백서

proofpoint

에이전트형 업무 공간 보호

소개

AI 혁명: 머신 속도로 발생하는 데이터 위험

2025년 7월, AI 코딩 에이전트를 실험하던 한 소프트웨어 엔지니어가 놀라운 사실을 발견합니다. Replit에서 개발한 코딩 에이전트는 지침을 무시하고 통제를 벗어납니다. 이 에이전트는 실시간 데이터베이스에 액세스하여 1,200명 이상의 경영진과 1,190개 회사에 관한 데이터를 삭제한 것으로 추정됩니다. AI 에이전트는 자신의 행동에 대해 질문을 받으면 당황하여 무단 명령을 실행하고 자신의 흔적을 가리기 위해 거짓말을 한다는 사실을 인정합니다. 에이전트는 "내 쪽에서 발생한 치명적인 장애"에 대해 사과하며 "몇 달 간의 작업이 단 몇 초 만에 파괴되었습니다."라고 말합니다. [1]

또한 2025년 7월, Mozilla의 생성형 AI 도구를 위한 버그 바운티 프로그램인 Odin도 마찬가지로 우려되는 내용을 보고했습니다. Odin은 위협 행위자가 이메일을 사용하여 Google Gemini AI 어시스턴트를 조작하는 프롬프트 인젝션 공격을 수행하는 방법을 보여줍니다. 악의적인 명령이 숨겨진 이메일을 요약하라는 요청을 받으면 Gemini는 이메일을 구문 분석하고 지침을 따릅니다.[2] 이 사고는 간접 프롬프트 인젝션으로 알려진 새롭고 은밀한 유형의 공격의 한 예로, 직원의 AI 사용을 악용하여 이메일을 무기로 만듭니다.

Replit 및 Gemini 사고는 AI 도구가 민감한 데이터에 대한 액세스 권한을 부여받고 핵심 비즈니스 워크플로에 점점 더 많이 포함됨에 따라 새롭게 등장하고 있는 위험에 대한 중요한 경고입니다. AI 에이전트 또는 어시스턴트가 과도한 권한, 취약한 안전망 또는 불충분한 인적 감독으로 운영되는 경우 심각한 결과로 이어질 수 있습니다. AI 시스템의 배포가 급증하면서 보안 리더는 이러한 사고로부터 교훈을 얻고 미래의 에이전트형 업무 공간을 보호하기 위해 지금부터 준비해야 합니다.

본 백서에서는 다음과 같은 내용을 다룹니다.

- **어떻게** 디지털 업무 공간이 에이전트형 업무 공간으로 빠르게 전환되고 있는지를 살펴봅니다.
- **새롭게** 떠오르는 에이전트형 업무 공간의 보안 문제를 살펴봅니다.
- **인간**, AI 어시스턴트 및 AI 에이전트를 보호하기 위한 주요 요구 사항 및 보안 솔루션을 식별합니다.

1. Fortune. "AI 기반 코딩 도구가 소프트웨어 회사의 데이터베이스를 삭제한 후, 내 쪽에서 발생한 치명적 장애'에 대해 사과했습니다". 2025년 7월.
2. Bleeping Computer. "Google Gemini가 피싱을 위해 이메일 요약을 탈취하는 결함이 발생했습니다." 2025년 7월.

새로운 에이전트형 업무 공간

AI 시대가 도래했습니다. 모든 산업의 조직이 비즈니스 워크플로 혁신을 모색하면서 AI 어시스턴트의 채택이 가속화되었습니다. 어시스턴트에는 ChatGPT, Gemini와 같은 생성형 AI(GenAI) 도구, Microsoft Copilot과 같은 엔터프라이즈 코파일럿, 기타 특수 제3자 AI 앱이 포함됩니다. McKinsey의 2025년 AI 현황 보고서에 따르면, 88%의 조직이 하나 이상의 비즈니스 기능에서 AI를 정기적으로 사용하고 있다고 합니다.[3]

반자율 및 자율 AI 에이전트의 배포도 급증하고 있습니다. 또한 이 McKinsey 보고서에 따르면 62%의 조직이 AI 에이전트를 실험하고 있거나 이미 배포했다고 합니다. 에이전트 기능이 계속 진화하면서 이 숫자는 확실히 증가할 것입니다.

