pwc

에이전틱 AI와 함께하는 Intelligent Enterprise

새로운 실행 주체인 에이전틱 AI를 성공적으로 구현하고 운영하기 위해 프로세스, 데이터, 조직, 보안 등 기업 전반에 어떤 준비가 필요한지를 제시합니다.

March 2026

Table of contents

들어가며

생성형 AI는 이미 많은 조직에서 일상적인 업무 도구로 자리 잡았습니다. 요약과 검색, 문서 초안 작성에서 출발한 AI 활용은 이제 스스로 계획하고 판단하며 실제 업무를 수행하는 에이전틱 AI로 빠르게 진화하고 있습니다. 그러나 기술의 발전 속도에 비해, 이를 조직 안에서 어떻게 활용하고 관리할 것인가에 대한 논의는 아직 충분하지 않습니다. 많은 기업이 AI를 도입했지만, 기대했던 수준의 성과로 이어지지 못하거나 파일럿 단계에 머무르는 이유도 여기에 있습니다.

에이전틱 AI는 단순한 도구가 아니라, 조직의 프로세스와 데이터, 의사결정 구조에 직접 개입하는 새로운 실행 주체입니다. 이는 AI를 잘 쓰는가의 문제가 아니라, AI가 실제로 일할 수 있도록 조직의 운영 모델을 어떻게 재설계할 것인가의 문제로 이어집니다. 특히 프로세스, 데이터, 기술, 조직, 그리고 보안과 거버넌스는 분리된 과제가 아니라 하나의 연결된 구조로 함께 고려되어야 합니다.

본 보고서는 이러한 문제의식에서 출발하여, 에이전틱 AI 시대를 준비하기 위해 기업이 점검해야 할 핵심 이슈와 실질적인 접근 방향을 살펴봅니다. 이 과정에서 에이전틱 AI 도입을 둘러싼 접근 방향을 시작으로, 구현 사례, 데이터와 조직의 변화, 그리고 확산 단계에서 필연적으로 고려해야 할 보안 리스크까지 구체적으로 다룹니다. 이를 통해 기업이 에이전틱 AI를 일회성 기술 도입이 아닌, 지속 가능한 운영 역량으로 정착시키기 위해 무엇을 준비해야 하는지 살펴보고자 합니다.

에이전틱 AI 시대 준비를 위한 접근 방향

왜 AI 도입 성과로 이어지지 못했는가?

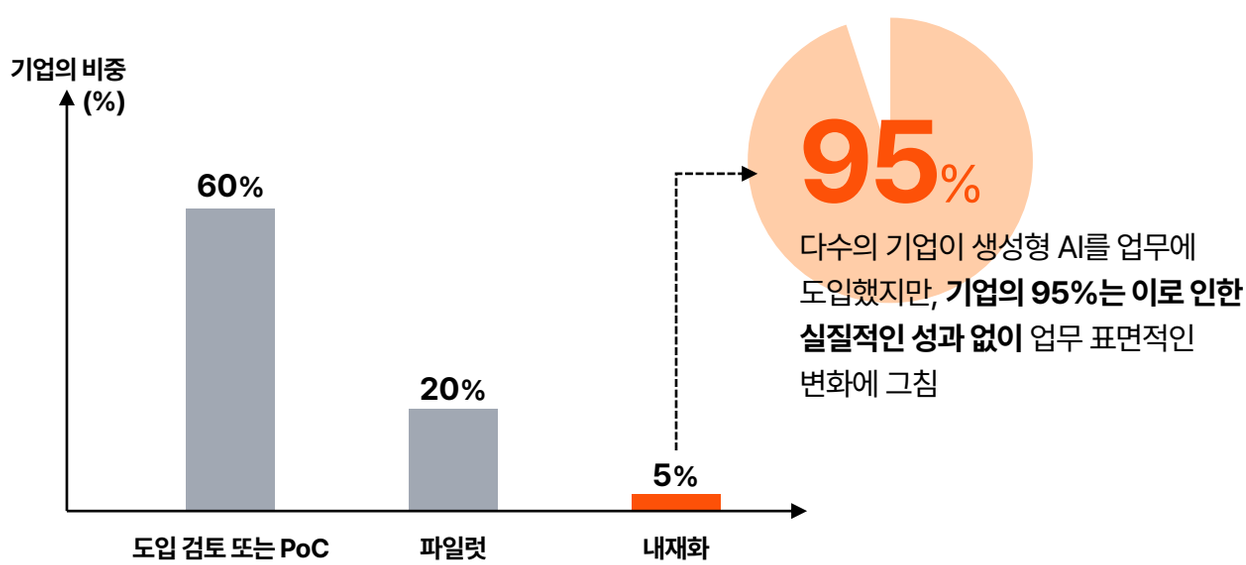
성과로 이어지지 않는 AI 투자

생성형 AI는 이미 많은 기업의 일상 업무에 깊숙이 스며들었습니다. 요약, 번역, 문서 초안 작성, 질의응답과 같은 기능은 개인 단위의 생산성을 일정 수준 개선하는 데 기여해 왔습니다. 특히 화이트칼라 업무 전반에서 반복적이고 시간 소모적인 작업을 줄이는 데 의미 있는 역할을 수행해 왔습니다.

그럼에도 불구하고 많은 기업은 실질적인 재무적, 운영적 성과를 경험하지 못하고 있습니다. 그 이유는 기술 자체의 한계라기보다는 접근 방식의 문제에 가깝습니다. 초기 AI 과제는 대체로 빠르게 적용 가능한 과제 위주로 선정되었으며, 기존의 업무 프로세스나 의사결정 구조에는 거의 변화를 주지 못했습니다.

이러한 과제들은 PoC 단계에서는 가시적인 데모 효과를 보여주었으나, 실제 업무 현장에 깊이 통합되지 못했습니다. 그 결과 AI는 일부 직원이 선택적으로 사용하는 보조 도구에 머물렀고, 조직 전체의 생산성이나 비용 구조를 변화시키는 수준에는 이르지 못했습니다.

생성형 AI 도입에 따른 성과 격차



출처: MIT Media Lab

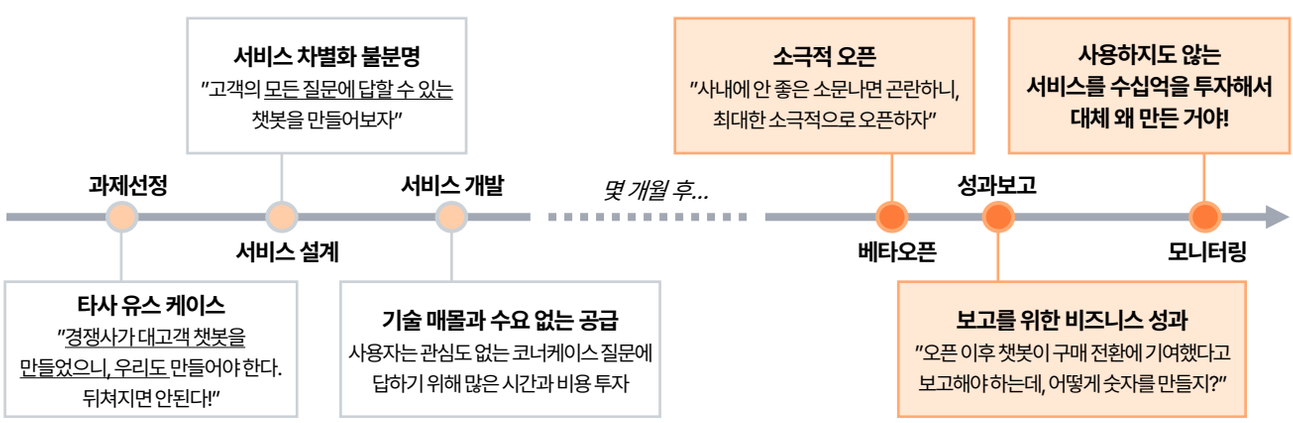
성과 미흡의 구조적 원인

AI 도입이 성과로 이어지지 못한 원인은 여러 기업에서 공통적으로 관찰됩니다. 예를 들자면, 첫째, 불명확한 목표입니다. AI를 도입해야 한다는 압박 속에서 수단과 목적이 뒤바뀌어 생성형 AI를 업무에 빠르게 적용하는 것 자체를 목적으로 진행된 과제들이 다수 존재합니다. 반면 정작 중요한 '이 과제가 어떤 성과를 창출해야 하는지'에 대한 합의는 부족했습니다. 이로 인해 성과 측정 기준 역시 모호해질 수밖에 없었습니다. 둘째, 핵심 경쟁력과의 연계 부족입니다. 비즈니스 니즈가 아닌 기술 접근성에서 출발한 대부분의 과제는 개인의 업무 효율성, 편의성을 지원하는 도구로는 의미가 있으나, 실제 업무에 통합되지 않아 사용도가 낮고, 가치 창출도 어렵습니다. 셋째, 변화관리 및 책임 구조의 부재입니다. AI 서비스는 도입되었으나, 이를 적극 활용하도록 유도하는 조직적 장치와 성과에 대한 책임 구조가 충분히 마련되지 않습니다.

생성형 AI의 도입 성과 부족의 원인

AI 도입 시 비즈니스 목표 불명확	인사규정 챗봇을 만들었는데, **목표가 뭐지?** 비용 절감? **그냥 있으면 좋은 기능인가?**
성과가 아닌 AI 적용이 용이한 과제 선정	**AI가 잘 할 수 있는 게** 뭐지? 요약, 초안 작성 같은 걸 잘 한다고 하던데...
핵심 경쟁력과 무관한 AI 도입	사내에 AI 어시스턴트만 10개가 넘는데, **없어도 일하는데 문제 없고...**
내 몸에 맞지 않는 타사의 유스 케이스	**타사는 AI로 뭘 했지?** 우리는 뒤쳐지면 안되는데...
사용성 및 변화관리 부족	**불완전한 결과물**을 전사에 공개했다가 사내 게시판에 불만이 올라오면 **누가 책임지나?**

차별적 목적 없이 타사의 유스 케이스를 도입하는 예시



출처: PwC

제언 ① 운영 전략과의 연계

명확한 목표 설정과 정량적 목표 설정

에이전틱 AI 도입의 출발점은 기술이 아니라 목적입니다. 산업과 기능별로 AI가 지향하는 바는 모두 다르기 때문에, 구성원 다수가 공감할 수 있는 명확한 방향성을 먼저 정의하는 것이 중요합니다. 즉, 생산성 향상, 원가 절감, 리스크 완화 등 목표를 설정한 뒤, 이를 정량화하여 목표로 명문화해야 합니다.

예를 들어, 한 제조 기업은 주 4일제 근무제에 선제적으로 대비하기 위해 AI를 통한 생산성 향상이라는 추상적 목표 대신, 2030년까지 인당 생산성을 30% 향상시키겠다는 구체적인 목표를 설정했습니다. 이후 모든 에이전트 과제는 해당 목표에 얼마나 기여하는지를 기준으로 평가되었고, 실행력도 크게 강화되었습니다.

집중 영역의 우선적 선별

모든 업무에 에이전트를 도입할 필요는 없습니다. 자원이 제한된 상황에서는 ROI가 가장 높은 영역에 집중하는 것이 중요합니다. 한 기업은 FTE 및 비용 분석을 수행한 결과, 상위 20% 프로세스가 전체 업무 부담의 80%를 차지한다는 사실을 확인했습니다. 이에 따라 해당 영역을 우선 집중하는 전략으로 전환해 성과를 극대화했습니다.

이 과정에서 "이 과제를 왜 하는가? 지금 가장 중요한 과제인가?"라는 질문에 답하려면 정량적으로 접근해야 합니다. 특히 1 FTE 이상 효과가 기대되는 영역에 집중하면 성과가 명확해지고 내부 공감도 얻을 수 있습니다. 또한 에이전틱 AI는 지속적으로 관리해야 하는 영역이므로, 지속적으로 업데이트할 수 있는 기준을 마련하는 것이 필수입니다.

에이전트 도입을 반영한 인사 계획 수립

에이전트 도입을 추진할 때 "에이전트의 운영 비용은 어느 정도인가?"와 더불어 "실제로 인간을 대체할 수 있는가?"를 질문으로 시작해야 합니다. 그렇지 않으면 업무는 자동화되지만, 기존 인력은 그대로 유지되거나 오히려 에이전트와 인간이 함께 늘어날 수 있습니다. 따라서 에이전트 도입은 HR 중장기 계획과 연계해 인력 증가를 억제하거나 재배치하는 방향으로 설계해야 합니다. 다만, 조직의 경쟁력을 유지해야 하는 핵심 영역을 명확히 구분한 뒤, 그 외 영역에서 에이전트를 활용해 생산성을 개선하는 방식이 바람직합니다.

PwC가 경험한 프로젝트 사례를 보면 한 기업은 과제의 실현 가능성을 고려해 연도별 에이전트 도입에 따른 업무 대체 효과를 수치로 예측했고, 2030년까지의 인사 계획을 조정했습니다. 이를 통해 구조적 생산성 개선이 가능했습니다.

제언 ② 내부 프로세스 분석기반의 접근

외부 사례보다 내부 프로세스를 우선시

에이전트는 인간이 하는 일을 도와주고 대체하는 역할을 하며, 외부 사례는 참고일 뿐 정답이 될 수 없습니다. 기업마다 시스템과 데이터가 다르기 때문에 같은 에이전트라도 효과가 달라질 수 있습니다. PwC는 다수 프로젝트에서 업무 활동 단위 분석을 통해 외부에서는 보이지 않던 병목과 비효율을 발견하고, 이에 적합한 에이전트를 설계해 왔습니다.

