스스로 행동하는 AI, 보안은 어떻게 달라져야 하나

글로벌 IT 리서치 기관 가트너(Gartner)는 지난 6월 1일부터 3일까지 미국 메릴랜드주 내셔널 하버 게일로드 컨벤션 센터에서 'Gartner Security & Risk Management Summit 2026(이하 가트너 서밋 2026)'를 개최했다. 이번 컨퍼런스에서는 가트너 애널리스트 포함 여러 연사들이 AI를 중심으로 다양한 인사이트를 공유했다.

이번 글에서는 가트너 서밋 2026에서 안랩이 직접 참석해 확인한 주요 보안 트렌드들을 다룬다.

이번 행사는 전 세계 보안 관계자 약 4000명이 참가했으며, 전시 업체 수 250여 개, 세션 수 470여 개를 기록했다. 행사의 중심에는 단연 에이전틱 AI(Agentic AI)가 있었고, AI를 활용한 공격 동향과 보안 방안, AI 에이전트 보안 전략 등에 대한 다양한 세션이 진행되었다. 이 외에도 CTEM(Continuous Threat Exposure Management), SASE(Secure Access Service Edge), CPS(Cyber-Physical System) 보안 등에 대한 논의도 이뤄졌다.

기조연설: AI 시대 보안, 차단에서 회복력으로

가트너 서밋 2026의 오프닝 키노트는 'Seize the Moment'라는 타이틀로, 가트너 Distinguished VP 애널리스트 레이 맥뮬런(Leigh McMullen)이 맡았다.

[그림 1] 레이 맥뮬런, 가트너 Distinguished VP 애널리스트 (출처: 가트너)

먼저, 레이 맥뮬런은 은 AI 기반 공격, 이른바 다크 AI(Dark AI)가 보안 업계의 핵심 화두로 부상하고 있다고 진단했다. AI는 공격자의 생산성을 크게 높이고 있으며, 기존 수주 또는 수개월이 걸리던 공격 준비 과정을 수분 단위로 단축시키고 있다. 그리고, AI를 활

용한 취약점 탐색과 AI 무기화 속도는 앞으로 더욱 빨라질 가능성이 크다고 예측했다.

그러나 이번 기조 연설의 핵심 메시지는 비관론이 아니었다. 레이 맥뮬런은 AI 시대의 경쟁을 공격 기술 경쟁이 아니라 자동화 경쟁(Automation Race)으로 바라봐야 한다고 제시했다. AI는 공격자뿐 아니라 방어자에게도 동일하게 주어진 기술이며, 방어자는 더 많은 데이터, 자원 및 협업 기반을 갖고 있다. 위협 인텔리전스, 탐지 정책, 대응 플레이북을 공유하고 자동화한다면 개별 공격자보다 더 효과적으로 방어를 구현할 수 있다는 관점이다.

또한, AI 에이전트가 확산되면서, 아이덴티티(Identity) 보안 전략에도 변화가 일어난다고 설명했다. 대규모 범용 에이전트가 아닌 작은 단위의 에이전트를 설계해 최소 권한을 적용하며, 에이전트 간 통신과 행동을 지속적으로 모니터링하는 체계가 필요하다는 것이다. 특히, 기계 아이덴티티(Machine Identity)는 휴먼 아이덴티티(Human Identity) 대비 최대 80배 이상 많을 수 있어, AI 에이전트 확산과 함께 관리 복잡성이 더욱 커질 전망이다. 향후 IAM(Identity and Access Management)은 사용자 인증 중심에서 에이전트와 기계 아이덴티티까지 포괄하는 방향으로 확장된다는 것이 그의 설명이다.

레이 맥뮬런은 위와 같은 변화와 함께 사이버 보안의 목표도 차단(Prevention)에서 회복력(Resilience)으로 이동하고 있다고 말했다. 모든 공격을 차단하는 것은 현실적이지 않으며, 이제 비즈니스 영향 최소화, 복구 시간 단축, 서비스 연속성 확보가 더 중요하다는 것이다. 이를 위해서는 비즈니스 프로세스별 영향 임계치(Impact Threshold)를 정의하고, 복구 리허설(Recovery Rehearsal)과 실제 충격을 주어 회복력을 테스트하는 카오스 엔지니어링(Chaos Engineering)을 일상적인 운영 프로세스로 정착시켜야 한다.

