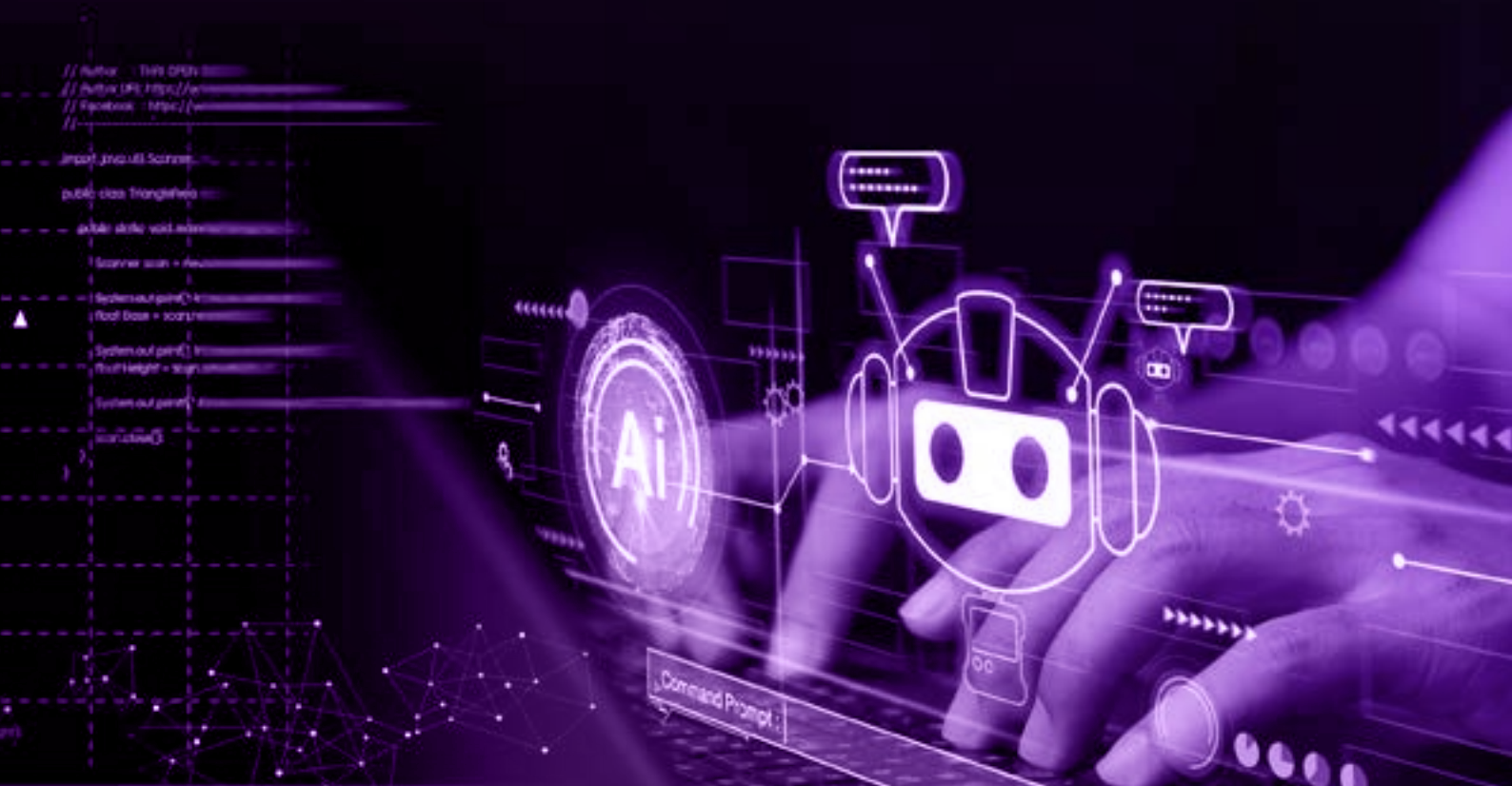
MENLO SECURITY

조직 내 생성형 AI 도입과 보안: 영향과 전략

ChatGPT와 유사한 생성적 인공지능 플랫폼이 직원들의 작업 방식을 혁신적으로 변화시키고 있습니다



SNAPSHOT

생성형 AI의 생산성 향상은 동시에 새로운 위험을 초래하고 있습니다

2022년 11월 출시 이후, ChatGPT는 역사상 가장 빠르게 성장하는 플랫폼 중 하나로 손꼽히며, 겨우 2개월 만에 1억 명 이상의 사용자를 모았습니다. 이는 TikTok이 9개월, Instagram이 2.5년이 걸린 것과 비교해보면 놀라운 행보입니다. ChatGPT는 전 세계적으로 사람들의 큰 관심을 끌어 일으키며, 이미 많은 생성적 인공지능 사이트 중 하나로 자리매김하고 있습니다. 이러한 플랫폼은 직원들이 새로운 아이디어를 도출하고 이메일을 개선하며 콘텐츠를 작성하고 생산성과 혁신을 높이는 데에 기여하고 있습니다.