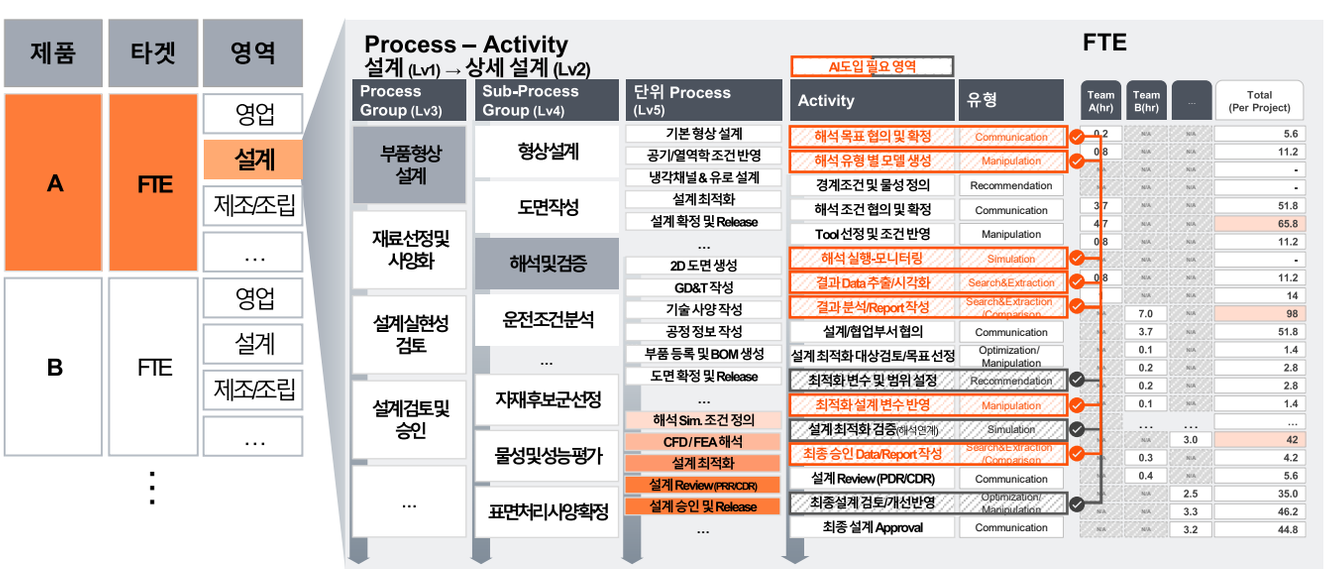
적정 수준의 에이전트 기획

불필요한 에이전트는 비용과 관리 리스크를 증가시킵니다. 한 기업은 고도화된 에이전트를 목표로 했지만 비용 대비 효과가 제한적이어서 단순 자동화 중심으로 방향을 전환하였습니다. 잘못된 에이전트 메모리 설계는 비용을 크게 증가시키고, 불필요한 컨텍스트로 토큰 비용이 폭증할 수 있습니다. 따라서 업무 성격과 우선순위에 따라 단순 자동화가 필요한 영역과 지능화된 AI가 필요한 영역을 구분해 에이전트 목표 수준을 정의해야 합니다.

AI 맵을 통한 통합 관리

AI 맵은 조직 내 모든 에이전트를 한눈에 관리할 수 있는 조직도 역할을 하므로 기업은 필요한 에이전트를 식별하고, 기존 도입된 에이전트를 통합 관리해 중복 투자와 관리 리스크를 최소화할 수 있습니다. 내부 프로세스와 결합되어 제시되는 AI 맵은 에이전트의 적용 프로세스와 기술 유형을 분류한 업무 기술서로 활용될 수 있습니다.

에이전트 도입을 위한 내부 업무 프로세스 파악 예시



출처: PwC

제언 ③ 프로세스 설계와 성과 관리

에이전틱 워크플로우의 자산화

에이전틱 워크플로우는 에이전트가 활용하는 업무 지침서이자 에이전트 확대의 구체적인 이정표입니다. 먼저 전체적인 청사진을 구상하고 단계별로 실현해 나가야 합니다. 그러려면 워크플로우를 기업의 자산으로 다루고 장기적으로 관리할 수 있는 체계가 필요합니다. 청사진이 없으면 담당자가 변경됨에 따라 기존 자산이 사장되는 경우가 비일비재합니다.

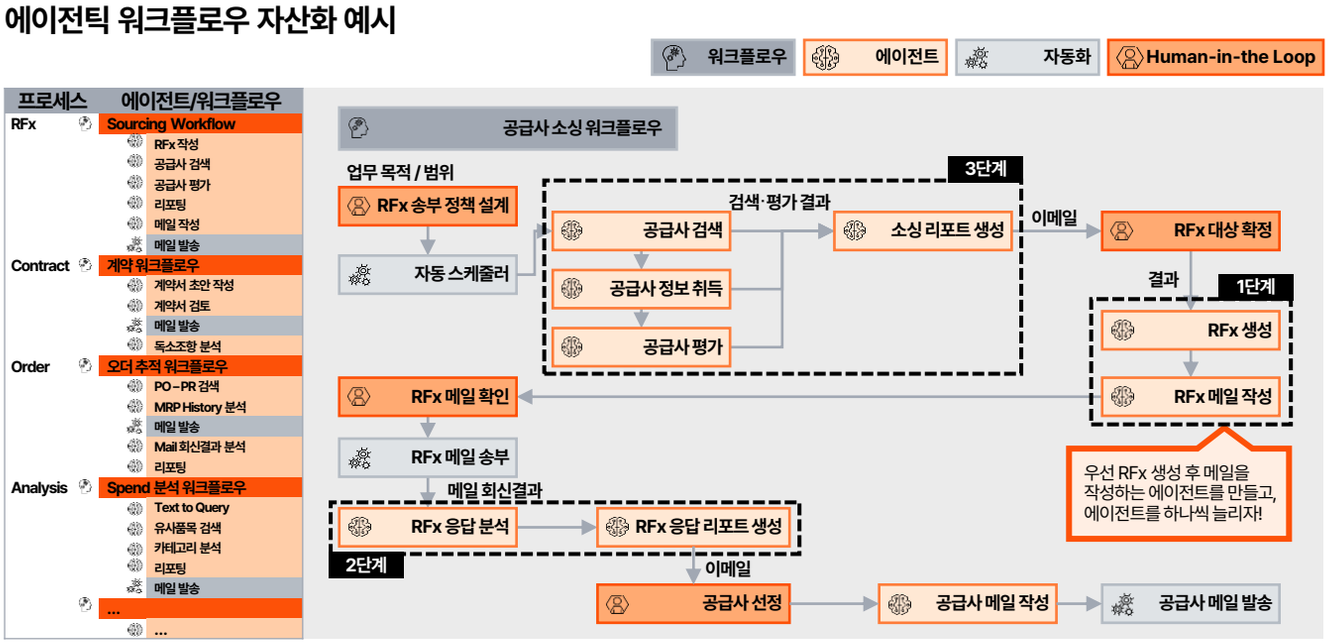
에이전트의 성능 지표와 성과 지표를 구분

에이전트가 실제 비즈니스 성과에 기여하지 못하는 상황을 방지하려면 에이전트의 성능 지표(예: 정확도, 안정성)와 성과 지표(예: 비용, 시간, 리스크)를 구분하여 관리해야 합니다. 성능 지표만 보면 데모에 속기 쉽고, 성과 지표만 보면 문제의 원인을 추적하기 어렵습니다. '잘 작동하는 에이전트'와 '비즈니스에 기여하는 에이전트'를 구분하고 에이전트 지표를 과제 지표, 전사 지표와 연계하여 설계해야 합니다.

지속적인 피드백 및 리뷰 프로세스 정립

에이전틱 AI를 완성된 시스템이 아니라 지속 교정해야 할 디지털 인력으로 봐야 합니다. 에이전틱 AI는 운영 중에도 업무 맥락이 계속 변하고, 에이전트의 판단이 실제 프로세스에 직접적인 영향을 미칩니다. 따라서 일반적인 DT 과제 관리 방식이 아닌 실험-검증-피드백이 상시 반복되는 체계가 필요합니다. 환각이나 컴플라이언스와 같은 리스크도 별도의 모니터링 체계로 관리해야 합니다.

에이전틱 워크플로우 자산화 예시



출처: PwC

PwC AI 애셋 및 에이전트 구현 사례

성공적인 AX를 위한 고려사항

에이전틱 AI의 성공적인 도입은 기술 자체의 도입을 넘어, 비즈니스 전략과 조직 역량의 근본적인 혁신을 요구합니다. AX 도입 효과를 극대화 하기 위해서는 "AI를 무엇을 위해 사용하는가?" 이를 위해 "비즈니스 프로세스는 최적화 되었는가?"와 같은 6가지의 핵심 질문을 고려해야 합니다.

PwC는 AX 도입 효과를 극대화하기 위해 'Discover', 'Build', 'Scale'의 3단계 접근법과 6가지 핵심 역량 영역에서의 혁신을 강조합니다. 본 장에서는 그 중 AX 기반으로 핵심역량을 확보하고 강화하는 방향으로 적합한 프로세스를 정의하고 과제를 발굴하고 구현하는 프로세스 영역에 대해서 다루고자 합니다.

AX 도입 효과 극대화를 위한 고려사항

AX 추진 관련 핵심 질문
AI를 무엇을 위해 사용하는가?
비즈니스 프로세스가 최적화 되었는가?
필요한 데이터에 접근 가능한가?
적절한 시스템과 툴을 보유하고 있는가?
적절한 조직 구성과 거버넌스를 보유하고 있는가?
AX 기반의 조직문화를 가지고 있는가?

접근법	역량	추진 방향·목적 및 PwC 강점
Discover **AI의 가치 발견** 「새로운 수익 창출 원천 선점」	비즈니스 의사결정	**명확한 AX 도입 목적 및 방향성 정의** → PwC BMR[1] 프레임워크 및 전략 전문 조직
	프로세스	**핵심 역량 확보 가능한 영역 정의/과제 발굴 및 구현** → 산업·밸류체인 전문성 및 PwC AI 애셋
Build **AI의 가치 확보** 「비즈니스에 연계된 AI 역량 내재화」	데이터	**비즈니스와 상호작용하는 AI를 위한 데이터 관리 체계 정의** → AI 및 데이터 전문조직과 거버넌스 프레임워크
	테크	**확장 가능한 AI 아키텍처, 최적화된 모델 및 인프라 확보** → PwC 에이전트 OS 및 AI 얼라이언스·파트너십
Scale **AI의 가치 확산** 「AI 상호작용으로 가치 확장 및 지속」	조직	**AI 가치 확산을 위한 조직 운영 방향성 (조직구성, R&R, 성과관리) 확립** → 워크포스 전문 조직 및 AI 성과 모니터링 툴
	인력 및 문화	**AI와 함께 일하는 문화 및 가치 확산 방안 구체화** → PwC AI 리터러시 러닝 플랫폼 및 워크숍

1) Business Model Reinvention
출처: PwC

에이전틱 AI의 효과적 적용을 위한 접근 방법

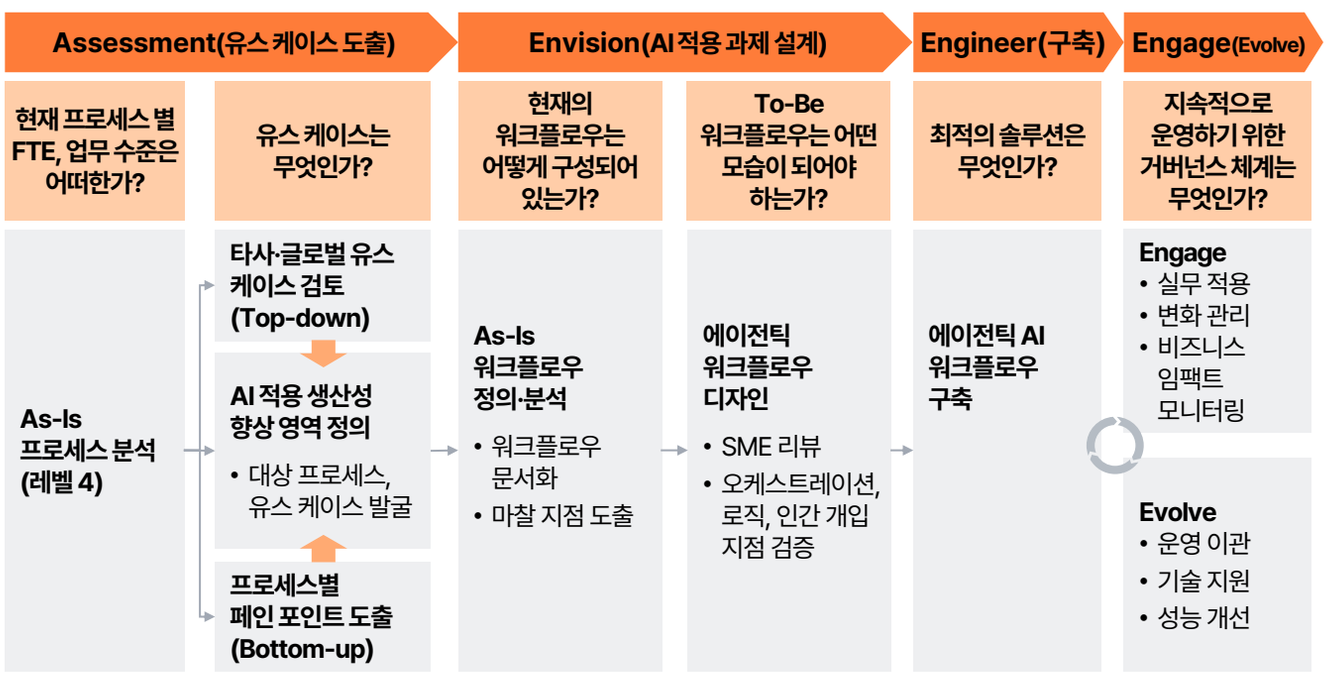
업무 생산성 제고를 위해 효과적인 AX나 에이전틱 AI 적용을 위해서는 도입 시 임팩트가 있고 적용 가능성이 고려된 유스 케이스를 신속히 파악 후, 기대 효과 등 우선순위에 따라 효율적으로 설계하고 적용하는 것이 바람직합니다.

일반적으로 에이전틱 AI는 4단계로 적용합니다. 첫째, 도입 시 생산성 증대와 가치 제고의 효과가 큰 과제를 잘 분석하고 도출하는 Assessment 단계입니다. 이 단계에서는 먼저 프로세스를 레벨 4 수준으로 상세 분석하여, 인력과 시간 투입이 많이 발생하는 영역을 확인하고, 이 영역에 대해 인터뷰 등을 통해 페인 포인트를 파악하고 타사 사례 기반으로 적용 시 효과가 큰 유스 케이스를 발굴합니다. 기존 방식과 다른 점은 단순히 타사 사례를 바로 적용하는 것이 아니라, 개선 효과가 큰 영역을 먼저 발굴하고 적용 시 효과가 큰 유스 케이스 발굴을 위해서 타사의 사례(유스 케이스 풀)를 참고한다는 점입니다.