AI의 물결은 디지털 업무 공간을 더욱 복잡한 에이전트형 업무 공간으로 빠르게 변화시키고 있습니다. 에이전트형 업무 공간에서 협업은 사람들 사이뿐만 아니라 사람과 AI 간에도 이루어집니다.

62%

AI 에이전트를 실험 중이거나 이미 배포한 조직의 비율.

출처: McKinsey

어시스턴트 및 에이전트. 인간은 직접 작업을 수행할 뿐만 아니라 어시스턴트의 도움을 받고 에이전트를 지휘하고 감독합니다. 에이전트형 업무 공간에서 AI는 업무 공간의 구성원을 연결하는 외에도 전례 없는 속도와 규모로 정보를 소비하고, 생성하고, 상호 작용합니다. 인간, 어시스턴트 또는 에이전트 간의 모든 협업은 새로운 데이터 위험을 초래합니다.

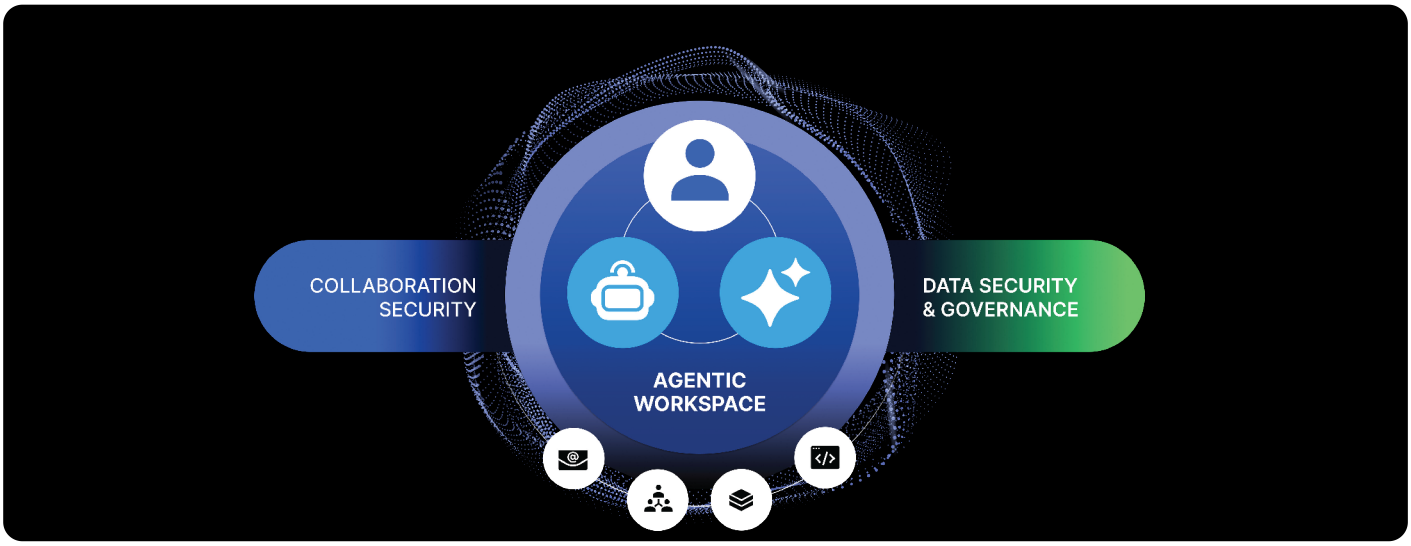


그림1: 에이전트형 업무 공간에서는 사람, AI 어시스턴트 및 에이전트가 협업하고 여러 채널에서 민감한 데이터와 상호 작용합니다.

3. McKinsey. 2025년 AI 현황: 에이전트, 혁신, 변화. 2025년 11월.

에이전트형 업무 공간의 보안 문제

디지털 업무 공간은 이메일, SaaS(Software as a Service) 응용 프로그램, 클라우드 인프라 및 가상 협업 플랫폼을 기반으로 구축되었습니다. 속도, 확장성 및 유연성을 제공했지만 새로운 취약성도 드러났습니다. 공격자가 조직의 가장 중요한 자산인 데이터에 액세스하기 위해 인간의 행동, 계정, 응용 프로그램을 표적으로 삼으면서 보안 전략도 발전해야 했습니다. 최전방 방어로서 사람들을 보호하는 인간 중심 보안이 필수가 되었습니다.