생산성과 혁신을 높이는 한편으로, ChatGPT와 같은 플랫폼은 사이버 보안에 대한 우려를 증폭시키고 있습니다. 이러한 도구들은 위협 주체들이 대규모 위협을 개발할 수 있는 가능성을 부각시켜 왔습니다. 특히, ChatGPT는 해커들이 더 정교하고 효과적인 피싱 공격을 실시하는 데 일조하고 있습니다. 사이버 보안 전문가들은 ChatGPT가 기업과 개인에게 미치는 위험을 경고하고 있지만, 이러한 생성적 인공지능 플랫폼과 챗봇이 미치는 더 즉각적인 부정적인 영향, 특히 독점 데이터나 지적 재산(IP)의 잠재적인 유실에 대한 우려는 간과되기 쉽습니다.

직원들이 ChatGPT나 Bard와 같은 생성적 인공지능 도구를 사용함에 따라, 민감한 기업 데이터가 공유되고 노출될 우려가 있습니다. 이는 고객 데이터, 영업 비밀, 기밀 정보, 그리고 지적 재산까지 모두를 포함할 수 있습니다. 생성적 인공지능을 통해 개인 데이터는 일반적인 데이터 유실 경로보다 훨씬 더 넓은 대중에게 전달될 수 있는 잠재력을 가지고 있습니다. ChatGPT 및 기타 생성적 인공지능 플랫폼은 채팅 기록과 같은 데이터를 저장하여 모델을 훈련하고 개선하는 데 사용합니다. 이는 입력된 모든 데이터가 모델 훈련에 활용될 수 있으며, 나중에는 다른 사용자들에게 노출될 수도 있다는 의미를 갖고 있습니다.

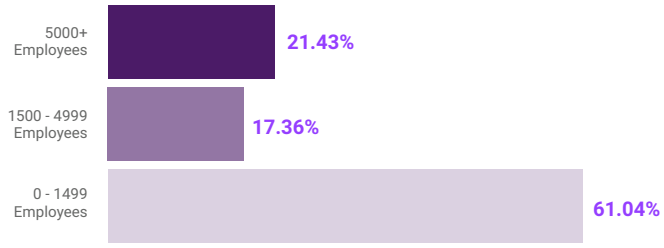
실제 사용 사례로, 삼성의 반도체 그룹 일부 엔지니어들이 최근에 회사가 개발 중인 새로운 기능의 소스 코드를 ChatGPT에 입력하여 효율성을 높일 수 있는지 확인한 것으로 알려져 있습니다. ChatGPT와 같은 생성적 인공지능 도구는 입력 데이터를 보존하고 자체적으로 훈련하는 특성을 가지고 있습니다. 이로써 삼성의 소스 코드는 다른 사용자들의 요청에 응답하기 위해 사용될 수 있게 되었습니다. 이러한 상황은 취약점을 찾는 위협 주체나 독점적인 정보를 탐색하는 경쟁업체 등에 대한 잠재적인 위험을 내포하고 있습니다.

일부 기업은 조직을 보호하기 위해 생성형 인공지능 사이트를 완전히 금지하는 등 다양한 조치를 취하고 있습니다. 예를 들어, 이탈리아는 데이터 개인 정보 보호 우려로 인해 ChatGPT를 전국적으로 일시적으로 금지한 적이 있었으며, 서비스는 약 한 달 후에 복원되었습니다. 생성적 인공지능 서비스에 대한 액세스를 차단하는 것은 보안 위험에 대한 단기적인 대책으로 보일 수 있지만, 장기적인 해결책이 아닙니다. ChatGPT와 다른 생성적 인공지능 플랫폼은 비즈니스 프로세스를 간소화하고 작업을 자동화하는 데 강력한 비즈니스 도구로 작용합니다. 이러한 사이트를 차단하면 생산성과 비즈니스 혁신이 제약될 수 있습니다.