둘째는 Envision 단계로서 대상이 되는 주요 영역(프로세스)과 유스 케이스에 대해 상세 분석(레벨 5~6 수준)을 하고 이를 기반으로 To-Be 에이전틱 워크플로우를 설계합니다. 이 때 에이전틱 워크플로우 구성요소에 대한 고려 등이 필요합니다. 이후 실제 구현하고 확산 및 운영하는 Engineer 및 Engage(Evolve) 단계를 진행합니다.

과거에는 이러한 과제 도출과 설계를 위한 현황 분석, 상세 업무 플로우 정의 및 AI 적용 효과성 판단과 구현 로드맵 등 정의 단계에서 현업과 컨설턴트들이 수작업하여 시간과 비용이 많이 소요되었습니다. 현재는 AI를 활용하여 신속하고 비용 효율적으로 진행할 수 있는 툴과 방법론 등이 도입되고 있는데, PwC AI 툴들을 예로 들 수 있습니다.

에이전틱 AI 적용 절차



출처: PwC

프로세스 분석 기반의 AX 적용 유스 케이스 파악 방법

많은 기업이 업무 효율이나 FTE 등 인력 구조를 개선할 수 있는 영역이 어디인지, 어디서부터 시작해야 할지를 파악하는데 어려움을 겪고 있습니다.

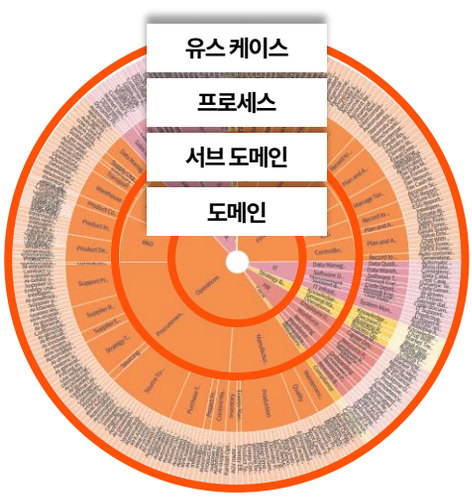
이러한 고민에 대해 신속히 지원하기 위해 PwC는 산업, 도메인, 프로세스별로 약 300개의 글로벌 AI 유스 케이스를 체계화한 AI 유스 케이스 컴퍼스를 통해서, FTE 분석 등을 통해 개선이 필요한 도메인 또는 유스 케이스를 신속히 파악하여, AX 적용이 가능한지를 신속히 검증, 의사결정할 수 있도록 지원합니다.

AI 유스 케이스 컴퍼스는 9개 산업, 14개 기술, 8개 도메인과 28개 서브 도메인, 60개 프로세스 체계로 구성돼 있어, 도입 배경, 솔루션, 효익까지 한눈에 보고 빠르게 후보군을 좁힐 수 있도록 도와줍니다. '우리와 비슷한 상황에서 어떤 케이스가 통했는가'를 즉시 확인하는 도구라고 보시면 됩니다. 이를 통해 초기 마스터 플랜 또는 프로세스 혁신(PI) 시작 단계의 리소스 투입을 최소화 하는데 기여합니다.

AI 유스 케이스 컴퍼스의 개념과 체계

PwC AI 유스 케이스 컴퍼스

- **9개 산업군의 도메인, 서브 도메인, 프로세스별 AI 기술이 사용된 유스 케이스 사례 수집**
  → 다양한 케이스 접근 가능
- **케이스별 도입 배경, 솔루션, 효익 정보 제공**



체계

9개 산업군	자동차, 에너지·유틸리티 및 자원, 금융, 산업재, 운송 및 물류, 제약 및 생명과학, 공공 서비스, 유통 및 소비재, 통신·미디어·기술
14개 기술	생성형 AI, 머신러닝, 자연어 처리, 시뮬레이션, 최적화, 그래프 분석, 가상·증강현실, 이상 탐지, 예측, 프로세스 마이닝, 컴퓨터 비전, RPA, IoT, 로보틱스

도메인	**최상위 영역, 8개** 예) 재무, 오퍼레이션, 세일즈 및 마케팅, IT, 인사, 법무 전략 및 경영, 지속가능성
서브 도메인	**도메인의 하위, 28개** 예) 재무 도메인 하위 6개 서브 도메인: 회계, 관리회계·기획·리포팅, 리스크 관리, 자금, 세무, ESG
프로세스	**서브 도메인의 하위, 60개** 예) 재무 도메인 하위 6개 프로세스: 계획 및 분석, Record to Report(R2R), Order to Cash(O2C), Procure to Pay(P2P), 자금 관리, 세무 관리
유스 케이스	**프로세스의 하위, 약 300개** 예) 영수증 자동 처리 및 비용 분류, 자동 분개 생성, 생성형 AI 기반 계획 수립, 운영비(OPEX) 예측, 생성형 AI 기반 재무 보고서 자동 코멘트 생성 등

출처: PwC

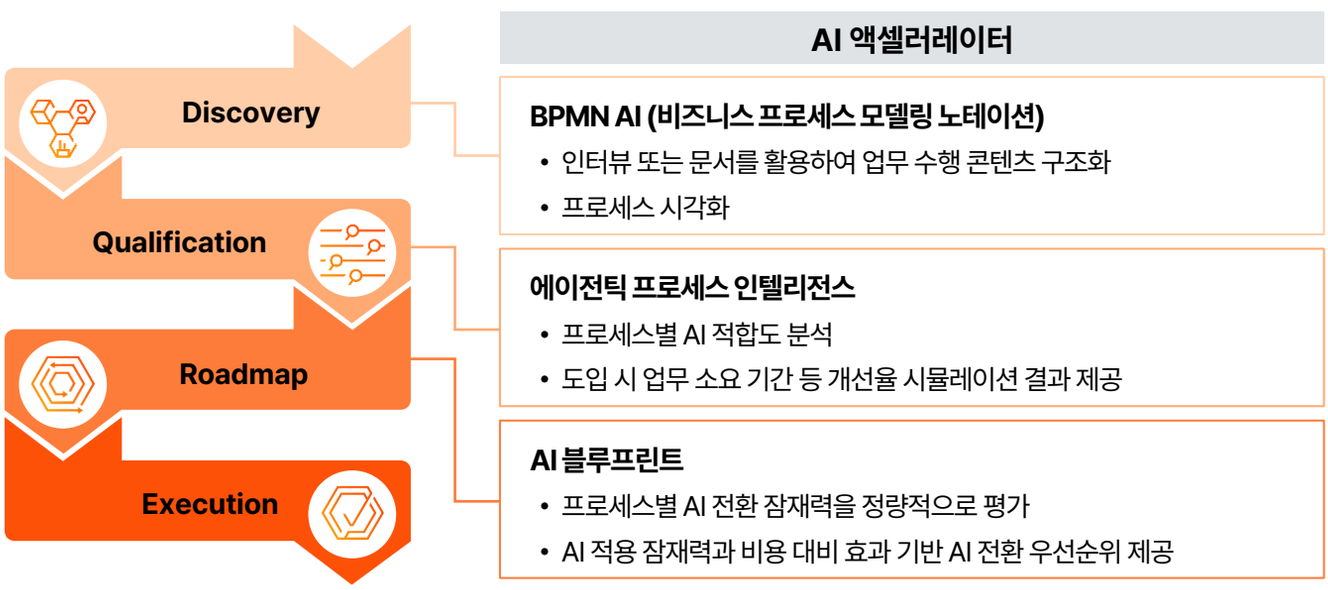
비즈니스 가치 평가 및 에이전틱 프로세스 설계 방안

다음은 빠른 적합도 분석과 우선순위화를 통해 실행력을 가질 수 있도록 체계적인 로드맵을 수립하는 것이 중요합니다.

Discovery, Qualification, Roadmap, Execution으로 이어지는 비즈니스 가치 평가 및 에이전틱 프로세스 설계 과정에서는 시간과 비용을 절감하고 효율적으로 진행할 수 있도록 다양한 방법론과 툴들이 활용되고 있습니다. PwC는 AI 기반 툴을 활용하여 AI 전환 기회의 신속한 식별, 적합도 분석과 효과 시뮬레이션 및 우선순위 선정 시 과제 도출과 설계를 효율적이고 효과적으로 진행할 수 있도록 지원합니다.

인터뷰나 문서를 활용하여 프로세스 구조화를 자동화하고(BPMN AI), AI 적합도와 기대 개선율을 시뮬레이션하며(에이전틱 프로세스 인텔리전스), 전환 잠재력과 비용 효과를 기반으로 우선순위 정량화를 지원합니다(AI 블루프린트).

프로세스 설계 단계별 액셀러레이터



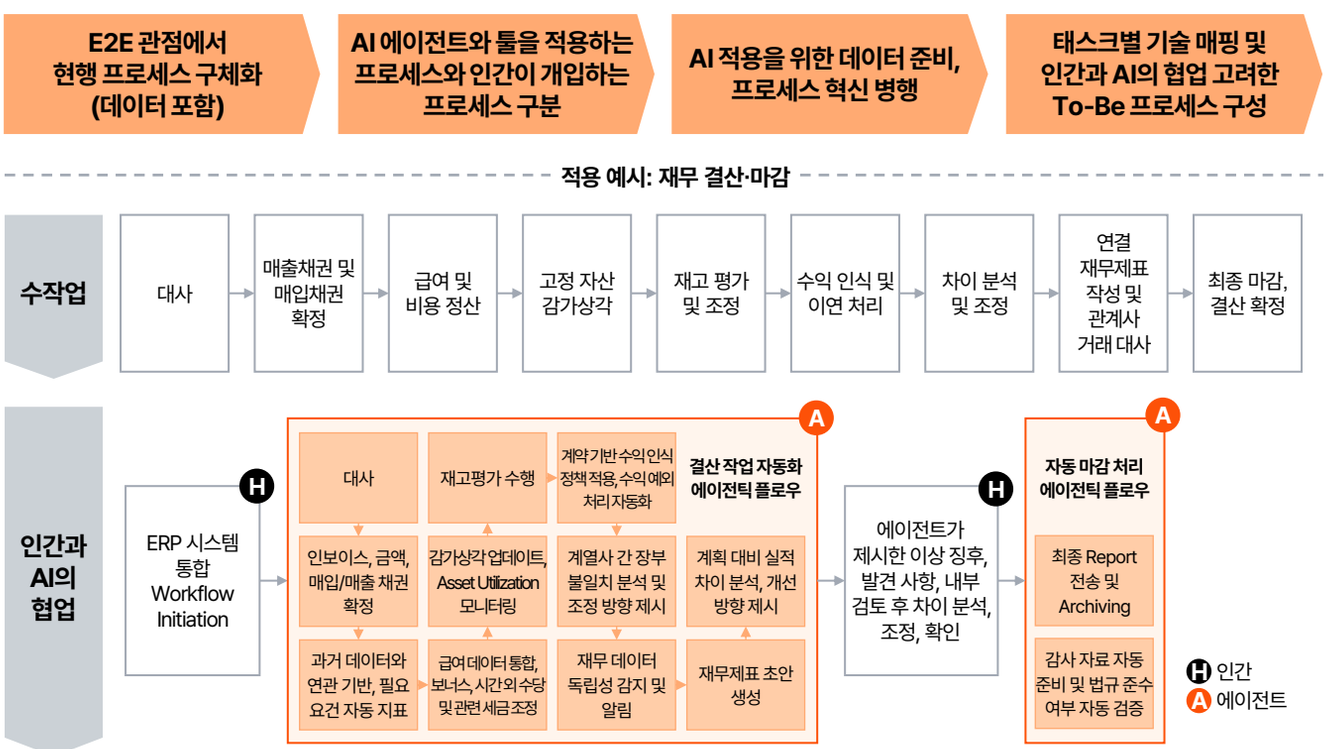
출처: PwC

에이전틱 AI 워크플로우 설계 및 적용 사례

앞서 과정들을 거쳐 생산성 제고 효과 등이 큰 과제가 도출이 되었다면, 이를 실제 구현해서 효과를 보기 위한 설계 방법 및 적용 사례를 소개합니다.

E2E(End-to-end) 업무 효율화와 자동화를 위해서는 프로세스와 데이터를 명확히 분석하고, AI와 인간이 협업하기 위한 최적의 To-Be 워크플로우를 구성해야 합니다. 이때 기존의 자동화 자산과 새로운 AI 에이전트를 통합해서 효율, 품질, 통제를 높이는 에이전틱 워크플로우를 잘 구성하는 것이 중요합니다. 이 과정에서 기존 프로세스나 데이터 체계를 AI가 잘 활용할 수 있도록 프로세스 혁신(PI)을 병행하는 과정이 필수입니다.

E2E 에이전틱 AI 적용을 위한 워크플로우 구성 PI



출처: PwC

재무 결산·마감을 예시로 보면, 과거엔 대사, 채권 확정, 비용 정산, 감가상각, 재고 평가, 수익 인식, 차이 분석, 연결 마감, 최종 확정까지 9단계를 모두 수작업으로 처리했었습니다. 많은 영역을 ERP 등 시스템을 통해 처리하지만, 인간이 입력, 처리, 확인하는 등의 수작업이 필수였습니다. 에이전틱 AI를 도입하면 이러한 업무에 있어서 대부분은 에이전트가 처리하고, 업무를 시작하는 트리거 포인트나 중요한 검증 또는 승인 절차 정도만 인간이 개입하여 E2E 업무가 자동화하는 형태로 발전할 것입니다.

에이전트에 의해 전표는 과거 데이터와 모델에 의해 자동 기표되고, 급여·세금 조정, 수익 인식과 이연, 재고 평가, 불일치 및 차이 분석, 감사자료 준비와 규정 준수 검증까지 에이전트가 수행합니다. 인간은 에이전트가 제시한 이상 징후와 불일치·재무제표 초안·차이 분석을 검토하고, 마지막으로 확정·보고 등을 승인합니다.

물론 이상은 궁극적인 모습이고 현재는 이 중 몇 단계를 AI가 대체하고 나머지를 인간이 할지는 프로세스나 시스템 수준에 따라서 많이 상이하고, 일부는 아직 AI의 기술적인 성숙도 미흡으로 적용이 어려운 부분도 있습니다. 하지만, 기술이 발전하면서 에이전트로 적용하는 영역이 점점 넓어지고 있고, 기업에 맞는 이러한 에이전틱 워크플로우를 잘 구성하는 것이 설계의 핵심입니다.