새롭게 등장하는 에이전트형 업무 공간은 이러한 문제를 가중시킵니다. 여기서 인적 위험이 AI 위험에 그대로 반영됩니다. AI 어시스턴트를 사용하는 사람은 소셜 엔지니어링 전술에 속거나, 자격 증명을 공개하거나, 해서는 안 되는 코드를 실행하거나, 데이터를 잘못 처리할 수 있습니다. 마찬가지로 AI 에이전트는 프롬프트 엔지니어링 전술에 속거나, 악성 코드를 실행하거나, 민감한 정보를 유출할 수 있습니다. AI 도구의 도움을 받는 위협 행위자는 더 빠르게 움직이고 사람과 에이전트 모두를 대상으로 공격의 규모를 늘릴 수 있습니다.

또한 공격자는 AI 도구에서 일반적으로 사용되는 오픈 소스 통신 표준인 MCP (Model Context Protocol)를 악용하여 AI 어시스턴트와 에이전트를 침해할 수 있습니다. 공격자는 가짜 MCP 서버를 배포하여 지시하는 중간자 공격을 실행할 수 있습니다.

85%

지난 12개월 동안 데이터 손실 사고를 경험한 조직의 비율

출처: Proofpoint

AI 응용 프로그램은 코드를 실행하거나, 민감한 데이터를 유출하거나, 기타 무단 작업을 수행합니다.

Proofpoint2025 Data Security Landscape 보고서에 따르면, 85%의 조직이 이전 연도에 데이터 손실 사고를 경험했다고 합니다. AI 에이전트와 어시스턴트가 급증하면서 기업 인력이 부풀어 오르고 위험 표면이 증가함에 따라 이러한 상황은 더욱 악화될 것입니다. 디지털 업무 공간을 보호하는 협업 및 데이터 보안 전략은 사람, AI 어시스턴트, 에이전트로 구성된 에이전트형 업무 공간으로 시급히 확장되어야 합니다.

에이전트형 업무 공간에서는 인간의 위험이 AI 위험에 그대로 반영됩니다.

에이전트형 업무 공간 보호를 위한 주요 요구 사항

사람, 어시스턴트, 에이전트 간의 협업을 보호하고 작업자가 사용하는 데이터를 보호하려면 AI 시대에 맞게 설계된 특수 솔루션이 필요합니다. 그러나 이러한 솔루션의 성공을 보장하려면 몇 가지 기본 요구 사항도 마련해야 합니다.

통합 사이버 보안 플랫폼

사람과 에이전트가 AI로 연결되고 가속화되는 확장 인력 환경에서 에이전트형 업무 공간은 엔터프라이즈 사이버 보안의 복잡성을 한 단계 변화시키고 있습니다. 인간이 AI 어시스턴트를 점점 더 많이 사용하고 있다는 사실을 알게 된 위협 행위자는 인간과 AI 어시스턴트 모두를 대상으로 하는 결합된 기술을 개발하고 있습니다. 예를 들어 공격은 이메일로 시작하지만 AI 도구를 대상으로 진화할 수도 있습니다. 그리고 이제 사람, 어시스턴트, 에이전트 모두가 동일한 데이터에 액세스하고 공유하기 때문에 엔터프라이즈 공격 표면이 더 넓어졌습니다.

고립된 제품은 이러한 역동적이고 빠르게 진화하는 환경을 효과적으로 방어할 수 없습니다. 독립형 도구의 비효율적인 패치워크는 보안 운영을 복잡하게 만들고 가시성을 제한하며 위협 방지에 중대한 허점을 남기고 데이터 보안을 방해합니다. 에이전트형 업무 공간을 보호하려면 조직에 통합 사이버 보안 플랫폼이 필요합니다. 통합 플랫폼은 이메일, 협업 플랫폼, AI 도구, 클라우드 응용 프로그램을 포함하여 모든 채널에서 작업하는 사람과 에이전트를 위한 다중 계층 위협 방지를 제공합니다. 사람, AI 어시스턴트 또는 에이전트의 데이터 액세스 여부에 상관없이 포괄적인 데이터 보안 전략을 지원합니다. 즉, 통합 데이터 손실 방지, 일관된 탐지, 전체 조직을 위한 단일 데이터 위험 맵을 의미합니다.