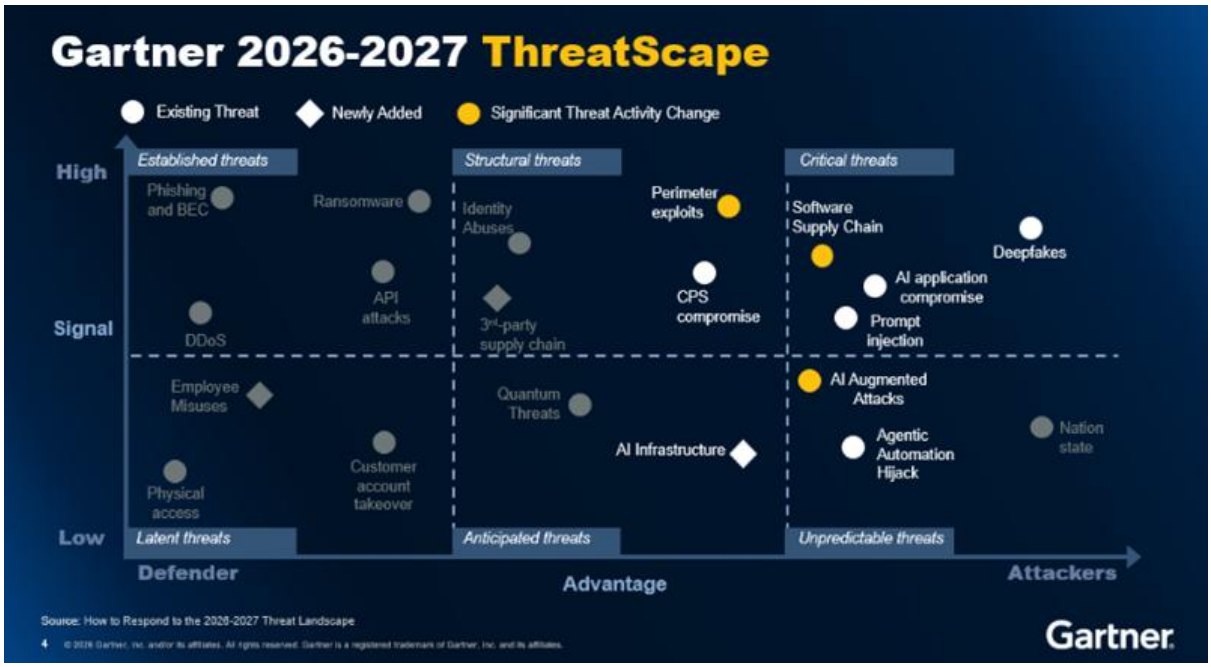
레이 맥뮬런은 AI 시대의 보안 조직은 더 많은 보안 툴을 구매하는 것만으로 경쟁력을 확보할 수 없으며, SOC 운영, 취약점 관리, 복구 절차, 테스트 환경, 에이전트 기반 업무 프로세스까지 자동화를 직접 설계하고 검증해야 한다고 강조했다. 자동화는 단순 효율화 수단이 아니라, AI 시대 보안 조직의 핵심 생존 전략이다.

에이전틱 AI를 활용한 공격과 보안 전략

AI는 사이버 공격의 방식을 완전히 새롭게 바꾸고 있다기보다, 기존 공격의 속도와 규모, 성공 가능성을 높이는 방향으로 작용하고 있다. 이번 컨퍼런스의 여러 세션에서는 랜섬웨어, 취약점 악용, 소셜 엔지니어링, 보안 자동화 관점에서 AI가 위협 환경과 보안 전략을 어떻게 바꾸고 있는지 다뤘다. 공통된 메시지는 명확했다. AI는 공격자에게 빠른 정찰, 정교한 피싱, 쉬운 악성코드 개발, 높은 수준의 자동화를 제공하지만, 동시에 방어자에게도 위협 인텔리전스 분석, 취약점 우선순위 설정, 보안 운영 자동화, 복원력 강화를 위한 새로운 기회를 제공한다.