에이전틱 AI 워크플로우의 핵심 구성 요소

에이전틱 AI 워크플로우 구성 시 다음의 에이전트, 인간, 툴, 워크플로우의 4가지 표준 구성 요소를 고려하여 적재적소에 배치하는 것이 중요합니다.

- **에이전트:** 수행 목적, 대상 업무, 인풋 데이터와 산출물 뿐만 아니라 활용할 툴 등에 대한 정의 필요
- **인간:** 과업 개시자, 검토·검증자, 의사결정자, 예외 처리자 등 다양한 역할을 수행하며, AI가 처리하기 어려운 예외적인 상황이나 최종적인 의사결정이 필요한 부분에 개입하여 AI와 협업
- **툴:** 에이전트가 과업을 수행하는 데 필요한 다양한 시스템, DB, API 등을 의미하며 기존의 Non-에이전틱 프로그램(RPA 등) 또한 에이전트가 호출하여 활용 가능
- **워크플로우:** 상기 구성 요소들이 상호작용하며 비즈니스 프로세스를 완수해 나가는 일련의 절차와 규칙

To-Be 에이전틱 AI 워크플로우 구성

에이전트	인간	툴 (Non-에이전틱)	워크플로우
에이전트를 정의하는 속성 • **목적:** 에이전트의 전반적인 역할을 정의하는 시스템 프롬프트 • **작업:** 에이전트가 수행하는, 명확히 구분되는 실행 가능한 최소 작업 단위 • **툴·기능:** 에이전트가 행동을 위해 사용할 수 있는 외부 시스템 연동 기능 (API 등) • **입력 및 출력:** 에이전트에게 주어지는 정보와 에이전트가 생성하여 내보내는 결과물 • **트리거:** 에이전트의 작동을 개시하는 특정 이벤트나 조건	**에이전틱 워크플로우에 개입하는 인간의 유형** • **과업 개시자**: 에이전트에게 필요한 컨텍스트와 정보를 제공하여 워크플로우를 시작 • **검토·검증자**: 에이전트가 생성한 결과물을 확인하고 피드백을 제공하거나 직접 수정 또는 개선 • **의사 결정자**: 에이전트가 제시한 여러 대안 중 하나를 최종 선택 • **예외 처리자**: 에이전트가 모호함 또는 실패에 직면했을 때 개입하여 해결	**에이전틱 워크플로우는 기존의 Non-에이전틱 프로그램을 호출할 수 있어야 함** **Non-에이전틱 툴 예시** • RPA(Robotic Process Automation) • Rule-based Task • Time-based Triggers	**기존 프로세스 계층 구조와 동일하게, 에이전틱 워크플로우 또한 다수의 하위 워크플로우를 연결하는 방식으로 구성** • **Chained Workflows**: 하나의 워크플로우가 종료된 후, 그 결과에 따라 즉시 하나 또는 그 이상의 워크플로우가 연이어 실행되는 구조 • **Nested Workflows**: 상위 워크플로우가 특정 작업을 처리하기 위해 서브 워크플로우를 호출하고, 서브 워크플로우는 작업을 완료한 후 그 결과를 다시 상위 워크플로우로 반환하는 구조

에이전틱 워크플로우의 4대 요소가 '노드 (Node)'로서 워크플로우 내에서 연계되어 역할 수행

출처: PwC

에이전트 AI, 인간과 AI의 협업을 통한 비즈니스 혁신

에이전틱 AI는 더 이상 먼 미래의 기술이 아닌, 기업의 경쟁력을 좌우하는 핵심 동력으로 자리 잡고 있습니다. 성공적인 도입을 위해서는 명확한 비전 수립, 데이터 거버넌스 확보, 조직 역량 내재화가 선행되어야 합니다. 또한, 위 설명에서 보듯이, 효과가 높은 과제를 선별하고 이를 실제 운영할 수 있도록 워크플로우를 설계하며 파일럿 프로젝트를 통해 성공 사례를 만들고 점진적으로 확산해 나가는 전략이 유효합니다.

궁극적으로 에이전틱 AI는 인간을 대체하는 것이 아니라, 인간이 더욱 창의적이고 전략적인 업무에 집중할 수 있도록 돕는 강력한 파트너가 될 것입니다. 에이전틱 AI와의 효과적인 협업 체계를 구축하는 기업만이 미래 비즈니스 환경에서 선도자가 될 수 있을 것입니다.

AI 시대를 위한 데이터의 새로운 기준

AI 도입의 당면 이슈

AI는 왜 기대만큼 성과를 못 내는가?

많은 기관들이 공통적으로 지적하는 바와 같이, AI 투자는 늘었지만 현장에서의 성과는 기대에 미치지 못하고 있습니다. 그 이유는 AI 분석 결과를 참고로 활용할 뿐, 실제 의사결정에는 제대로 사용하지 못하기 때문입니다. 결국 "AI를 정말 믿어도 될까?"라는 의심이 문제의 핵심입니다.

본 챕터에서는 특히 데이터를 관리하는 방식에 초점을 맞추어 논의하고자 합니다. 기업 내부에 데이터는 많지만, 시스템 간 분절되어 있는 사일로 현상과 시스템 상 동일 용어에 대해 상이한 정의를 사용하는 문제가 존재합니다. 예를 들어 '불량률' 하나를 두고도, 생산팀은 재작업을 하면 괜찮다고 하고, 품질팀은 무조건 불량으로 정의하는 상황에서, AI에게 "불량률이 얼마야?"라고 질문할 경우, 어느 기준으로 답해야 할지 혼란에 빠질 수 있으며, 업무 맥락이 정의되지 않은 상태에서는 AI가 근거 있는 답변을 도출하기 어렵습니다. 이런 데이터로는 AI가 과거 패턴을 분석해서 "곧 문제가 생길 것 같습니다"라고 알람을 주는 수준에 머물 수 밖에 없습니다. 왜 그런 결론이 나왔는지, 그 근거가 무엇인지 설명하지 못하니, 현업에서는 당연히 그 결과를 신뢰하고 행동으로 옮기기 어려운 것입니다.

현장에서 접하는 현실

AI 도입의 현실

45% **AI 확산을 가로막는 요인으로 부정확성과 편견을 꼽은 비율 (1위)**

25% **기대한 ROI를 달성했다는 기업의 비율**

95% **생성형 AI 파일럿 실패율**

출처: MIT, IBM

AI 투자는 늘었지만...
현장에서의 실질 성과는 제한적

다수의 제조 기업이...
- AI PoC 및 파일럿 프로젝트 수행
- 설비, 공정, 품질 등의 데이터 인프라 구축

그러나, 실제 운영에서는...
- AI의 답변이 의사결정의 참고 자료 수준
- 현장의 판단 또는 조치의 기준으로 사용하기에는 부담

이러한 상황에서 가능한 AI 활용은...
- 분석 결과 제공 가능
- 예측, 알람 생성 가능
- **판단 기준으로 사용하기 불안**
- **"왜 이 결론인가?"는 설명 불가**

우리가 지향하는 AI의 궁극적인 모습, 에이전틱 AI

에이전틱 AI는 스스로 사고하고 행동하는 AI 비서입니다. 기존의 AI가 단순히 주어진 데이터를 입력 받아 정해진 답을 출력하는, 이른바 자판기와 같은 존재였다면, 에이전틱 AI는 목표가 주어졌을 때 스스로 계획을 수립하고 행동하는 능동적인 주체라고 할 수 있습니다. 이러한 특성으로 인해 에이전틱 AI는 "이상이 있는가?"와 같은 단순한 질문에 응답하는 수준을 넘어, "무엇을, 왜, 어떻게 조치해야 하는가?"와 같은 복합적인 질문에 대해서도 종합적인 해결책을 제공할 수 있습니다. 이는 우리가 지향하는 지능적인 AI의 모습입니다. 그렇다면 이러한 에이전틱 AI를 구현하기 위해서는 필요한 데이터는 무엇인가? 바로 조직이 수행하는 업무(정보, 워크플로우 등)가 디지털 언어로 구조화된 형태의 데이터입니다. PwC는 이러한 데이터를 AI-ready Data라고 정의합니다.

에이전틱 AI가 잘 작동하기 위한 조건과 기존 AI와의 차이

기존 AI 와 비교하여 에이전틱 AI 는...	에이전틱 AI가 잘 작동하기 위해서는...	
기존 AI가 **"이상 여부를 알려줘"**에 답할 수 있다면, **에이전틱 AI**는 **"지금 이 공정 상태에서 무엇을, 왜, 어떻게 조치해야 하는가?"**에 답할 수 있음	**단순 입력 기반 출력 탈피**	**목표 기반의 행동**
	주변 상황을 인지하고 판단	**스스로 계획, 추론, 행동 선택**

출처: PwC

AI-ready Data란?

AI가 이해, 추론, 판단, 설명까지 가능한 구조로 정렬된 데이터

많은 에이전틱 AI-ready Data의 첫 번째 특징은 업무 맥락(Context)입니다. AI가 현재 어떤 상황에 놓여 있는지를 이해할 수 있어야 합니다. 예를 들어, AI가 '온도 5도 상승'이라는 데이터를 입력 받았다고 가정해 보겠습니다. 해당 수치가 단순한 일상적 변동인지, 아니면 심각한 문제의 전조인지는 현재 상황이 통상적 운영 상태인지, 신제품 테스트 단계인지 또는 라인 전환 대기 상태인지에 따라 전혀 다르게 해석될 수 있습니다. 이처럼 지금 어떤 업무를 수행하고 있는지에 대한 맥락 정보가 없을 경우, 데이터는 단순한 숫자에 불과합니다.

두 번째 특징은 객체 간의 관계(Relationship)입니다. 즉, '무엇이 무엇에 영향을 미치는지'에 대한 연결 구조를 AI가 이해해야 합니다. 예를 들어, 특정 설비의 레시피를 변경했을 때 어떤 제품의 품질과 수율에 어떤 영향을 미치는지, 그리고 해당 정보가 MES, SPC, QMS와 같은 시스템과 어떻게 연계되는지에 대한 인과관계의 지도가 필요합니다. 이러한 관계 정보가 구조화되어 있어야 AI가 복합적인 상황을 종합적으로 판단할 수 있습니다.

세 번째 특징은 판단 기준(Rule & Constraint)입니다. 이는 '무엇은 허용되며, 무엇은 절대 허용되지 않는지'에 대한 명확한 기준을 의미합니다. 품질 기준의 상·하한선, 납기 준수를 위해 반드시 지켜야 하는 운영 규칙, 그리고 숙련된 현장 작업자들의 경험 속에 축적되어 있던 '이 경우에는 이렇게 대응한다'는 암묵적인 노하우까지 모두 포함됩니다. 이러한 판단 기준이 데이터로 체계화되어야만 AI가 비합리적이거나 엉뚱한 결정을 내리지 않고, 인간과 유사한 수준의 합리적인 의사결정을 수행할 수 있습니다.

마지막으로, 반복 사용 및 확장 가능성이 확보되어야 합니다. 한 번 구축된 데이터는 특정 문제 해결에만 국한되지 않고, 다른 문제나 다른 AI에도 재사용될 수 있어야 합니다. 나아가 AI 간 상호 검증이 가능한 구조로 확장될 수 있을 때, AI-ready Data의 가치가 극대화될 수 있습니다.

클린 데이터와 AI-ready Data의 차이, 체계적 접근의 필요성

에이전틱 원본 데이터(Raw Data)를 1차 정제하면 '클린 데이터(Clean Data)' 혹은 'AI가 읽을 수 있는 데이터(AI-readable Data)'가 됩니다. 많은 기업들이 이 상태의 데이터를 AI 혹은 ML 등 데이터 분석에 많이 활용하고 있습니다. 여기에 업무 의미와 관계를 부여하는 구조화, 맥락화, 지능화, 표준화 등의 과정을 더해야만, 비로소 AI가 스스로 학습하고 추론하는 AI-ready Data가 완성됩니다.

AI-ready Data의 4가지 특징

업무 의미가 명확히 정의된 데이터

- 숫자, 값이 아니라 업무 관점의 의미가 연결
- 이 수치는 무엇을 판단하기 위한 것인가를 반영

객체 간 관계가 연결된 데이터

- 설비, 공정, 품질, 계획 데이터를 통합적으로 고려
- 원인–영향–결과를 추적 가능

판단 기준이 명시된 데이터

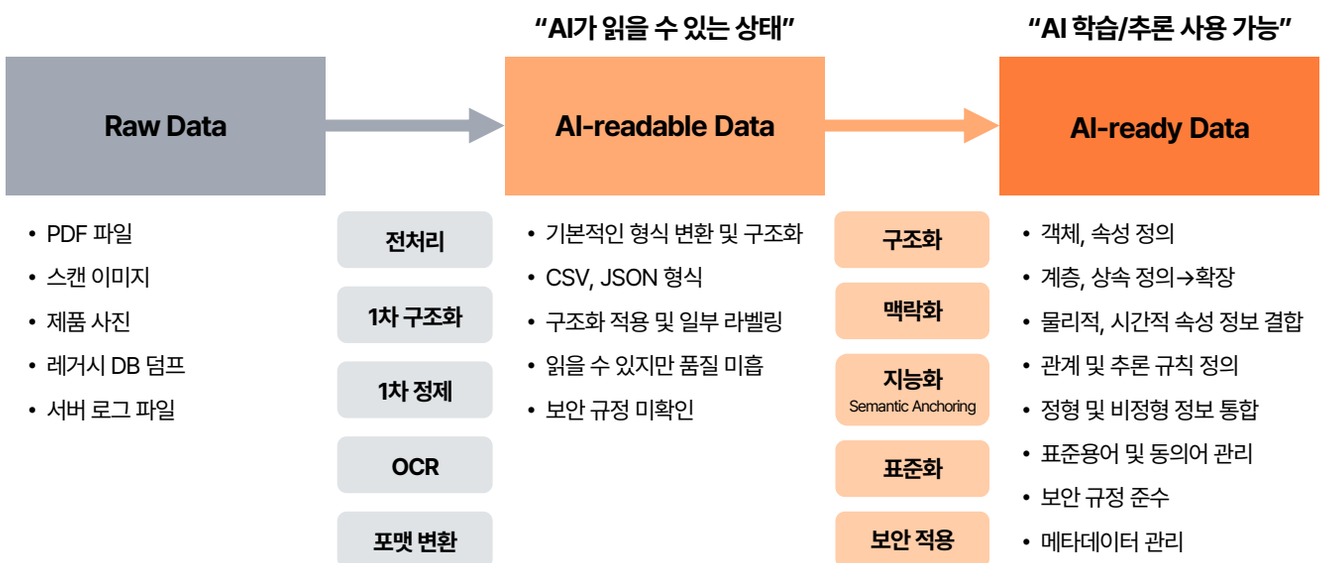
- 기준, 룰, 제약 조건이 데이터 자산으로 관리됨
- 현업의 판단을 AI가 재현 가능

반복 사용과 확장이 가능한 데이터

- PoC용 데이터가 아닌 유스 케이스 확장 가능 구조
- 새로운 AI 에이전트에도 재사용 가능

출처: PwC

Raw Data에서 AI-ready Data로의 전환 과정



출처: PwC

AI-ready Data의 1차적 요건의 중요성

1차적으로 AI-ready Data는 네 가지 핵심 속성을 갖추어야 합니다. 첫째, 계층성입니다. AI가 데이터를 쉽게 검색하고 이해할 수 있도록 계층을 구조화해야 합니다. 둘째, 일관성입니다. 부서 간 다르게 쓰이는 용어를 AI가 해석할 때 오류가 발생하지 않도록 용어 표준이 필요합니다. 셋째, 연계성입니다. 시스템별 채널별 개념이 단절되지 않고 AI가 정확한 이해와 추론을 할 수 있도록 개념 매핑이 중요합니다. 마지막으로 표준성입니다. 문서와 데이터의 작성 방식과 분류 기준이 다르면 AI가 필요한 정보를 찾아오기 어려우므로 표준 분류 기준이 필요합니다.