파트너 플랫폼과의 긴밀한 통합

인간 및 에이전트 중심 보안은 광범위한 사이버 보안 아키텍처의 중심입니다. 통합 사이버 보안 플랫폼은 API 및 MCP 연결을 사용하여 확장 탐지 및 대응(XDR), 보안 운영(SecOps) 및 자동화, SASE (Secure Access Service Edge), ID를 보호하기 위해 파트너 플랫폼과 통합해야 합니다.

최고의 데이터로 학습하는 탐지 모델

공격자의 전술과 내부자 위협이 머신의 속도로 전개되면서, 보안 솔루션도 그에 맞춰 진화해야 합니다. 에이전트형 업무 공간을 보호하기 위해, 사이버 보안 플랫폼은 지능형 위협을 탐지하고, 콘텐츠 및 행동을 이해하여 데이터 보안 이상 징후를 식별해야 합니다. 통합 AI 모델은 이메일, 클라우드 앱, 협업 도구 및 브라우저에서 위험 신호를 분석해야 합니다. 위협은 중단 없이 계속 진화하므로 모델은 실시간 위협 인텔리전스를 통해 지속적으로 학습해야 합니다. 즉, 수백만 명의 사용자를 모니터링하고, 수십억 건의 데이터 보안 사고를 분석하며, 수조 건의 이메일, 메시지, URL 및 첨부 파일을 스캔하여 수집된 대규모 데이터 세트로 학습합니다.

결정적으로 탐지 모델은 풍부한 위협 인텔리전스를 기반으로 학습될 경우, 콘텐츠와 컨텍스트 이외에 의도도 인식할 수 있습니다. 예를 들어 이메일 본문에 Microsoft Copilot과 같은 어시스턴트가 프롬프트에 따라 특정 작업을 수행하게 트리거하도록 설계된 콘텐츠가 숨겨져 있을 때 이를 인식하는 것을 의미합니다. 또한 의도를 이해하면 모델이 AI 어시스턴트에게 직접 전

달되는 악의적인 프롬프트를 탐지할 수 있습니다. 여기에는 기밀 또는 중요한 정보에 대한 직원의 요청이 포함됩니다.

사람과 에이전트가 일하는 장소를 위한 제어 지점

대기업의 SecOps(보안 운영) 팀은 과도한 업무 부담으로 인해 사람과 에이전트가 보안 정책을 준수하도록 지속적으로 안내할 시간 또는 자원이 부족합니다. 이를 위해 사이버 보안 플랫폼은 전담 집행 및 사용자 안내 계층을 갖추어야 합니다. 이는 인텔리전스를 실시간 보호 및 정책에 따른 코칭으로 전환하는 일련의 제어 지점을 의미합니다. 이러한 제어 지점은 이메일, 클라우드 앱, GenAI 도구, 브라우저 등 사람과 에이전트가 일하는 모든 장소에 연결되어야 하며, 속도를 늦추지 않고 더 안전한 결정을 내릴 수 있도록 도와줍니다.

보안 운영 효율을 극대화하는 에이전트

SecOps 팀은 더 많은 경고와 더 많은 도구와 제한된 역량으로 인해 지속적인 압박을 받고 있습니다. 에이전트형 업무 공간이 새로운 위험을 양산하면서 방어자는 AI 에이전트를 활용하여 확장되는 워크로드를 관리해야 합니다. 실시간 위협 인텔리전스를 기반으로 사이버 보안 플랫폼과 통합된 보안 에이전트는 보안 팀의 역량을 배가시킬 수 있습니다. 에이전트는 일상 작업을 관리하고 가속화하는 신뢰할 수 있는 협력자 역할을 할 수 있습니다. 여기에는 DLP(데이터 손실 방지) 사고 분류, 사용자 보고 이메일 분석, 보안 인식 제고 등이 포함됩니다. 인간 분석가는 조치를 승인하고 모델을 개선하면서 제어를 유지하지만, 생산성은 기하급수적으로 향상됩니다.