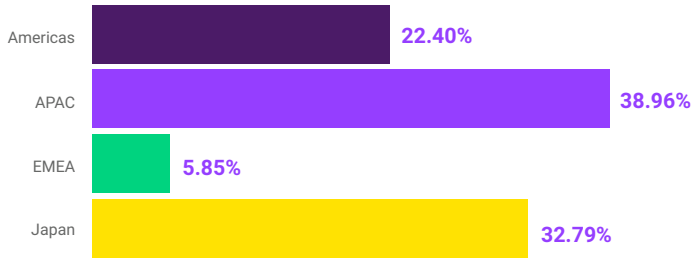
방법론

Menlo Security는 생성적 인공지능이 사이버 보안에 미치는 영향을 이해하기 위해 전 세계 500개 기업을 대상으로 샘플링한 데이터를 기반으로 분석을 수행했습니다. 이 스냅샷은 직원들이 얼마나 자주 생성적 인공지능을 활용하고 있으며, 이로 인해 발생하는 데이터 유실이 어떤 실제적인 영향을 미치고 있는지에 중점을 두고 있습니다.

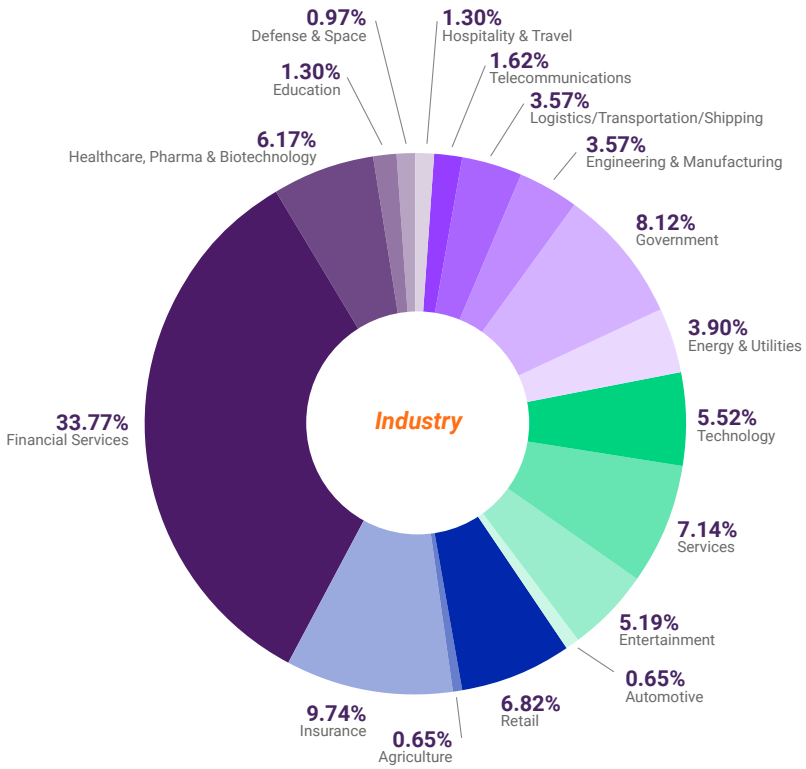
Business Size



Regions



sample size of 500 global organizations



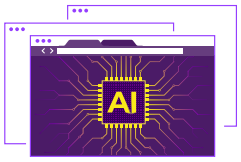
인사이트 1.

직원들은 생성형 AI를 기하급수적으로 활용하고 있습니다

직원들이 생성적 인공지능을 더 많이 사용할수록 기업이 직면하는 잠재적인 위험도 증가합니다 실제로, 0 년 11월부터 03 년 월까지 생성적 인공지능 사용량이 100 증가하고 있으며, 이 수치는 매일 증가하고 있습니다

또한, 이러한 직원들은 생성적 인공지능 사이트를 빈번하게 활용하고 있으며, 각 사용은 잠재적인 데이터 유실의 가능성을 갖고 있습니다.

지난 30일 동안 다음과 같은 사례들이 있었습니다:

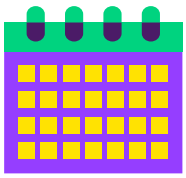


2.5 Million

visits to Generative AI sites for 500 organizations

78,825

Users at these organizations are utilizing Generative AI sites



That means...

32x

a month per user

사용자들이 어떤 사이트를 방문하고 있는지 알아볼까요?