존 왓츠(John Watts) 가트너 VP 애널리스트는 'Outlook for Cybersecurity Threats:

Prioritizing With Gartner's 2026-2027 Threatscape' 세션에서 2026 ~ 2027년 위협 환경의 핵심 변화로 AI 기반 익스플로잇(Exploit) 생성과 공격 규모 확대를 제시했다. 존 왓츠는 미토스(Mythos), 글래스윙(Glasswing)과 같은 AI 기반 취약점 분석 및 익스플로잇 생성 기술은 취약점 발견부터 공격 코드 생성, 실행까지의 시간을 크게 단축시킬 수 있다고 설명했다.



[그림 2] 2026 ~ 2027 위협 환경 (출처: 가트너)

특히, AI 시대에는 피싱보다 취약점 악용이 초기 침투 경로로 더 중요해질 가능성이 크다고 강조했다. 공격자는 신규 취약점이 공개되면 인터넷 전체를 자동 스캔하고, 취약 시스템을 찾아 대규모 공격을 전개한다. 과거에는 특정 목표를 정하고 취약점을 활용했다면, 향후에는 취약점을 중심으로 가능한 모든 대상을 공격하는 것이다. 이에 대해, 패치 주기를 단축하는 것만으로는 충분하지 않다. 노출 자산을 파악하고, 실제 악용 가능성을 기준으로 우선순위를 정해 공격이 시작됐을 때 조기에 탐지하고 차단할 수 있는 탐지 & 대응 역량이 필요하다.

[그림 3] 존 왓츠(John Watts) 가트너 VP 애널리스트 (출처: 가트너)

존 왓츠는 이러한 위협 환경의 변화가 CTEM(Continuous Threat Exposure Management)의 중요성을 높인다고 말했다. CTEM은 단순히 취약점 목록을 관리하는 것을 넘어, 조직의 실제 노출 상태와 공격 가능성을 지속적으로 확인하고 우선순위를 재조정하는 운영 모델이다. 위협이 우리 조직과 관련이 있는지, 즉시 대응이 필요한지, 실제 공격에 활용되고 있는지, 대응 효과를 측정할 수 있는지 등을 기준으로 보안 역량을 집중해야 한다. 이에 대해 존 왓츠는 AI 시대 보안은 더 많은 알림을 처리하는게 아니라, 수 많은 노이즈 속에서 의미 있는 시그널을 찾아내는 문제에 가깝다고 강조했다.

한편, 기조연설을 맡았던 레이 맥뮬런은 'Use AI Like a Threat Actor' and Other Strategies for AI in Cyber Defense 세션을 통해 보안 조직들이 공격자의 AI 활용 방식을 반대로 적용해야 한다는 관점을 제시했다. 공격자는 AI를 타깃 선별, 공격 은닉, 반복 작업 자동화 등에 사용하고 있다. 그리고, 방어자도 같은 방식을 활용할 수 있다.

예를 들어 AI를 이용해 EDR 텔레메트리(Telemetry)를 기반으로 새로운 방화벽 룰을 생성하고, 검증 후 배포하는 방식이 가능하다. AI를 모든 판단의 주체로 두는 것이 아니라, 특정 문제를 해결하는 스크립트와 자동화 코드를 생성하도록 활용하는 접근이다.

위협 인텔리전스 영역에서도 AI의 활용 가능성이 크다. CVE 피드, 다크웹 정보, 산업별 위협 정보 등을 수집해 RAG(Retrieval-Augmented Generation) 파이프라인에 연결하면 조직 맞춤형 위협 인텔리전스 리포트를 생성할 수 있다. 특정 산업을 주로 공격하는 위협 행위자의 TTP와 CVE 유형을 분석해 패치 우선순위 설정에 적용할 수도 있다.

레이 맥뮬런은 AI 기반 공격과 방어의 핵심은 자동화의 방향성에 있다고 조언했다. 공격

자는 AI를 이용해 반복 작업을 빠르게 수행하고, 방어자는 이를 역이용해 위협 인텔리전스, CTEM, SOC 우선순위 설정 등을 자동화해야 한다.

또한, AI 자체를 목표로 삼아서는 안되며, 중요한 것은 AI 로드맵이 아니라 사이버보안 로드맵이라 강조했다. 패치 속도 개선, 탐지 품질 향상, 취약점 관리 효율화, 복원력 강화라는 기존 보안 목표를 먼저 정의하고, AI는 이를 달성하기 위한 수단으로 활용해야 한다는 것이다.

늘어나는 AI 에이전트, 보안은 어떻게?