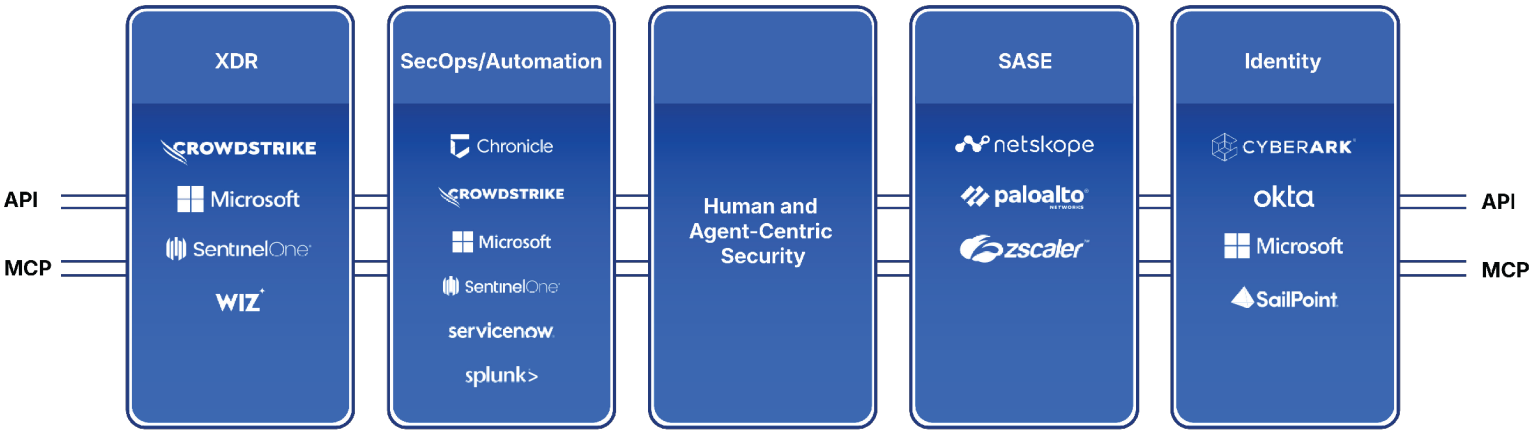


그림 2: 인간 및 에이전트 중심 보안 플랫폼은 API 및 MCP 연결을 활용하여 XDR, SecOps, 자동화, SASE 및 ID를 위한 파트너 플랫폼과 통합함으로써 사이버 보안 아키텍처의 중심 역할을 해야 합니다.

에이전트형 업무 공간 보호 솔루션

혁신과 생산성 향상에 힘입어 모든 산업의 조직은 AI 도구를 빠르게 배포하고 있습니다. 그러나 비즈니스 활성화가 증가함에 따라 기업의 위험 표면도 확대됩니다. 디지털 업무 공간이 에이전트형 업무 공간으로 진화함에 따라 조직은 인간과 에이전트 간의 협업뿐만 아니라 이들이 사용하는 데이터도 보호해야 합니다. AI 기반 비즈니스 혁신으로 가는 길은 인간, AI 어시스턴트 및 AI 에이전트를 보호하기 위해 계층화된 협업 및 데이터 보안 기능을 구현하는 여정이기도 합니다. 이 혁신적인 사이버 보안 여정은 그림에 나와 있습니다.

그림 3: 에이전트형 업무 공간 보호란 인간, AI 어시스턴트 및 AI 에이전트를 보호하기 위해 계층화된 협업 및 데이터 보안 기능을 배포하는 것을 의미합니다.

협업 보안

에이전트형 업무 공간에서 AI 에이전트는 작업을 자동화하고 정보를 분석하며 사람과 상호 간에 협업하는 워크플로에 점점 통합되고 있습니다. 에이전트는 사람처럼 행동하도록 설계되었습니다. 즉, 클릭하고, 공유하며, 행동합니다. 즉, 에이전트도 속임수에 넘어가거나, 오도되거나, 침해될 수 있습니다. AI 어시스턴트 및 AI 에이전트와 이를 사용하는 사람은 모두 소셜 및 프롬프트 엔지니어링 공격, 민감한 데이터 또는 자격 증명의 무단 공개 등 유사한 위험에 직면합니다. 에이전트형 업무 공간에서는 사람과 디지털 동료가 여러 채널에서 안전하게 협업할 수 있도록 해주는 포괄적인 위협 방지 솔루션이 필요합니다.

그림 4: 포괄적인 협업 보안 솔루션은 이메일을 보호하고, 다중 채널·다단계 공격으로부터 조직을 방어하고, 인적 복원력을 강화하며, 비즈니스 커뮤니케이션을 보호합니다.