AI-ready Data가 갖추어야 할 요건들

계층성(완전성)

정보의 계층구조 설명

✓ AI가 이해하려면 해당 계층 구조 및 관계 (부모–자식, 대체 관계, 고객사별 파생 스펙)를 명시적으로 구조화 필요

일관성

동의어 이해

✓ 동일한 용어가 영업, 기술, 품질에서 다르게 쓰이므로, 동의어 사전과 용어 표준(Glossary) 필요

현업에서 자주 쓰는 표현	실제 의미(공식 용어)
재고 눌림	→ 수요 예측오류로 인한 과잉재고
클레임	→ 고객 품질 이슈 (Claim Case)
스펙 이탈	→ 제품 규격값 벗어남 (Spec Out)

연계성

관용어구 관리

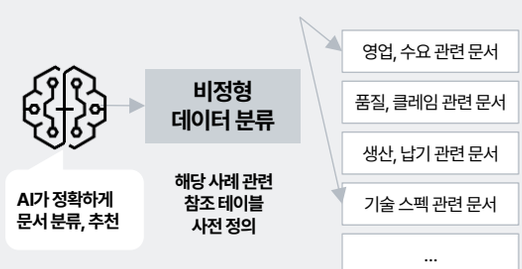
✓ 문서 ,ERP, RDB마다 용어 표현이 달라, AI에게 동일 개념임을 알려주는 용어 매핑 테이블 필요

제조업 예시	
고객사 생산 중단 위험	ERP에는 라인다운 리스크로 기억
매출 기여도 높은 고객	시스템에는 전략고객 플래그로 저장
리드타임 단축 요청	메일에서는 납기 CR(Change Request)로 표현

표준성

표준활용 사례

✓ AI가 문서를 자동 분류하고 필요한 정보를 추출할 수 있는 기준(Topic–Subtopic–Tag) 정의 필요



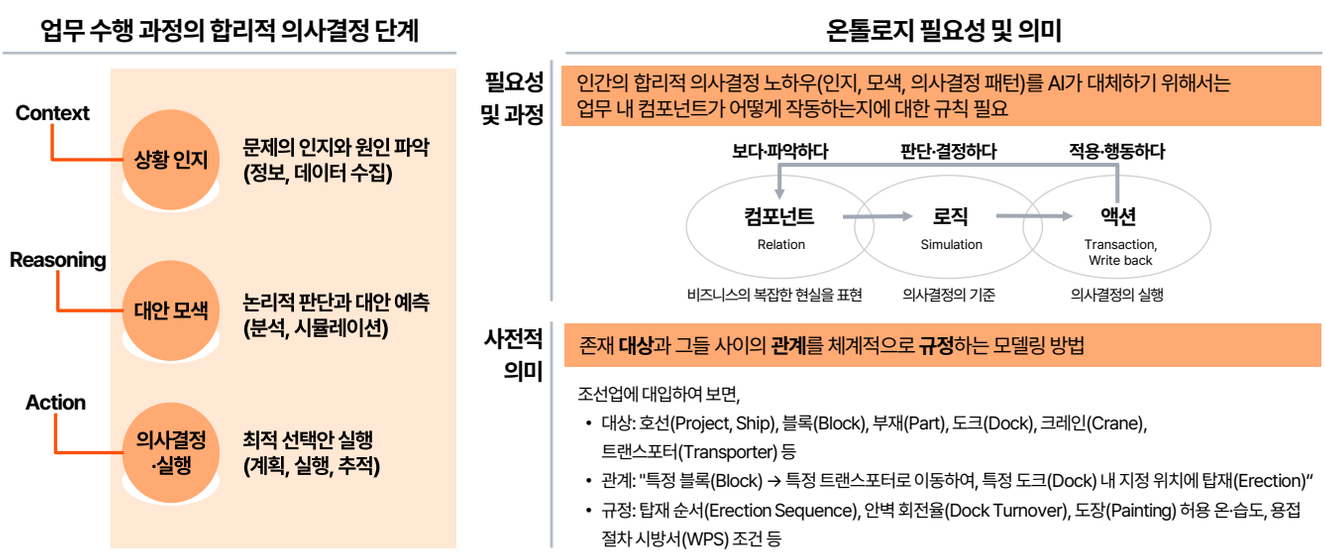
출처: PwC

온톨로지 체계 하에서 완성되는 AI-ready Data

AI-ready Data의 핵심 기술은 바로 온톨로지(Ontology)입니다. 온톨로지는 AI를 위한 지식 지도(Knowledge Map)로서, 예를 들어 반도체 설비라는 개념을 정의할 경우, 이에 연관된 공정, 제품, 품질, 운영 규칙, 재고 정보 등 모든 지식과 관계를 거미줄처럼 연결한 지도라고 할 수 있습니다. 이는 흩어져 있는 지식과 업무 노하우를 AI가 이해할 수 있는 디지털 언어로 정렬해 주는 상세한 내비게이션의 역할을 합니다.

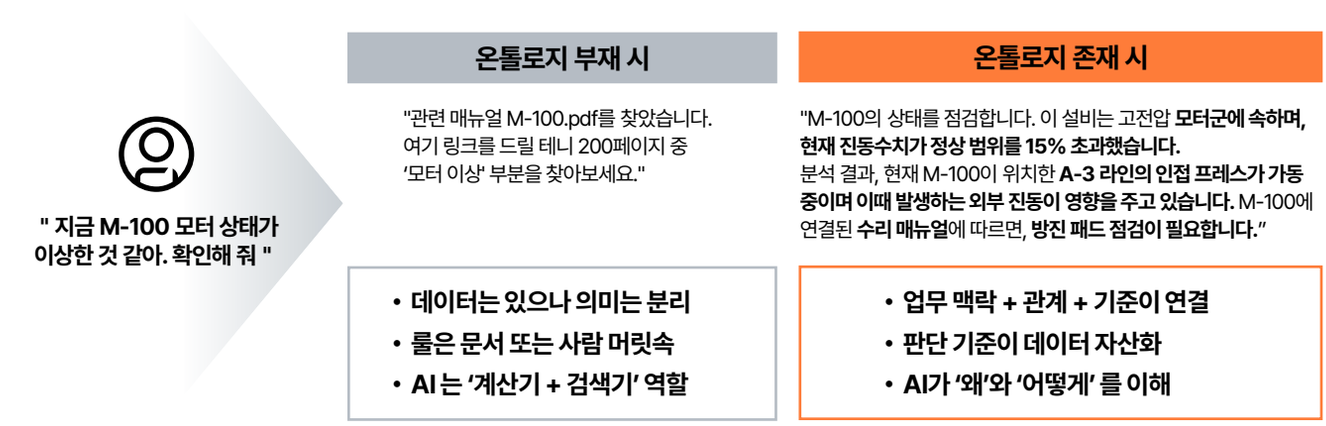
온톨로지는 AI의 역할과 답변 수준을 근본적으로 변화시킵니다. 온톨로지가 없을 경우, AI는 단순한 계산기나 검색기의 역할에 머무를 수밖에 없습니다. 반면, 온톨로지가 구축되어 있으면 데이터에 맥락과 관계, 판단 기준이 유기적으로 연결되어, AI는 비로소 '왜?', '어떻게?'에 대해 스스로 이해하고 설명할 수 있는 파트너로 거듭나게 됩니다.

합리적 의사결정 단계와 온톨로지의 필요성



출처: PwC

온톨로지의 유무에 따라 달라지는 AI 답변 수준



출처: PwC

온톨로지 기반 에이전틱 AI의 작동 방식

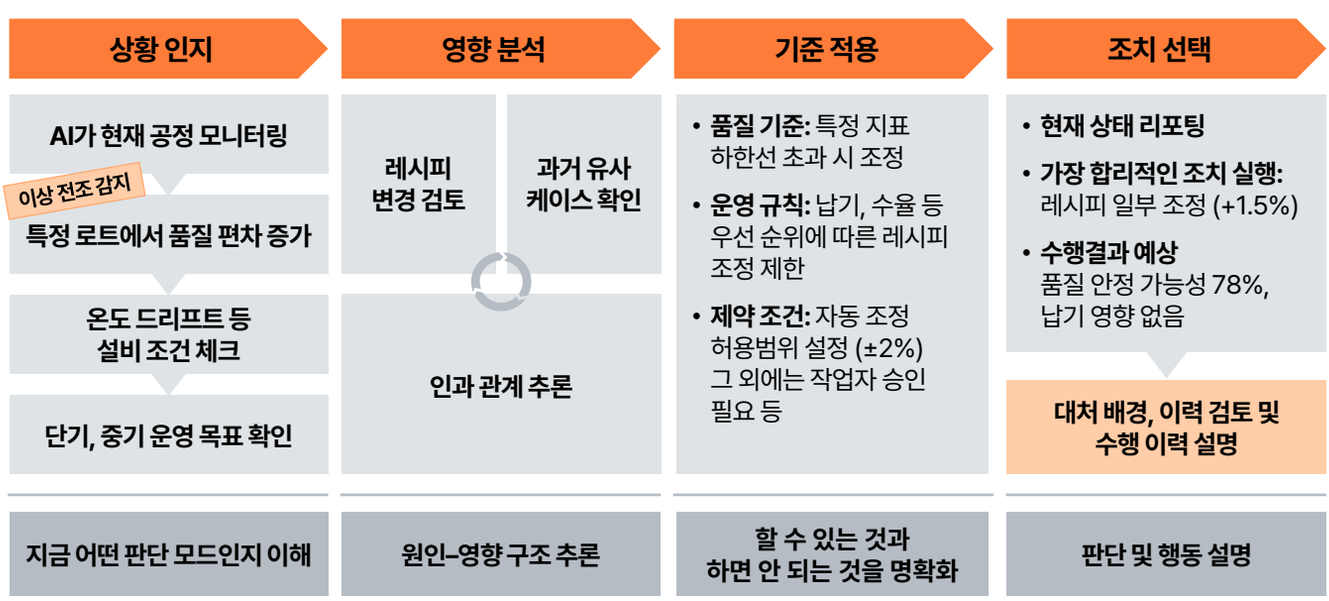
온톨로지 기반의 에이전틱 AI의 작동 방식은 네 단계로 구성되어 있습니다. 1단계는 상황 인지 단계입니다. 마치 24시간 쉬지 않는 베테랑 현장 관리자처럼 AI는 모든 공정 데이터를 실시간으로 모니터링합니다. 그 과정에서 특정 로트에서 품질 편차 값이 미세하게 증가하고 있음을 감지합니다. 이는 이상 징후가 발생하고 있다는 문제의 전조를 인지하는 단계입니다.

2단계는 영향 분석 단계입니다. AI는 즉시 과거의 유사 사례와 현재 설비 조건을 확인하고, 온톨로지에 정의된 인과 관계 지도를 따라 원인을 추론합니다. 예를 들어, 해당 품질 편차가 3번 설비의 미세한 온도 변동과 연관되어 있을 가능성을 도출하고, 이 상태가 지속될 경우 수율에 영향을 미칠 수 있음을 판단합니다. 이를 통해 문제의 원인과 잠재적인 파급 효과를 종합적으로 분석합니다.

3단계는 기준 적용 단계입니다. AI는 해결 방안을 도출하기 위해 온톨로지에 자산화된 판단 기준을 적용합니다. 품질 기준상 특정 지표는 하한선을 초과해서는 안 되며, 현재 운영 우선순위가 납기보다 수율 안정화에 있음을 고려하여 레시피 조정이 가능하다는 결론에 도달합니다. 다만, 자동 조정의 허용 범위는 ±2% 이내로 제한되어 있으므로, 이를 초과하는 조정에 대해서는 작업자 승인이 필요하다는 점까지 명확히 구분합니다. 이 단계에서는 허용 가능한 조치와 허용되지 않는 조치를 분명히 정의합니다.

마지막 4단계는 조치 선택 및 설명입니다. AI는 분석된 모든 정보를 종합하여 가장 합리적인 조치를 선택하고 실행합니다. 예를 들어, "레시피를 +1.5% 조정하겠습니다. 이 조치를 통해 품질 안정 가능성은 78%이며, 납기에는 영향이 없습니다."라고 제안하는 것입니다. 여기서 끝이 아니라, 왜 그러한 판단을 내렸는지 그 배경과 근거까지 인간이 이해할 수 있도록 명확하게 설명해주며, 바로 이 설명 가능성이 에이전틱 AI에 신뢰성을 보강하는 핵심입니다.