최근, AI 에이전트가 빠르게 확산되고 있다. 업무 효율성 개선에 큰 혁신을 불러올 것으로 기대가 되지만, 동시에 기존 보안 모델로는 다루기 어려운 새로운 리스크를 만들고 있다.

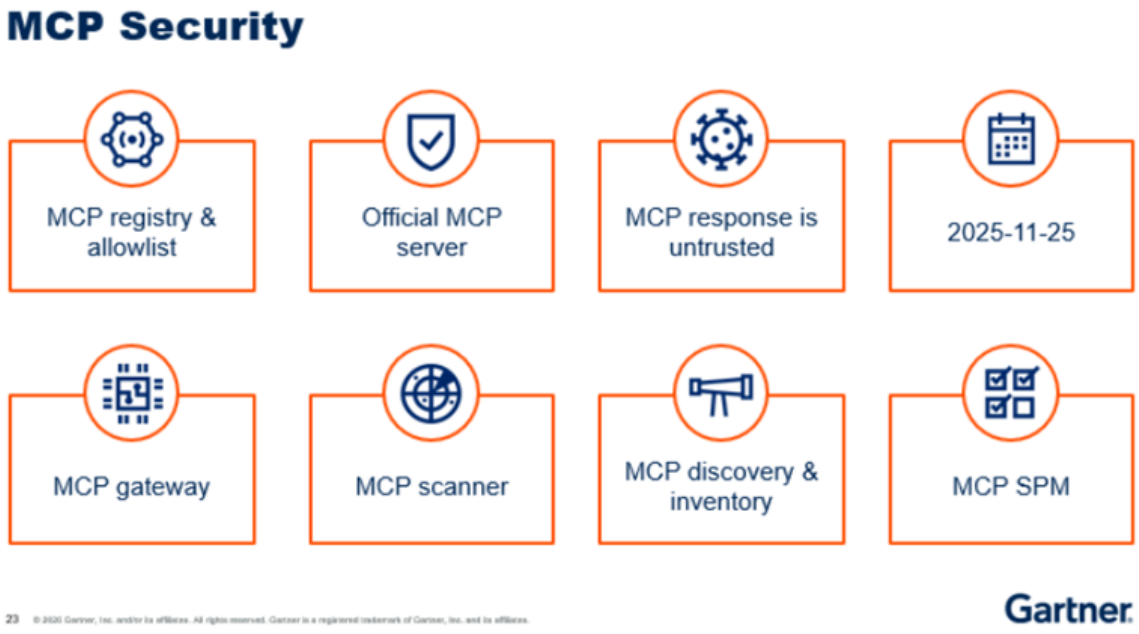
데니스 주(Dennis Xu) 가트너 리서치 부사장(Research Vice President)는 'Technical Insights: Secure AI Agents Before They Go Rogue'에서 AI 에이전트가 단순 자동화 도구가 아니라 높은 권한과 자율성(Autonomy)을 가진 실행 주체라는 점에 주목했다. 그는 AI 에이전트가 사용자 메일함을 삭제하거나, 코딩 에이전트가 운영 데이터베이스를 9초 만에 삭제한 사례가 실제로 발생했음을 언급하며, AI 에이전트가 현실의 운영 리스크가 되었다고 말했다.

[그림 4] 데니스 주(Dennis Xu) 가트너 리서치 부사장 (출처: 가트너)

데니스 주는 AI 에이전트가 크게 두 가지 측면에서 리스크를 발생시킨다고 했다. 하나는 민감 데이터와 시스템에 접근할 수 있는 권한이고, 다른 하나는 아직 완전하지 않은 추론 능력이다. 현재의 AI 모델은 프롬프트 인젝션(Prompt Injection)과 제일 브레이크

(Jailbreak)를 100% 차단할 수 없다. 따라서, AI 에이전트는 데이터 접근 권한, 이메일 발송 권한, 시스템 제어 권한을 갖고 의도하지 않은 행동을 수용하거나 공격자에게 악용될 수 있다. 특히, 높은 자율성을 부여 받은 AI 에이전트는 실행 시점에 어떤 도구를 호출하고 어떤 데이터에 접근할지 스스로 결정하기 때문에, 기존 애플리케이션 보안 방식만으로는 통제가 어렵다.