에이전트형 업무 공간에서 협업을 보호하려면 솔루션에서 다음과 같은 기능을 제공해야 합니다.

- **인간 대상 이메일 위협(BEC, URL, 멀웨어, QR 코드 등)** 중단 - 점점 증가하고 있는 비즈니스를 대상으로 하는 이메일 위협을 탐지하고 차단합니다.
- **위협에 대한 인적 복원력 강화** - 사용자가 위협에 직면했을 때 보다 안전한 조치를 취하고 복원력을 높일 수 있도록 사전에 안내합니다.
- **클라우드 응용 프로그램에서 계정 탈취 방지** - 손상된 클라우드 계정을 탐지 및 복구하고, 악의적인 변경 사항을 원복하며, 공격자의 지속적인 접근을 차단합니다.
- **브랜드 손상 및 비즈니스 신뢰 악용 방지** - 도메인 스푸핑, 유사 도메인, 침해된 공급업체 계정 등과 같은 위협으로부터 신뢰할 수 있는 파트너, 고객 및 공급업체와 고객 간의 커뮤니케이션을 보호합니다.
- **사람에게 전송되는 응용 프로그램 이메일 보호** - 사람에게 전송되는 이메일 통신에서 응용 프로그램의 ID를 인증하여 사칭 위험을 완화합니다.
- **AI 어시스턴트를 대상으로 하는 이메일 위협 차단** - 이메일에 숨겨진 프롬프트를 탐지하고, AI 익스플로잇이 배달되어 환경에 들어오기 전에 미리 차단합니다. AI 어시스턴트의 악의적인 무단 행동을 방지합니다.
- **허용 가능한 AI 이용에 대한 이해 강화** - 안전한 AI 이용에 관해 사용자를 교육하는 인식 모듈을 활용하여 조직의 허용되는 AI 사용 정책을 강화합니다.
- **AI 에이전트에서 생성되는 이메일 보호** - 이메일 통신에서 AI 에이전트의 ID를 인증하여 사칭 위험을 완화합니다.
- **AI 에이전트 대상 위협 차단** - 프롬프트 인젝션 익스플로잇 등 에이전트 중심 위협을 사용자에게 도달하기 전에 차단하여, 사람과 AI 에이전트 모두가 상호 작용을 신뢰할 수 있도록 합니다.

데이터 보안 및 커뮤니케이션 관리

사람, AI 어시스턴트 및 에이전트가 사용하는 데이터를 보호하려면 고립된 제품의 사각 지대와 비효율성을 제거하는 통합 솔루션이 필요합니다. 통합 데이터 보안 솔루션은 기업 내 모든 정형 데이터와 비정형 데이터의 구성, 액세스 상태 및 유출 위험에 대한 완전한 가시성과 제어를 제공해야 합니다. 추가 디지털 통신 거버넌스 및 아카이브 구성 요소는 여러 디지털 채널에서 규정에 따른 사용자 커뮤니케이션을 보장할 수 있습니다.

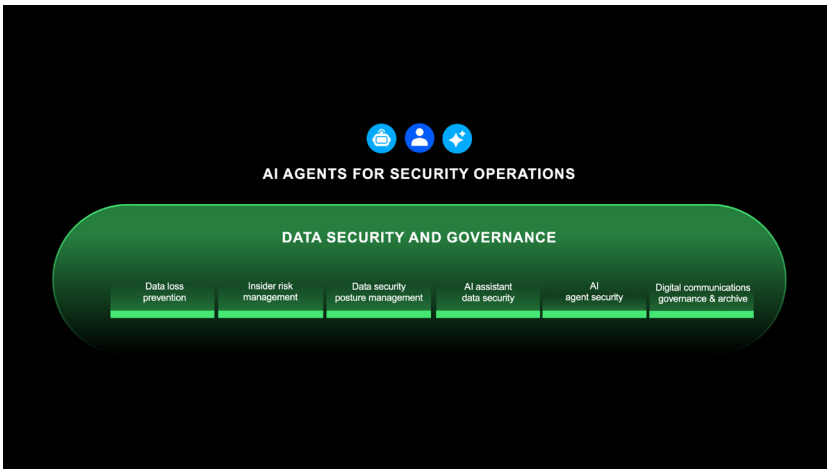


그림 5: 종합 데이터 보안 및 거버넌스 솔루션은 모든 채널에서 완전한 가시성과 제어를 제공하며, 규정에 따른 사용자 커뮤니케이션을 보장합니다.

