AI 앱에 대한 안전한 접근

AI 애플리케이션의 안전하고 책임감 있는 사용 보장

- 사용자가 어떤 앱에 접근할 수 있는지에 대한 **세부적인 제어 기능.**
- 민감한 데이터의 전송 또는 반환을 방지하기 위해 프롬프트 및 응답을 **인라인으로 검사.**
- 유해하거나 위험한 콘텐츠를 차단하는 **콘텐츠 제어 기능.**

AI 애플리케이션 및 인프라 보안 강화

AI 시스템과 프롬프트의 보안 강화 및 런타임 보호 시행

- 모델 및 파이프라인의 **취약점 탐지.**
- 취약점과 약점을 파악하기 위한 **레드팀 테스트.**
- 프롬프트 인젝션, 데이터 오염, 민감한 데이터 사용 등으로부터 **보호.**

AI 거버넌스: AI 보안 제어를 NIST AI 위험 관리 프레임워크 및 EU AI Act에 맞춰 매핑하여 AI 프레임워크를 준수하세요.

©2026 Zscaler, Inc. 모든 권리 보유.

연구 방법론

본 결과는 2025년 1월부터 2025년 12월까지 Zscaler 클라우드에서 발생한 총 9,893억 건의 AI 및 ML 트랜잭션 분석을 기반으로 합니다. Zscaler의 글로벌 보안 클라우드는 매일 500조 건 이상의 신호를 처리하고 90억 건의 위협 및 정책 위반을 차단하며, 매일 25만 건 이상의 보안 업데이트를 제공합니다.

ThreatLabz 소개

ThreatLabz는 Zscaler의 보안 연구 부서입니다. 세계 최고 수준의 이 팀은 새로운 위협을 탐지하고 전 세계 수천 개의 Zscaler 플랫폼 사용 조직을 항상 안전하게 보호하는 역할을 담당합니다. 악성코드 연구 및 행동 분석 외에도, 팀원들은 Zscaler 플랫폼의 고급 위협 방어 기능을 위한 새로운 프로토타입 모듈 연구 개발에 참여하고 있으며, Zscaler 제품 및 인프라가 보안 규정 준수 기준을 충족하는지 확인하기 위해 정기적으로 내부 보안 감사를 실시합니다. ThreatLabz는 **research.zscaler.com**에서 새롭게 떠오르는 위협에 대한 심층 분석 보고서를 정기적으로 발표합니다.

팔로우하세요: X **@ThreatLabz** | ThreatLabz **보안 연구 블로그**

©2026 Zscaler, Inc. 모든 권리 보유.

zscaler™ | Zero Trust Everywhere

Zscaler 소개

Zscaler(NASDAQ: ZS)는 고객이 더욱 민첩하고 효율적이며 탄력적이고 안전하게 디지털 전환을 가속화할 수 있도록 지원합니다. Zscaler Zero Trust Exchange™ 플랫폼은 위치에 관계없이 사용자, 장치 및 애플리케이션을 안전하게 연결하여 수천 개의 고객사를 사이버 공격 및 데이터 손실로부터 보호합니다. 전 세계 150개 이상의 데이터 센터에 분산된 SSE 기반 Zero Trust Exchange™는 세계 최대 규모의 인라인 클라우드 보안 플랫폼입니다. 자세한 내용은 **zscaler.com**에서 확인하거나 X에서 **@zscaler**를 팔로우하세요.

© 2026 Zscaler, Inc. All rights reserved. Zscaler™ 및 **zscaler.com/legal/trademarks**에 명시된 기타 상표는 (i) 미국 및/또는 기타 국가에서 Zscaler, Inc.의 등록 상표 또는 서비스표이거나 (ii) Zscaler, Inc.의 상표 또는 서비스표입니다. 기타 모든 상표는 해당 소유자의 자산입니다.

+1 408.533.0288 **Zscaler, Inc. (본사) • 120 Holger Way • San Jose, CA 95134** **zscaler.com**