이에 대해, 데니스 주는 AI 에이전트 보안을 식별(Discovery)과 형상 관리(Posture Management)부터 시작해야 한다고 조언했다. 조직 내 어떤 AI 에이전트가 존재하는지, 어떤 플랫폼에서 개발됐는지, 어떤 역할을 수행하는지, 어떤 툴과 MCP(Model Context Protocol)에 연결되어 있는지, 어떤 데이트에 접근할 수 있는지를 지속적으로 파악해야 한다는 것이다. 특히, AI 에이전트는 배포 후에도 메모리, MCP, 연결 툴 등이이 계속 변경될 수 있기 때문에 일회성 평가가 아니라 지속적인 형상 관리가 필요하다.



[그림 5] MCP 보안 방안 (출처: 가트너)

이어서, AI 에이전트의 리스크를 테스트하고 검증하는 레드 티밍(Red Teaming)도 중요한 통제 수단이라고 말했다. 현재 기술 수준에서, 프롬프트 인젝션, 제일 브레이킹 등 모든 시나리오를 자동으로 검증하는 것은 한계가 있기 때문에, 레드 티밍의 목적을 완전한 방어가 아니라 리스크 우선순위 식별 및 대응 역량 강화에 둬야 한다고 조언했다. 리스크 검증은 AI 에이전트의 입력과 출력만 보는 방식으로는 충분하지 않고, 실제 어떤 도구를 호출하고 어떤 권한을 사용하는지까지 확인이 필요하다.

그 다음으로 필요한 역량으로 런타임 모니터링(Runtime Monitoring)과 의도 기반 접근 제어(Intent-Based Access Control)를 소개했다.

- 런타임 모니터링: AI 에이전트 실행 중 프롬프트 인젝션을 탐지하고, 고위험 명령과 비정상 행동, 위험한 데이터 등을 감시
- 의도 기반 접근 제어: 권한을 고정적으로 부여하기보다 실제 수행하려는 작업의 의도(Intent)에 따라 동적으로 조정. 예를 들어 사용자가 일정 조회를 요청했다면 Calendar Read 권한만 부여하고 Calendar Write 권한은 제거.

결국 AI 에이전트 보안의 핵심은 에이전트를 신뢰하는 것이 아니라, 권한과 행동을 지속적으로 검증하고 통제하는 데 있다.

CTEM과 TDIR의 통합으로 보는 차세대 보안 모델

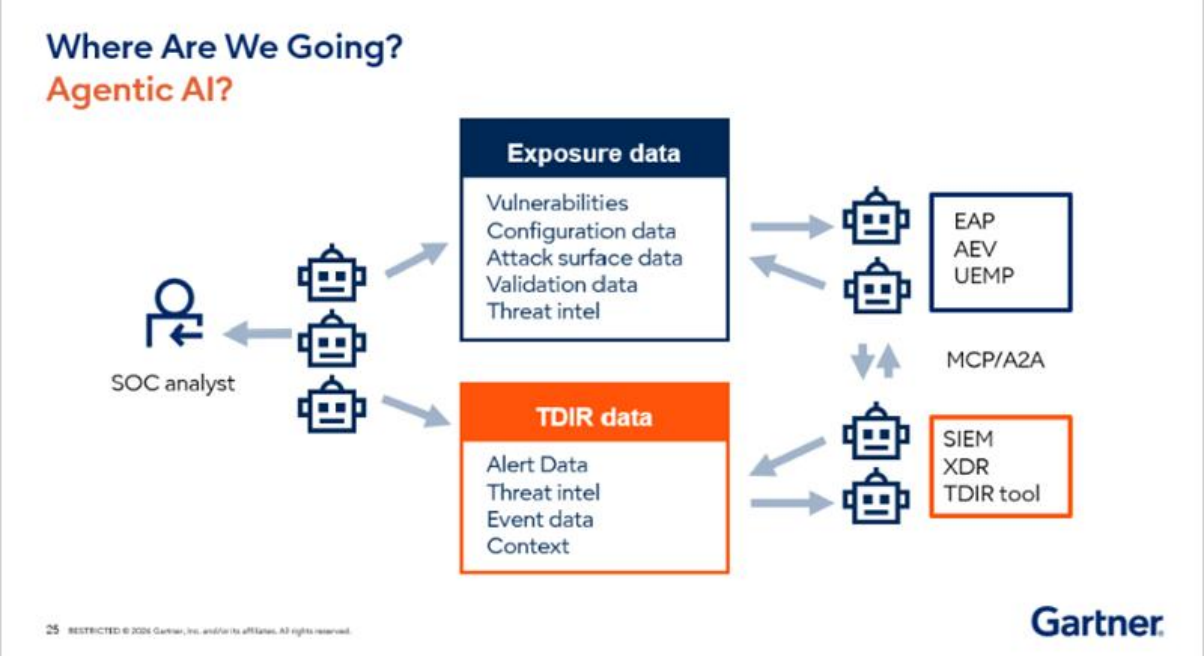
이번 컨퍼런스에서는 에이전틱 AI와 함께 기존 보안 기술 영역의 진화도 함께 다뤄졌다. 공통된 방향은 단순한 보안 툴의 확장이 아니라, 더 많은 컨텍스트를 연결하고 실제 리스크에 기반해 의사 결정을 내리는 구조로 전환하는 것이다.