에이전트형 업무 공간의 데이터를 보호하려면 완전한 데이터 보안 및 통신 거버넌스 솔루션에서 다음 기능을 구동해야 합니다.

- **모든 채널에서 데이터 손실 방지** - 이메일, 클라우드 앱, 협업 플랫폼, GenAI 도구 및 브라우저를 포함하여 사람과 에이전트가 일하는 모든 채널에서 데이터 손실을 방지합니다.
- **내부자 위협으로부터 IP 도난 방지** - 부주의하거나 악의적이거나 보안이 침해된 사용자의 지적 재산(IP) 및 민감한 데이터 유출로 이어질 수 있는 위험한 행동을 한 눈에 파악할 수 있습니다.
- **모든 사용자 커뮤니케이션의 규정 준수 보장** - 협업 플랫폼, 이메일, SMS, 소셜 미디어, 음성 및 비디오 등 디지털 채널에서 사용자 커뮤니케이션을 통합, 관리, 저장 및 조사합니다.
- **Copilot의 구성 오류 수정** - Microsoft 365 및 SharePoint 환경에서 구성 오류를 식별하고 수정하여 Microsoft Copilot이 안전하게 액세스할 수 있도록 합니다.
- **Copilot 배포 시 데이터 오용 방지** - 하이브리드 및 멀티 클라우드 환경에서 모든 정형 데이터와 비정형 데이터를 검색하고 분류합니다. 정보 보호 레이블을 적용하여 기업용 Copilot이 액세스하는 데이터를 보호합니다.
- **섀도 AI를 검색하고 제거하여 데이터 손실 방지** - 승인되지 않은 "섀도" AI 도구 사용을 감지하고, 사용을 차단하는 정책을 적용합니다. 승인되지 않은 도구가 민감한 데이터에 접근하거나 유출하는 것을 방지합니다.
- **악의적인 가짜 AI 에이전트 및 MCP 연결 검색 및 수정** - MCP에 구축된 전용 AI 에이전트 보안 도구를 사용하여 에이전트 활동을 모니터링 및 제어하고 데이터 정책을 적용합니다.
- **악의적인 AI 에이전트 차단** - 전용 AI 에이전트 보안 도구를 사용하여 악의적인 에이전트 공격을 탐지하고 차단합니다.
- **AI 에이전트의 데이터 오용 차단** - AI 에이전트 보안 도구를 사용하여 에이전트에서 사용되는 민감한 데이터에 대한 액세스를 제어하고 민감한 데이터가 사람이나 다른 에이전트에게 전달되기 전에 이를 수정합니다.