온톨로지 기반 에이전틱 AI의 작동 단계

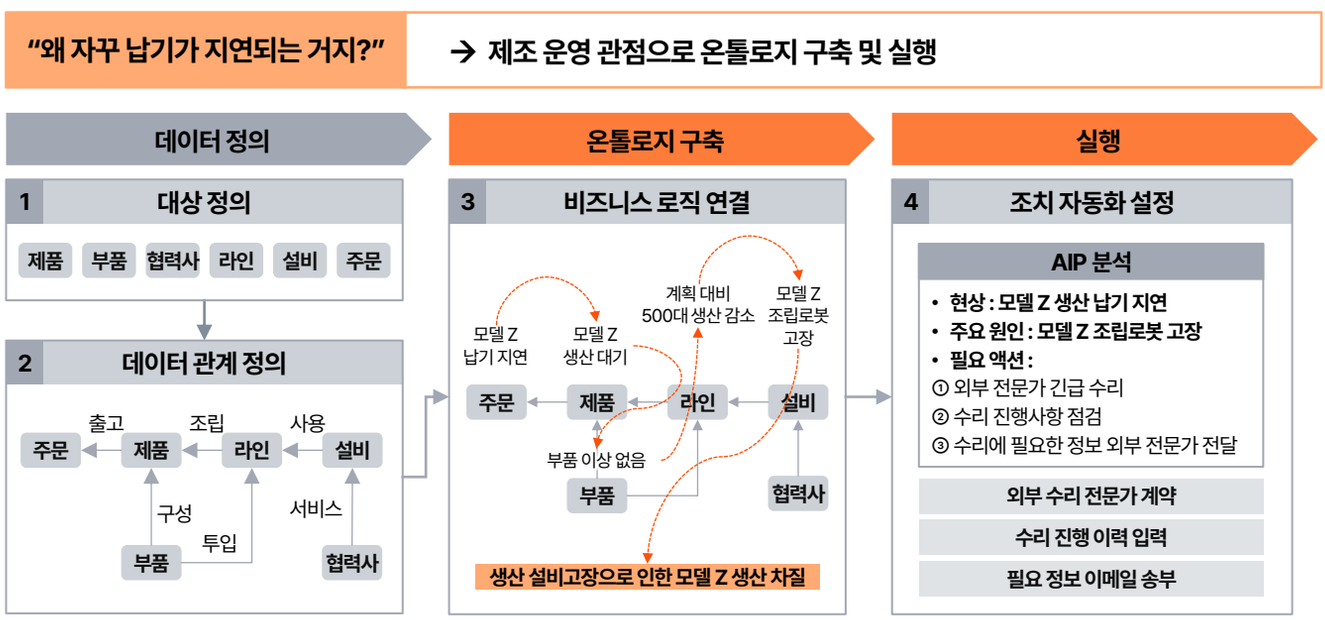


출처: PwC

테슬라의 온톨로지 활용 사례

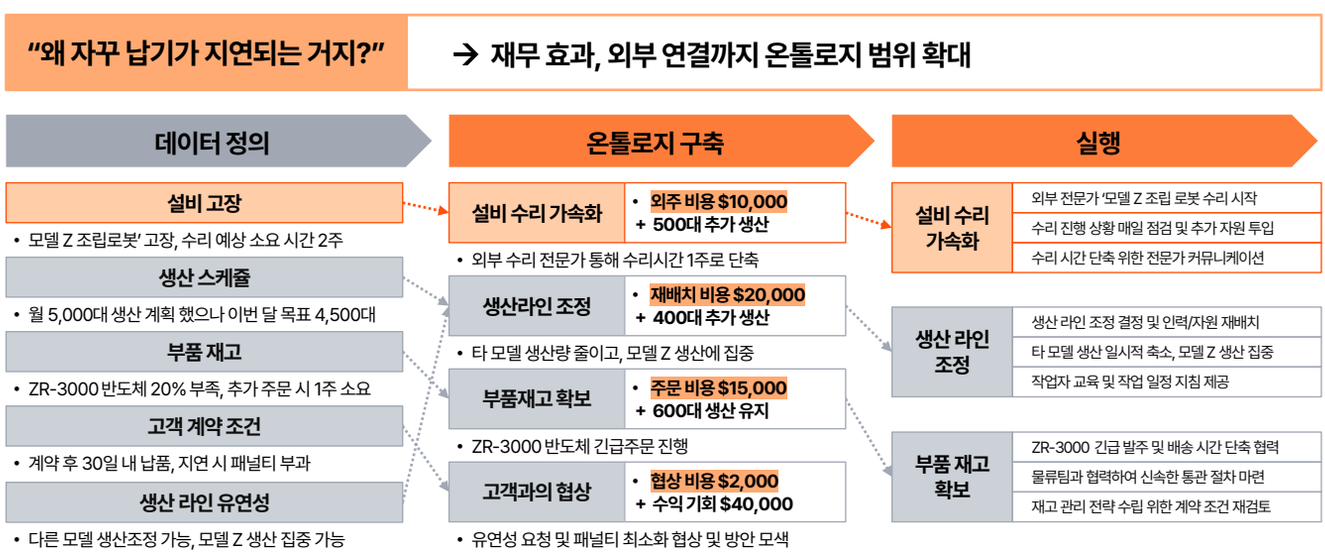
테슬라는 다양한 데이터를 온톨로지에 연결하여 데이터 기반 의사결정에 활용하고 있습니다. 예를 들어 납기 지연 문제를 해결하기 위해 생산, 재고, 계약 등 모든 데이터를 제조 온톨로지로 연결하여 '로봇 고장 → 모델 Z 생산 차질 → 계획 대비 500대 생산 감소 → 납기 지연 위험'처럼 원인과 결과의 맥락을 한눈에 확인할 수 있습니다. 더 나아가 온톨로지 기반으로 수리 시간 단축, 라인 조정, 부품 주문 등 여러 대안 및 시나리오를 소요 비용과 효과(ROI)를 고려하여 최적의 실행 계획을 즉시 확보하고 실행하고 있습니다.

테슬라의 온톨로지 유스 케이스 1



출처: PwC

테슬라의 온톨로지 유스 케이스 2



출처: PwC

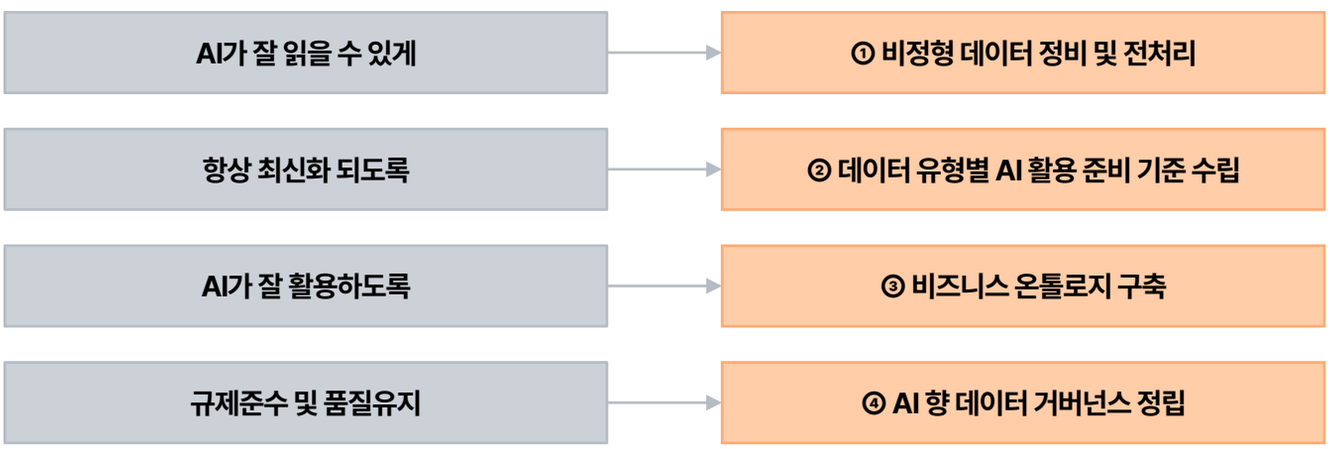
앞으로 준비해야 하는 것

AI-ready Data는 어떻게 준비해야 하는가?

AI-ready Data를 준비하기 위해서는 AI가 바로 읽고, 바로 판단할 수 있는 데이터가 필요합니다. 이를 위해 앞서 제시한 네 가지 핵심 속성(계층성, 일관성, 연계성, 표준성)에 따라 데이터를 구조화하고 운영할 수 있는 체계가 필요합니다. 즉, 다양한 원천 데이터가 AI가 바로 활용할 수 있는 형태로 표준화, 전처리, 거버넌스화되어야 하며, 변경 이력과 품질도 함께 관리되어야 합니다.

이러한 준비 과제는 결국, 비정형 데이터의 정비와 전처리 비정형 데이터 수집 파이프라인 구축, 업무 단위의 아토믹 액션(Atomic Action) 분석, 그리고 AI 향 데이터 거버넌스 체계 수립으로 구체화됩니다. 다음 장에서는 위 4가지를 중심으로, 기업이 실제로 어떤 방식으로 AI-ready Data를 구축해야 하는지 살펴봅니다.

온톨로지 기반 에이전틱 AI의 작동 방식 4단계



출처: PwC

① 비정형 데이터 정비 및 전처리

AI-ready Data를 준비하기 위한 첫번째는 비정형 데이터의 정비와 전처리입니다. AI가 데이터를 쉽게 검색하고 이해할 수 있도록, AI가 의미를 이해하기 어려운 비정형 문서는 요약, 검색, 추론에 활용할 수 있도록 변환해야 합니다. 하지만 비정형 문서 작성 방식은 AI 활용을 저해할 수 있습니다. PwC는 문서 유형별 AI 활용 이슈를 레이아웃 불규칙, 비선형 연결 구조, 비텍스트 요소 포함, 복잡한 표 형식의 4가지로 분류합니다. 예를 들어, 다단 편집이나 글머리 기호 계층 불규칙은 문장 순서 혼란을 초래해 요약과 검색 품질을 저하시킬 수 있으며, 각주, 미주 분리나 별첨 참조 등 비선형 연결 구조는 문맥 이해를 방해하여 근거와 정의 연결 실패를 유발할 수 있습니다. PPT, PDF 캡처 삽입, 도형, 화살표 다이어그램 등 비텍스트 요소는 텍스트 추출이 불가하거나 의미 손실이 발생하며, 과도한 셀 병합, 다단표, 숫자 단위 누락 등 복잡한 표 양식은 구조 파악과 데이터 매칭을 불가능하게 만들어 데이터 이해 시 오류를 유발합니다. 또한 텍스트박스나 위첨자처럼 인간에게는 직관적인 표현도 AI 관점에서는 본문과 분리되거나 순서가 뒤섞여 연결 대상이 불명확해지는 문제가 발생할 수 있습니다.

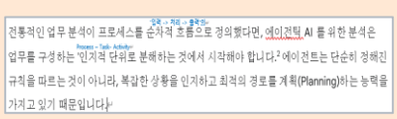
문서 유형별 AI 활용 이슈

AI 활용이 어려운 비정형 문서 작성 방식

구분	문서 포맷 사례	AI에 미치는 영향
레이아웃 불규칙	• 다단 편집 • 글머리 번호·기호, 계층 불규칙	• 문장 순서 혼란 • 요약 및 검색 오류
비선형적 연결 구조	• 엔터키로 줄바꿈 처리 • 각주, 미주 본문 분리 • 별첨 참조, 하단 표 참고 식 연결	• 문맥, 문장 이해 단절 • 근거, 정의 연결 실패
비텍스트 요소 포함	• PPT, PDF 캡처 삽입 • 도형, 화살표 등 다이어그램 삽입 • 그림, 표의 제목 누락	• 텍스트 추출 불가 • 의미 손실 • 그림 표 이해 불가
복잡한 표 양식	• 과도한 셀 병합 • 다단 표 (표 안의 표) • 표 내 숫자 단위 누락	• 구조 파악 불가 • 데이터 매칭 불가 • 데이터 이해 오류

기타

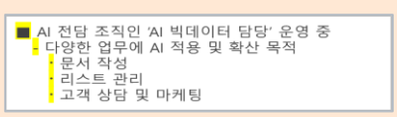
- **텍스트 박스, 위 첨자로 의미 전달**



인간이 읽기에는 매우 직관적이나....

AI 관점에서는 본문과 분리되어 추출되거나 순서가 혼란되어, 해당 코멘트가 어디에 연결되는지 불명확

- **글머리 기호 계층 불규칙**



문서 계층 및 맥락 구분이 어렵고, 문서 참조 시 고유 식별자가 없어 모호한 참조만 가능

출처: PwC

② 데이터 유형별 AI 활용 준비 기준 수립

두번째는 데이터 유형별 비정형 데이터 수집입니다. AI가 활용 가능한 형태로 만들기 위해서는 데이터의 수집, 정제, 연계 기준을 체계적으로 정의하고, 기업 내 데이터 특성에 따라 유형별로 준비해야 합니다. 기업의 데이터는 내부 데이터와 외부 데이터로 구분되며, 내부 데이터는 다시 비정형 데이터와 정형 데이터로 세분화됩니다. 비정형 데이터는 사내 게시판의 공식 문서나 사용자가 업로드한 문서와 이미지를 포함하며, 정형 데이터는 IT 메타 시스템 내 DB 메타와 IT 내부 시스템 내 API 스펙 형태로 존재합니다. 외부 데이터는 DART, KRX, NICE 신용평가 등과의 API 연동 형태로 제공되는 외부기관 연동형과 웹 크롤링 형태로 수집되는 외부 웹사이트형으로 구분됩니다.

데이터 수집 시 고려해야 할 점은 시스템 연계, 최신화, 컴플라이언스의 세 가지 기준으로 분류할 수 있습니다. 먼저, 시스템 연계 관점에서는 공식 문서가 AI 시스템과 연계 가능한지 여부를 확인하고, 사용자 문서는 비정형 포맷 미지원, 권한 제한 등 기술적 제약을 검토해야 합니다. 외부 데이터의 경우에는 API 제공 여부와 연동 방식, 접근 방식 등을 점검하는 기준을 제시합니다. 최신화 관점에서는 내부 데이터의 개정, 폐지, 버전 관리 방안을 수립하고, 일부 내용이 변경되는 경우의 처리 방식을 정의하고, 외부 데이터의 경우 주기적인 업데이트가 이루어지는지 여부를 확인해야 합니다. 마지막으로 컴플라이언스 관점에서는 문서 유형별 AI 학습 가능 여부와 저작권을 검토하고, 공유 범위와 사용 권한을 결정하며 직책자 승인 절차를 마련해야 합니다.