생성적 인공지능 사이트의 목록은 광범위하지만, 이 스냅샷에서는 상위 6개의 인공지능 응용 프로그램에 초점을 맞추기로 결정했습니다:

chat.openai.com; bing.com; bard.google.com; writesonic.com; copy.ai; jasper.ai

상위 3개의 생성적 인공지능 사이트:

OpenAI (ChatGPT)	1977607 visits
Microsoft (Bing)	410850 visits
Google (Bard)	96181 visits

그 중 지난 30일 동안, OpenAI가

>생성형 AI방문을 50% 이상을 차지하였습니다.

인사이트 2:

데이터 유실은 다양한 경로를 통해 발생할 수 있습니다.

직원들이 생성적 인공지능을 사용하는 다양한 방식을 고려하는 것이 중요합니다. 그들이 검색 창에 질문이나 요청을 입력하고 있는지, 다른 소스에서 정보를 복사하여 붙여넣고 있는지, 아니면 요청과 함께 파일을 업로드하고 있는지 등을 고려해야 합니다.

대부분의 사용자는 직접 질문이나 요청을 입력하지만, 파일 업로드와 복사 및 붙여넣기는 데이터 유실의 경로 중에서 가장 큰 영향을 미칠 수 있습니다. 직원들이 이를 통해 많은 양의 민감한 데이터를 훨씬 빠른 속도로 노출시킬 수 있습니다. 몇 가지 예시로는 다음과 같습니다:

- 소스 코드, 고객 목록 또는 로드맵 계획을 복사하여 붙여넣기
- 수백 개의 열을 포함한 스프레드시트를 업로드하기

파일 업로드
—10190 업로드
(지난 30일간)

복사 붙여넣기
—3394 클릭
(지난 30일간)

따라서, 이러한 이벤트의 수는 전체 생성적 인공지능 방문 수에 비해 상대적으로 낮을 수 있지만, 잠재적인 데이터 유실 관점에서는 가장 큰 영향을 미칩니다.

(일부 생성적 인공지능 사이트에는 기본 파일 업로드 옵션이 없을 수도 있습니다. 그러나 사용자는 플러그인을 통해 업로드할 수 있습니다.)

지난 30일간의 파일 업로드 유형별 통계는 다음과 같습니다:

알수없음 (ex. Text files)	1237
PDF	133
Word	48
Excel	22
PowerPoint	21
Script	13
WinEXE	3
ZIP	2

직원들은 다양한 파일 형식을 생성적 인공지능 플랫폼에 업로드하고 있습니다. 가장 빈번한 파일 형식은 알 수 없음 과 PDF입니다.

또한, Excel과 같은 파일 형식은 우발적인 데이터 유실로 이어질 수 있습니다. 직원은 파일이 안전하다고 생각할 수 있지만, 민감하거나 비밀 데이터가 포함된 숨겨진 행이나 열이 있을 수 있습니다.

인사이트 3:

생성형 인공지능으로 인한 데이터 유실 위험은 현실입니다.

데이터 유실 방지(DLP)는 사이버 보안 분야에서 새로운 개념이나 기술은 아니지만, 조직의 통제를 벗어나는 데이터 이동이 일어날 경우의 잠재적인 영향으로 인해 데이터 유실을 제한하거나 차단하는 시도는 모든 산업과 기업 규모에게 계속해서 중요한 문제입니다. 생성적 인공지능을 통해 이러한 우려가 더욱 증가하고 있습니다.

생성적 인공지능 사이트로 전송되는 민감한 데이터의 범위는 이 문제가 공개된 소스 코드 우려를 넘어서는 것을 보여줍니다. 다음 표는 조직에서 정기적으로 발생하는 잠재적인 데이터 유실 사례를 보여주며, 직원들이 생성적 인공지능을 사용함에 따라 의도적이거나 의식하지 못한 상태에서 민감하거나 독점적인 데이터를 입력하거나 업로드하려고 하는 경우를 나타냅니다.

직원들이 생성형 인공지능에 입력하는 데이터의 유형		
개인 식별 정보 (PII)	5.4%	저희는 직원들이 민감하고 비밀스러운 정보를 생성적 인공지능 플랫폼에 입력하려는 시도를 얼마나 자주 하는지 분석했습니다. 지난 일 동안 다음과 같은 범주에 해당하는 데이터를 포함한 DLP 이벤트가 있었습니다: 가장 빈번한 잠재적인 노출은 개인 식별 정보였습니다. 이러한 조직들은 이러한 사례를 차단하기 위한 정책을 시행하고 있습니다.
지적 재산 정보	50.4%	
기밀 정보	24.6%	
의료 정보	2.2%	
금지된 정보	1.5%	
그 외	15.9%	