피트 쇼어드(Pete Shoard) 가트너 VP 애널리스트는 'Breaking Boundaries: Uniting Exposure Management & Threat Detection & Incident Response' 세션에서 노출 관리(Exposure Management)와 TDIR(Threat Detection, Investigation & Response) 통합의 필요성을 강조했다.

[그림 6] 피트 쇼어드(Pete Shoard) 가트너 VP 애널리스트 (출처: 가트너)

현재 대부분의 조직에서 노출 관리를 담당하는 팀은 공격 표면 분석, 취약점 관리, 리스크 우선순위 설정을 담당하고, SOC는 SIEM, XDR 및 위협 인텔리전스 기반 탐지와 대응을 수행한다.

피트 쇼어드는 이러한 분리가 보안 운영의 한계로 이어진다고 지적했다. SOC는 이메일, 엔드포인트, 네트워크, 클라우드 전반의 로그를 수집하고 분석할 수 있지만, 공격 표면, 자산 중요도, 공격 경로, 취약점, 구성 정보, 접근 권한과 같은 컨텍스트는 충분히 활용하지 못한다. 반대로 노출 관리 부서는 취약점, 공격 표면 및 비즈니스 컨텍스트를 활용해 실제 리스크 우선순위를 설정할 수 있지만, 실제 사고 대응 데이터, 탐지 규칙, SOC 인사이트와 충분히 연결되지 못한다는 것이다.



[그림 7] 에이전틱 AI 기반 Exposure-TDIR 통합 (출처: 가트너)

그는 이에 대한 향후 방향성으로 데이터 통합이 아닌 인사이트 통합을 제시했다. 그리고, 노출 관리 플랫폼이 생성한 위험도, 공격 그래프, 검증 결과와 SOC의 탐지 결과, 위협 인텔리전스, 공격 진행 상황 인사이트를 연결해야 한다고 강조했다. 그리고, 이 과정에서 AI 에이전트가 노출 플랫폼, SIEM, XDR, 위협 인텔리전스를 실시간으로 조회하고 연결하는 인터페이스 역할을 할 수 있다고 설명했다. 핵심은 데이터를 중앙으로 모두 이동시키는 것이 아니라 필요한 시점에 필요한 컨텍스트를 조회하고, 이를 기반으로 판단과 대응을 고도화하는 것이다.

맺음말: 에이전틱 AI와 함께하는 사이버 보안의 미래

가트너 서밋 2026은 수 많은 세션과 벤더 부스를 통해 에이전틱 AI를 중심으로 나아가는 사이버 보안의 방향성, 그리고 이에 대한 최신 인사이트를 확인할 수 있는 자리였다.

[그림 8] 가트너 서밋 2026 부스 전경

안랩 역시 글로벌 트렌드에 발맞춰 에이전틱 AI 보안 플랫폼 'AhnLab AI PLUS'를 중심으로 AI를 활용한 보안과 AI 모델에 대한 보안까지 폭 넓은 AI 보안 역량을 제공한다. 이를 통해, 여러 조직들이 추진하고 있는 AI 전환(AX)에 신뢰와 안정성을 더하는 파트너로 자리매김하고 있다.

▶**AhnLab AI PLUS 자세히 알아보기**

- **AhnLab**

콘텐츠마케팅팀 신재만 매니저