데이터 유형별 수집 시 고려 사항

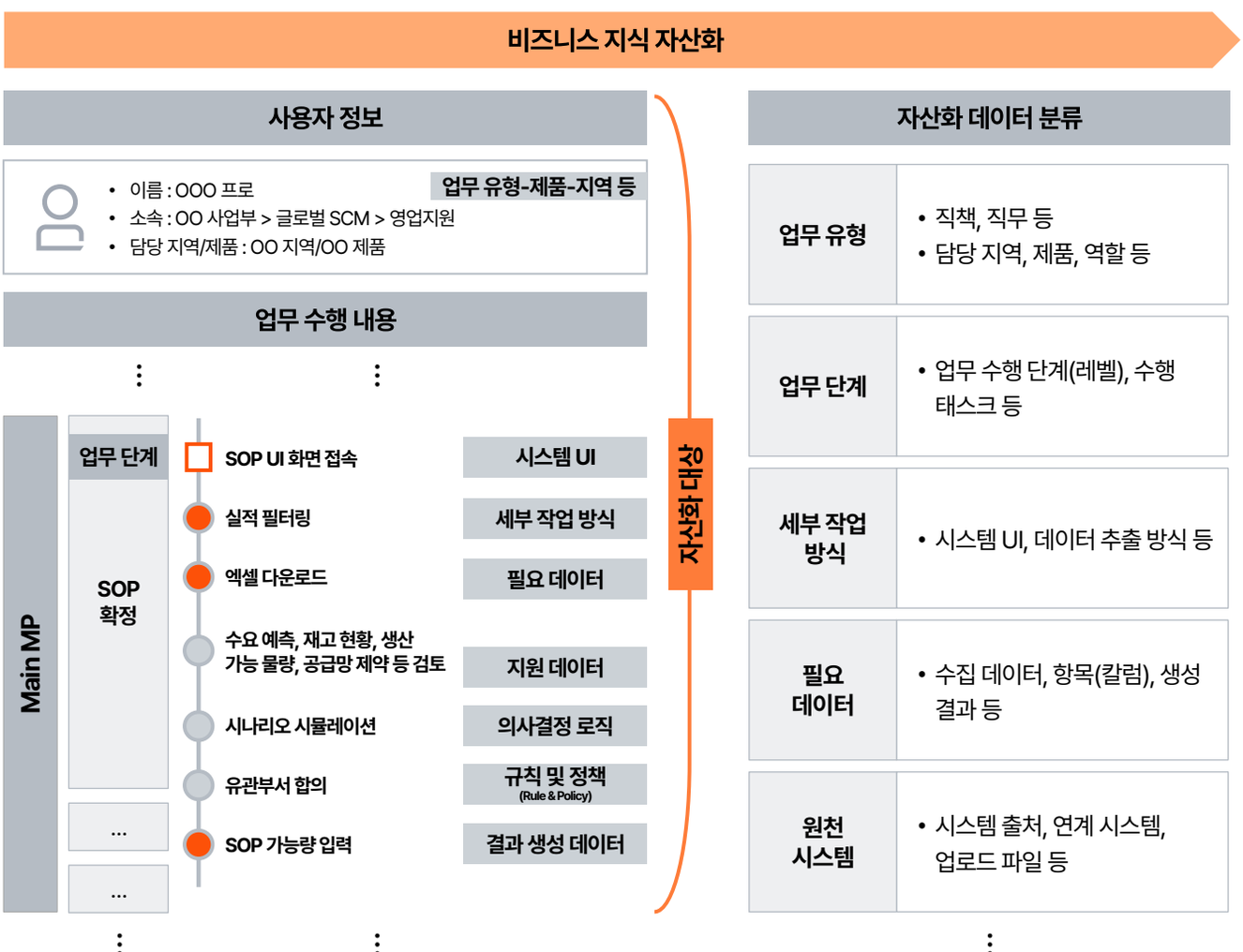
데이터 유형 분류 및 원천 확인				유형별 수집 시 고려 사항		
구분	유형	형태	원천 시스템	시스템 연계	최신화	컴플라이언스
내부 데이터	비정형 데이터	공식 문서	사내게시판, KMS, 그룹웨어	• 연계 가능여부 확인 필요	• 개정, 폐지, 버전관리 방안 • 문서가 부분변경 되는 경우	• 문서유형별 AI학습 가능 여부, 저작권 검토 • 공유범위 결정, 직책자 승인절차 검토
		사용자 문서	사용자 직접 업로드	• 한글 코드 상이, 미지원 포맷 등 기술적 문제 검토		• AI 학습 활용 허용은 사용자의 책임임을 고지
		이미지	이미지 관리 시스템(스캔문서 등)	• OCR로 인식 • 이미지 유형 등 메타정보 필요		• 불필요한 개인정보는 삭제
		음성	콜센터	• STT로 전환		• 개인정보 삭제 필요
	정형 데이터	DB 메타	IT 메타 시스템	• IT 메타 솔루션 연계 가능여부 확인 필요	• IT 내부 최신화 프로세스와 연계	
		API Spec	IT 내부 시스템		• IT 내부 최신화 프로세스와 연계	
외부 데이터	외부기관 연동	API 연동	DART, KRX, NICE 신용평가	• 시스템별 연계 방안 확인 필요	• 주기적 업데이트 수행	
	외부 웹사이트	크롤링	외부 웹사이트 등		• 주기적 업데이트 수행	• 크롤링 가능한 Site 인지 확인 필요

출처: PwC

③ 비즈니스 온톨로지 구축

세번째 단계는 아토믹 액션입니다. 이 단계는 준비된 데이터를 연결하고 맥락화하여 AI가 이해할 수 있는 형태로 구조화하기 위해 필요합니다. 이를 위해 사용자 정보와 업무 수행 내용을 업무 유형, 업무 단계, 세부 작업 방식, 필요 데이터, 원천 시스템 등으로 분류하여 지식 자산화를 진행합니다. 예를 들어 글로벌 SCM 영업지원 사원의 경우, 직책, 담당 지역, 제품군 등을 기반으로 업무 유형을 정의하고, SOP 기반의 UI 접속, 실적 필터링, 엑셀 다운로드, 수요 예측 등 각 업무를 최소 실행 단위로 분해합니다. 이러한 요소들은 모두 자산화 대상이 되며, AI는 이를 바탕으로 업무를 단순히 이해하는 수준을 넘어 실행 가능한 절차로 재구성할 수 있습니다. 즉, 최소 실행 단위 기준으로 시스템, 데이터, 업무를 연결하면 지식이 AI가 처리 가능한 구조로 정리되어, AI-ready Data에 가까워집니다.

지식 자산화 과정



출처: PwC

④ AI 향 데이터 거버넌스 정립

마지막으로, 데이터 관리와 활용의 근본이 되는 명확한 전사 데이터 거버넌스 체계 하에 AI를 활용해야 합니다. AI 데이터 거버넌스와 운영 고려사항은 인력 및 조직, 프로세스 및 정책, 플랫폼 및 기술의 세 가지 영역으로 나뉘며, 구성 요소별로 운영 원칙과 책임 범위를 명확히 해야 합니다. 모든 활동은 투명성, 오너십, 안전성, 윤리성을 지향하며, AI 판단 근거의 품질을 보장하기 위한 기반이 됩니다. 이를 통해 AI 판단 결과에 대한 신뢰를 확보하고, 결과에 대한 관리 주체의 책임을 명확히 할 수 있습니다.

AI 향 데이터 거버넌스 및 운영 고려사항

출처: PwC

전문가의 일하는 방식 및 업무 처리 방식의 자산화

이상 네 가지와 더불어 온톨로지 체계의 AI-ready Data를 만들기 위해서는 먼저 '어떤 목적으로 AI를 쓸 것인가'하는 명확한 목표를 정의해야 합니다. 그리고 그 목표를 달성하기 위해 전문가들이 '어떤 순서와 기준으로 문제를 인지하고, 대안을 찾고, 의사결정을 하는지' 시나리오를 분석하고 이를 바탕으로 용어와 개념을 설계해야 하며 객체, 관계, 규칙을 정의하여 온톨로지 모델을 설계하며, 실제 데이터를 변환하여 저장하고 관리하는 순서로 온톨로지를 구축해야 합니다. 더불어, AI의 의사결정 결과를 평가하고 다시 학습시켜 온톨로지를 점차 더 똑똑하게 만들어가는 자산화 과정 역시 매우 중요한 항목입니다.

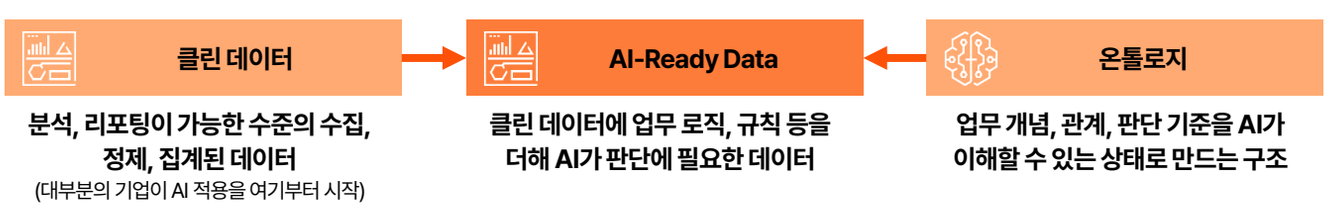
에이전틱 AI의 경쟁 우위는 범용 모델이나 솔루션을 얼마나 빨리 도입하는지로 결정되지 않습니다. 핵심은 지식이 체계적으로 수집, 정제되고, 업무 개념 그리고 개념 간의 관계가 정의되고, 판단 기준이 AI가 이해할 수 있도록 정립될 때 비로소 AI는 조직의 업무를 수행하는 에이전틱 AI로 진화한다는 것입니다.

결국 에이전틱 AI 시대에 승리하는 기업은 좋은 AI 모델 사용에 집중하는 기업이 아니라, 조직의 지식을 구조적으로 자산화하는 데 집중하는 기업일 것입니다.

온톨로지 기반 시나리오 정의 및 온톨로지 상세 설계

출처: PwC

구조적 자산화



출처: PwC

엔터프라이즈 에이전트의 구현

엔터프라이즈 에이전트 도입을 위한 고려사항

에이전틱 AI 시대의 도래

AI 기술은 이제 단순한 자동화를 넘어, 스스로 판단하고 계획하며 실제 시스템을 조작해 업무를 수행하는 에이전틱 AI 시대로 빠르게 이동하고 있습니다. 특히 엔터프라이즈 환경에서 에이전틱 AI가 갖는 의미는 챗봇 고도화가 아니라, 기업이 일을 수행하는 방식 자체, 즉 실행 구조(Operating Model)를 재설계하는 변화입니다. 기존 IT 신기술 도입이 늘 기업 경쟁력을 높였던 것은 아닙니다. ERP는 프로세스 표준화와 데이터 일원화를 통해 업무를 시스템 기준으로 재정의하면서 조직, 권한, 통제 체계 변화까지 동반했고, 클라우드는 인프라 조달 방식을 바꾸며 운영 체계(보안, 거버넌스, 비용 통제)의 변화를 요구했습니다. 반면 RPA나 분석(BI, 고급분석)은 부분 최적화나 분석과 실행의 단절로 인해 기대 효과가 제한되는 경우가 있었습니다.

마이크로소프트 CEO 사티아 나달리가 "코파일럿은 하나의 앱이 아니라 일하는 방식의 새로운 인터페이스"라고 말한 이유도 여기에 있습니다. 결국 기업은 인간과 더불어 '디지털 워커'로 이루어진 새로운 운영 모델을 준비해야 하며, 이를 준비하지 못한 상태에서 에이전트를 확산하면 혁신이 아니라 운영 리스크가 됩니다. 따라서 어떤 모델을 쓰는가를 논하기보다, 엔터프라이즈 에이전트를 실제로 구현하고 확산 및 운영하기 위한 필수 조건(프로세스, 데이터, 거버넌스, 플랫폼, 개발 체계)을 제시하고자 합니다.

실행 가능한 체계 수립

엔터프라이즈 에이전트의 성공 조건은 LLM의 성능이 아니라 업무가 실제로 돌아가게 만드는 실행 체계입니다. 업무 프로세스는 자동화 가능한 구조(Automation-ready) 관점에서 먼저 재설계되어야 합니다. 에이전트는 표준 흐름에서는 강점을 보이지만, 예외 상황에서 오류가 발생하기 쉽고 이때 예외 처리 기준, 승인 경로, 책임자가 불명확하면 자동화 수준이 높아질수록 운영 리스크가 확대됩니다. 따라서 현행 프로세스를 그대로 자동화하기보다, 표준 흐름, 예외 케이스, 승인 조건, 책임 분기를 사전에 정의해 '운영 가능한 프로세스'로 정리하는 것이 선행되어야 합니다. 성과 측정 역시 시간 절감 중심에서 벗어나 운영 품질과 통제력을 반영하는 KPI로 재정의해야 합니다.

엔터프라이즈 환경에서 신뢰를 만드는 지표는 오류율, 재작업률, 리드타임, 응답 품질, 규정 준수(컴플라이언스)와 같이 정확하고 안전하게 반복 가능한가를 보여주는 항목입니다. 에이전트는 실행 주체이기 때문에 빠른 처리보다 정확, 안전, 재현 가능성이 본질이며, 이를 측정·관리하지 못하면 확산 단계에서 품질 편차와 통제 붕괴가 발생합니다. 또한 변화관리는 부수 과제가 아니라 확산의 전제 조건입니다. 에이전트 도입 실패가 기술이 아니라 인간과 조직에서 비롯된다는 점처럼, 에이전트가 업무를 수행할수록 인간의 역할은 단순 처리자에서 감독자, 검증자, 설계자(예외 처리 포함)로 이동합니다. 대체 관점이 아니라 재배치와 고부가 업무 전환 관점에서 운영 모델을 재구성해야 합니다.

데이터와 지식 기반 정비는 옵션이 아니라 필수 인프라입니다. 마스터 데이터 정합성, 메타데이터·정의 표준화, 문서·규정·업무 룰의 구조화가 미흡하면 자동화가 고도화될수록 작은 오류가 즉시 잘못된 판단과 실행으로 증폭되어 신뢰를 붕괴시킵니다. 에이전트는 말이 아니라 행동하기 때문에 데이터와 지식의 품질 문제는 곧 운영 사고로 연결됩니다. 권한, 보안, 감사 체계(Audit Trail)도 반드시 함께 설계되어야 합니다.