솔루션

데이터 유실과 생성형 인공지능의 안전한 사용을 고려해야 하는 조직의 관점

기존의 데이터 유실 방지(DLP), CASB 및 기타 내부 위협 대응 솔루션은 새로운 기술의 세부 사항을 다루기에 충분하지 않습니다. 이러한 솔루션은 탐지 및 대응 접근 방식을 사용하여 조직 외부로 흐르는 방대한 양의 트래픽 중에서 키워드나 문구를 찾습니다. 이는 보안 전문가나 제품 소유자가 수동으로 입력해야 하는 과정으로, 매우 수동적인 프로세스이며 모든 것을 탐지하는 것은 거의 불가능합니다. 심지어 솔루션이 데이터 유출을 감지하더라도 이미 너무 늦을 수 있습니다. 정보가 입력되었고 되돌릴 수 있는 실행 취소 버튼이 없습니다. 정보는 생성형 인공지능 플랫폼 내에서 영원히 남아 있으며 계속해서 응답에 영향을 줄 것입니다.

조직은 이러한 생성적 인공지능 플랫폼과 챗봇으로의 정보 입력을 방지해야 하지만, 직원들의 이러한 유용한 도구 사용을 방해하지 않는 방식으로 처리해야 하며, DLP는 유용하지만, 조직은 일관된 해결책에 초점을 두는 대신 다층 접근 방식이 필요합니다.

먼저, 조직은 입력 필드에 붙여넣을 수 있는 내용을 제한할 수 있습니다. 예를 들어 문자 수를 제한하거나 알려진 코드를 차단하는 것입니다. 수천 줄의 소스 코드를 수동으로 입력하는 사람은 없을 것이므로, 붙여넣기 기능을 제한함으로써 이러한 유형의 데이터 유실을 효과적으로 방지할 수 있습니다. 또한, 사용자들이 입력하려고 하는 정보에 대해 한 번 더 생각하도록 유도할 수 있습니다.

가장 중요한 것은 조직이 ChatGPT와 기타 생성적 인공지능 플랫폼과의 상호작용을 웹 브라우저에서 제한하는 것입니다. 클라우드의 원격 브라우저에서 앱 명령을 실행함으로써 사용자와 인터넷 사이에 추가적인 보호 계층을 제공하여 조직이 악의적인 활동(의도적인지 아닌지에 관계없이)이 데이터 유출이 발생하기 전에 중단할 수 있는 기회를 얻을 수 있습니다.

조직은 또한 이벤트 로깅이나 브라우저 녹화 시작과 같은 추가 보안 제어 조치를트리거하는 보안 정책을 적용할 수 있습니다. 이는 해결과 사후 이벤트 분석을 돕기 위한 것입니다. 내부자에 의한 침해 조사는 의도를 증명해야 한다는 점을 기억하는 것이 중요합니다. 이벤트와 브라우저 세션을 기록하는 것은 사용자가 악의적이거나 부주의한지 여부에 대한 가시성과 통찰력을 제공할 수 있습니다.

www.menlosecurity.com/ko-kr/

멘로시큐리티는 웹에서 악성 코드의 위협을 제거함으로써 조직을 사이버 공격으로부터 보호합니다.

멘로의 클라우드 보안 플랫폼은 조직에 대한 위협을 차단하고 데이터 및 애플리케이션 접근을 안전하게 제공합니다. 엘라스틱 격리 코어(Elastic Isolation Core™)는 사용자, 콘텐츠, 애플리케이션 간에 분리를 생성하여 보안, 정책, 그리고 가시성을 제공합니다. 브라우저 내 심층 가시성을 활용한 적응형 정책은 위협이 발생하기 전에 예방하며, 모든 위협(HEAT 포함)을 웹, 이메일, SaaS 애플리케이션 및 개인 애플리케이션에서 매우 탐지 어려운 수준에서 제거합니다.

멘로시큐리티는 HEAT 공격에 대한 취약성을 향상시키기 위해 침투 평가를 제공합니다. 이 평가는 다양한 HEAT 공격을 활용하여 조직의 노출 정도를 파악하고, 최상위 웹 위협에 얼마나 취약한지 확인할 수 있도록 도와줍니다.

조직이 현재 얼마나 위험에 노출되어 있는지 파악하는 것이 중요하며, 더 나아가 이러한 위협을 사전에 방지하는 방법을 찾기 위해 멘로시큐리티에 문의하시기 바랍니다.