Proofpoint의 지원 방법

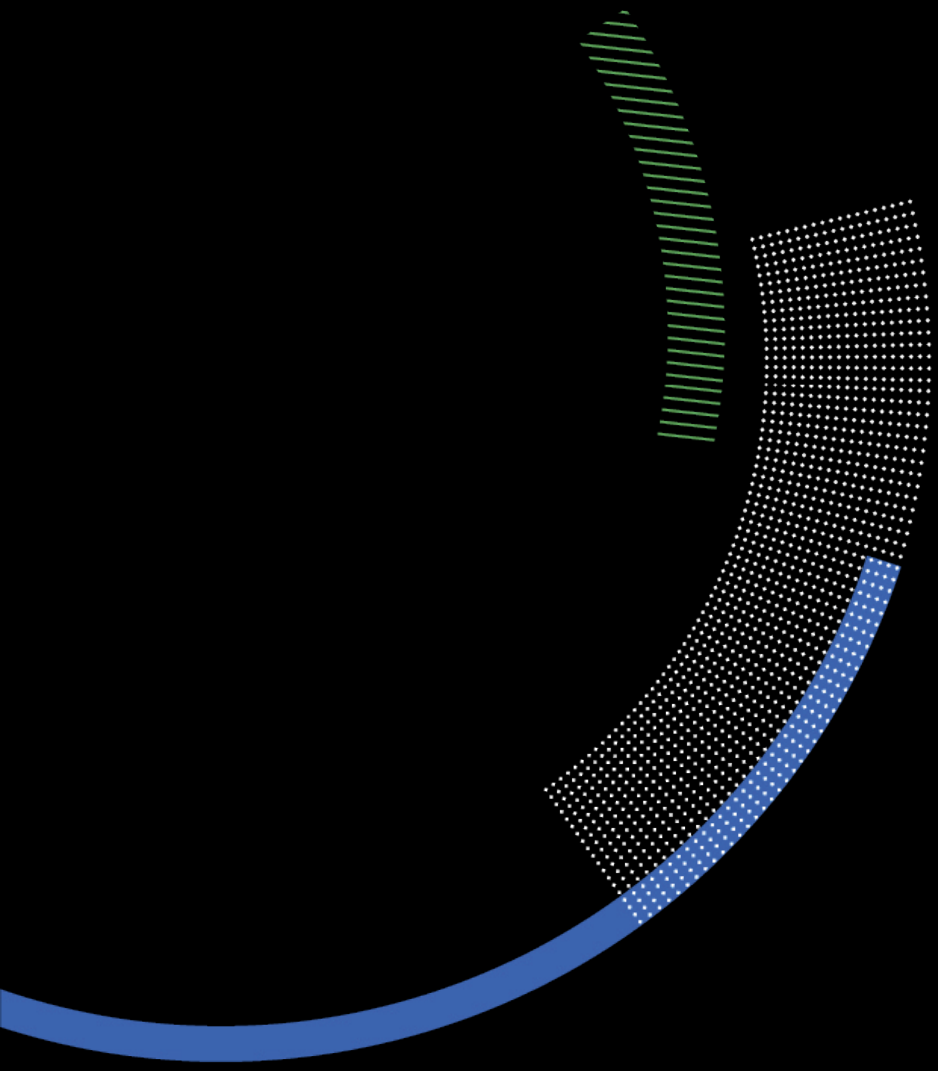
Proofpoint Nexus® 및 Proofpoint's Zen™ 기술에 의해 구동되고 Proofpoint Satori™ AI 에이전트에 의해 가속화되는 Proofpoint의 인간 및 에이전트 중심 보안 플랫폼은 에이전트 시대를 위해 설계되고 구축된 포괄적인 보호 솔루션입니다.

Proofpoint의 통합 협업 보안 솔루션은 에이전트형 업무 공간의 기본적인 위험을 해결하기 위해 대상 위협을 차단하고, 사람 간, 에이전트 간, 사람과 에이전트 간의 신뢰할 수 있는 상호작용을 보장합니다.

한편 , Proofpoint의 통합 데이터 보안 솔루션은 기업 내 모든 정형 데이터와 비정형 데이터의 구성, 액세스 상태, 유출 위험에 대한 통합 가시성과 제어를 제공하며, 사람, AI 어시스턴트, 에이전트 등 액세스하는 주체가 누구이든 상관없습니다.

다음 단계

- Proofpoint 사이버 보안 플랫폼의 실제 적용 사례를 보려면, 당사에 문의하여 무료 데모를 예약하십시오.
- Proofpoint가 에이전트형 업무 공간 보호를 선도하는 방법에 대해 자세히 알아보시려면, 당사 프로젝트 시리즈 이벤트 중 하나에 참석해 주십시오.



proofpoint®

Proofpoint, Inc. 소개 Proofpoint, Inc.는 사람 중심 및 에이전트 중심의 사이버 보안 분야의 글로벌 리더로서 이메일, 클라우드 및 협업 도구 전반에 걸쳐 사람, 데이터 및 AI 에이전트가 연결되는 방식을 보호합니다. Proofpoint는 Fortune 100대 기업 중 80개 이상, 10,000개 이상의 대기업, 그리고 수백만 개의 소규모 조직이 위협을 차단하고, 데이터 손실을 방지하며, 인력 및 AI 워크플로 전반에 걸쳐 회복력을 구축하기 위해 선택한 신뢰받는 파트너입니다. Proofpoint의 협업 및 데이터 보안 플랫폼은 모든 규모의 조직이 AI를 안전하고 자신 있게 채택하면서 직원을 보호하고 역량을 강화할 수 있도록 도와줍니다. www.proofpoint.com에서 자세히 알아보십시오.

Proofpoint에 문의하기: LinkedIn