에이전트가 인간을 대신해 판단하고 실행하는 순간, 결정 근거(왜), 실행 기록(무엇을), 책임 경계(누가 승인했는지)가 남지 않으면 운영·감사·규제 대응이 불가능합니다. 역할·상황 기반 IAM, 민감정보·대외 발송·금액 임계치 기반 승인 게이트, 실행 근거와 로그를 남기는 권한·보안·감사 체계가 함께 구현되어야 합니다.

마지막으로 전사 확산의 승부처는 플랫폼 운영 통제(Observability)입니다. 확산 단계에서 가장 자주 발생하는 실패는 비용 폭증과 운영 장애이며, 호출 비용·지연시간·실패율 같은 운영 KPI, E2E 모니터링, 장애 대응 런북(Runbook)이 없으면 규모가 커질수록 통제가 무너집니다. 결국 엔터프라이즈 에이전트의 성공 기준은 PoC 성공이 아니라, 운영 가능한 구조로 표준화되고 통제된 상태에서 확산 가능한가에 달려 있습니다.

엔터프라이즈 에이전트 구현 및 운영 전략

효과적 구현을 위한 접근 방안

엔터프라이즈 에이전트 구현은 단일 모델, 단일 앱 개발이 아니라, 사용자 접점부터 실행·통제·운영까지 연결되는 플랫폼 아키텍처를 설계하는 것입니다.

예를 들어, 사용자 경험 계층은 다양한 채널과 사용자 역할에 따라 권한 기반의 일관된 상호작용을 제공하고, 오케스트레이션 계층은 요청을 업무 단위로 분해하여 적절한 에이전트, 툴, 워크플로우를 조합함으로써 실제 운영 흐름을 통제합니다. 추론·정책 계층은 기업 정책과 권한 체계를 반영한 통제된 의사결정을 수행하며, 실행 계층은 ERP, CRM, ITSM, API 등과 연계해 판단을 실제 업무 실행으로 연결하되 보안·감사 체계와 결합되어야 합니다. 지식 계층은 문서, 정형 데이터, 관계정보 기반 컨텍스트를 제공해 판단 근거를 형성하고, 데이터·연계 계층은 내외부 데이터 연결과 품질 관리를 통해 최신 정보를 활용하게 합니다. 여기에 보안·통제 계층이 권한, 프라이버시, 오남용 리스크를 통제하고, 관측성 계층은 실행 상태, 성과, 비용을 가시화하고 운영·감사·개선을 가능하게 합니다.

결국 엔터프라이즈 에이전트 구현은 모델을 적용한 챗봇 구축이 아니라, 실행과 통제까지 포함한 8가지 계층으로 구성된 운영 플랫폼을 구축하는 일이며, 이러한 구조가 갖춰져야 에이전트가 증가하더라도 운영 복잡도와 리스크를 억제한 상태에서 표준화된 확산이 가능합니다.

AI 에이전트 구성 핵심 8대 계층

1	**사용자 경험 계층** Experience Layer	사용자와 AI 에이전트가 상호작용하는 접점으로, 채널, 역할, 권한에 따라 일관된 경험 제공
2	**오케스트레이션 계층** Orchestration Layer	사용자 요청을 업무 단위로 분해하고, 적절한 툴, 워크플로우를 조합해 전체 실행 흐름 통제
3	**추론·정책 계층** Reasoning·Policy Layer	AI 에이전트의 판단 기준을 정의하며, 정책, 규칙, 권한을 반영한 통제된 의사결정 수행
4	**실행 계층** Tool·Action Layer	AI 에이전트의 판단 결과를 실제 업무 실행으로 연결하는 역할 담당
5	**지식 계층** Knowledge Layer	AI 에이전트가 판단 시 근거가 되는 지식과 업무 컨텍스트 관리
6	**데이터·연계 계층** Data·Integration Layer	기업 내외부 데이터와 시스템을 연결하여 AI 에이전트가 최신의 정확한 정보 활용 지원
7	**보안·통제 계층** Security·Guardrail Layer	AI 에이전트 사용 시 발생할 수 있는 보안, 프라이버시, 오남용 리스크를 통제하는 안전 장치
8	**관측성 계층** Observability Layer	AI 에이전트의 실행 상태, 성과, 비용을 가시화하고 운영, 감사, 개선 실행

출처: PwC

구현을 위한 기술 아키텍처

엔터프라이즈 에이전트는 관찰–추론–계획–행동–도구 활용으로 구성된 에이전트 코어를 중심으로, 오케스트레이션(워크플로우, ReAct), 통신 표준(MCP, A2A), 데이터 인프라(RAG), 서빙 구조(클라우드, 온프레미스), 보안 체계가 통합된 아키텍처로 구현되어야 합니다.

여기서 모델 선택의 핵심은 특정 모델 우열이 아니라 속도, 정확도, 비용을 지속적으로 관리하는 운영 체계입니다. 속도는 압축, 양자화, 캐싱 등으로 최적화하고, 정확도는 목적 기반 평가와 모델 조합 전략으로 확보하며, 비용은 인프라 구조와 호출 패턴을 반영한 동적 확장과 통제로 관리합니다. 즉, 모델은 선택의 대상이 아니라 운영 관리의 대상입니다.

오케스트레이션은 단순 연결이 아니라 실제 업무 수행 로직입니다. 워크플로우가 트리거와 제어 흐름을 구성한다면, ReAct는 관찰 결과에 따라 다음 행동을 조정하는 반복 실행 패턴으로 업무 완료까지 이어집니다. 통신 표준 측면에서 MCP는 에이전트와 외부 도구를 연결해 실행 능력을 확장하고, A2A는 에이전트 간 협업을 가능하게 하여 다중 에이전트 워크플로우를 형성합니다. 이러한 표준이 없으면 에이전트 증가가 곧 운영 복잡도와 장애 리스크 증가로 이어집니다.

데이터 인프라는 문서 중심(Vector RAG)과 관계 중심(Graph RAG)의 특성을 고려해 혼합 전략으로 설계하는 것이 일반적이며, 서빙 구조는 클라우드의 확장성과 온프레미스의 통제권을 균형 있게 선택해야 합니다. 또한 프롬프트 인젝션, 데이터 유출, 권한 상승과 같은 위협에 대응하기 위해 입력 검증, 최소 권한 통제, 로깅 및 감사 체계를 포함한 보안 구조가 필수입니다. 결국 엔터프라이즈 에이전트 구현의 본질은 개별 기술의 조합이 아니라, 보안·감사·관측이 내장된 통합 아키텍처를 구축하는 것이며, 이러한 구조가 확보되어야만 안정적인 확산이 가능합니다.

AI 에이전트 구성 기술적 요소

출처: PwC

개발 옵션 및 확산: 자체 개발 vs 상용 에이전트

클라우드 기반 인프라 전환은 초기에는 운영비(Opex) 절감 효과를 가져올 수 있으나, 실제 운영 단계에서는 워크로드 특성(상시·변동), FinOps 성숙도, 데이터 전송(Egress) 비용, AI·GPU 사용량 증가 등에 따라 비용이 다시 상승하는 경우가 빈번하게 나타납니다.

특히 에이전트 기반 서비스가 확대될수록 호출 빈도와 처리량이 기하급수적으로 증가하여 모델과 실행 환경에 대한 통제권을 외부에 의존하는 구조에서는 비용 예측 가능성이 급격히 낮아집니다. 결국 에이전트와 모델 운영 주도권을 잃는 것은 비용 통제력을 잃는 것이며, SaaS, 패키지형 에이전트에 전적으로 의존할 경우 확산 단계에서 비용과 성능을 동시에 관리하기 어려워집니다.

자체 구축은 에이전트 로직, 추론, 워크플로우에 대한 완전한 통제와 함께 멀티 에이전트 및 오케스트레이터 설계를 자유롭게 구성할 수 있다는 장점이 있습니다. 또한 IAM, 승인 체계, 감사 로그, HITL(Human-in-the-loop)과 같은 통제 구조를 내재화할 수 있으며, ERP, ITSM 등 운영 시스템과의 깊은 연계나 복합 트랜잭션, 롤백 처리와 같은 요구사항을 안정적으로 충족할 수 있습니다. 반면, 상용 플랫폼은 표준화된 패턴과 커넥터를 기반으로 빠른 구축과 낮은 초기 부담, 기본 운영 기능을 제공한다는 측면에서 효율적이지만, 핵심 업무로 확장될수록 기업 고유의 정책, 권한, 감사 요구를 반영하기 어렵고 벤더 종속과 장기적 비용 통제 한계가 발생할 수 있습니다.

따라서 현실적인 접근은 양자택일이 아니라 하이브리드 전략입니다. 범용 업무와 공통 기능은 상용 플랫폼을 활용해 도입 속도를 확보하되, 권한, 감사, 프로세스 경쟁력과 직결되는 핵심 영역은 내부 통제 하에 구현하는 것이 가장 합리적입니다. 엔터프라이즈 에이전트는 결국 기술 도입이 아니라 운영 체계의 문제로 귀결되며, 기업은 표준 API, 정책 코드화, 테스트·관측 체계, 런북, 데이터·지식 관리 등 어떤 영역을 내부 자산으로 확보할 것인지 정의한 뒤 단계적 확산 전략을 수립해야 합니다.

자체 개발 vs 상용 에이전트 활용

	자체 개발	상용 에이전트 활용
개발 유연성	• 에이전트 로직, 추론 방식, 워크플로우 완전 통제 • 멀티 에이전트, 오케스트레이터 자유 설계	• 표준화된 에이전트 패턴 즉시 사용 • 설정 기반 커스터마이징
	➢ LangGraph / LangChain, CrewAI, Temporal / Airflow	➢ Microsoft Copilot Studio, Salesforce Einstein GPT
보안/권한	• IAM·SSO·권한, 승인 로직 완전 내재화 • 감사, 로그, HITL 자유 설계	• 기본 보안·권한 프레임 제공, SaaS 보안 인증 활용
	➢ Keycloak / OPA(Open Policy 에이전트), Secrets Manager	➢ Copilot Studio Approval Flow, Salesforce Shield
실행(Action) 시스템 연계	• ERP, ITSM, 운영시스템과 깊은 연계 • 복합 트랜잭션, 롤백 설계	• 표준 커넥터 기반 빠른 연계
	➢ FastAPI / OpenAPI, Kafka / Event-driven Architecture	➢ SAP BTP Integration, ServiceNow Integration Hub
운영, 확장, 비용	• 벤더 락인 최소화, 장기 TCO 통제 가능	• 초기 구축 비용 최소화, 운영 부담 낮음
	➢ OpenTelemetry, Prometheus / Grafana	➢ Azure AI Studio Monitoring, Salesforce Analytics

출처: PwC

엔터프라이즈 에이전트 성공의 관건

AI 에이전트는 더 이상 고도화된 챗봇이 아니라, 조직의 업무를 실제로 수행하는 디지털 워커입니다. 따라서 성공의 기준 역시 모델 성능이 아니라 아키텍처와 거버넌스에 있습니다. 우수한 LLM을 적용하는 것만으로는 안정적인 운영이 이루어지지 않으며, 추론, 오케스트레이션, 툴, 지식, 보안, 관측성이 결합된 실행 구조가 갖춰질 때 비로소 조직의 업무 체계 안에서 신뢰 가능한 주체로 작동합니다. 또한 대부분의 기업에게 현실적인 선택은 전면 자체 개발이나 전면 상용 의존이 아닌 하이브리드 구조이며, 범용 영역은 활용하되 핵심 업무에 대한 통제권은 내부에 유지해야 품질, 비용, 리스크를 장기적으로 관리할 수 있습니다. 결국 AI 에이전트는 단순한 기술 도입 과제가 아니라 기업 경쟁력의 근간이 되는 운영 인프라입니다.

이를 내재화하지 못한 조직은 업무 판단과 실행의 주도권을 외부 플랫폼에 의존하게 되고, 이는 곧 프로세스 경쟁력의 약화로 이어질 수 있습니다. 따라서 기업은 PoC 중심의 단발성 적용을 넘어, 정책, 보안, 관측이 포함된 표준 에이전트 플랫폼을 먼저 구축하고, 프로세스, 데이터, 지식, 개발 체계를 통합한 운영 기반 위에서 자율성을 단계적으로 확대해야 합니다. 향후 기업의 디지털 전환은 "어떤 AI를 사용하는가"가 아니라 "AI가 어떻게 조직 안에서 일하도록 설계되어 있는가"로 평가받게 될 것입니다. 에이전트 인프라를 전략 자산으로 구축하고 운영 역량을 내재화한 기업만이, 자동화를 넘어 지속적으로 학습하고 개선되는 운영 체계를 확보하게 됩니다. 에이전틱 AI의 도입은 효율화 프로젝트가 아니라 조직이 스스로 진화할 수 있는 구조를 설계하는 일이며, 이 구조를 먼저 갖춘 기업이 미래 경쟁에서 우위를 선점하게 될 것입니다.

AI 에이전트 개발 시 핵심 고려 사항

통제 범위 책임 구조	• 실행(Action), 권한 변경, 대외 발송, 결재 연계가 포함될수록 통제 가능한 구조(자체 개발 또는 하이브리드) 필요 • 에이전트 전략은 기술 선택이 아니라 '책임 범위 설계'의 문제
보안·권한 규제 대응	• 재무, 인사, 품질, 운영 등 규제 민감 업무에 대해서는 권한, 승인, 감사 구조를 직접 통제할 수 있어야 함 • SaaS 상용 에이전트는 기본 프레임은 제공하나 세부 정책, 예외 처리에는 한계 존재
확장성 및 기술 수용성	• 상용 에이전트는 초기 확산에는 유리. 단, 기업 특화 고유 기능 개발 및 다양한 기술 수용성 측면의 한계 존재 • 오픈소스 기반 기술 스택과 혼합된 하이브리드 기술 아키텍처 필요
운영 비용 FinOps	• 상용 에이전트는 초기 비용은 낮으나, 사용량 증가 시 비용 급증 위험(에이전트, GPU 등 리소스 과금) • 하이브리드 구조는 모델, 툴, 관측 비용을 전략적으로 분산하고 최적화 가능
기술 자산화	• 핵심 업무 로직은 장기적으로 내재화 또는 하이브리드로 관리하는 것이 유리(인소싱) • 범용 기능은 상용 에이전트 활용으로 효율화

출처: PwC

에이전틱 AI 기반 전사 HR 제도의 혁신

인간과 AI 에이전트의 협업

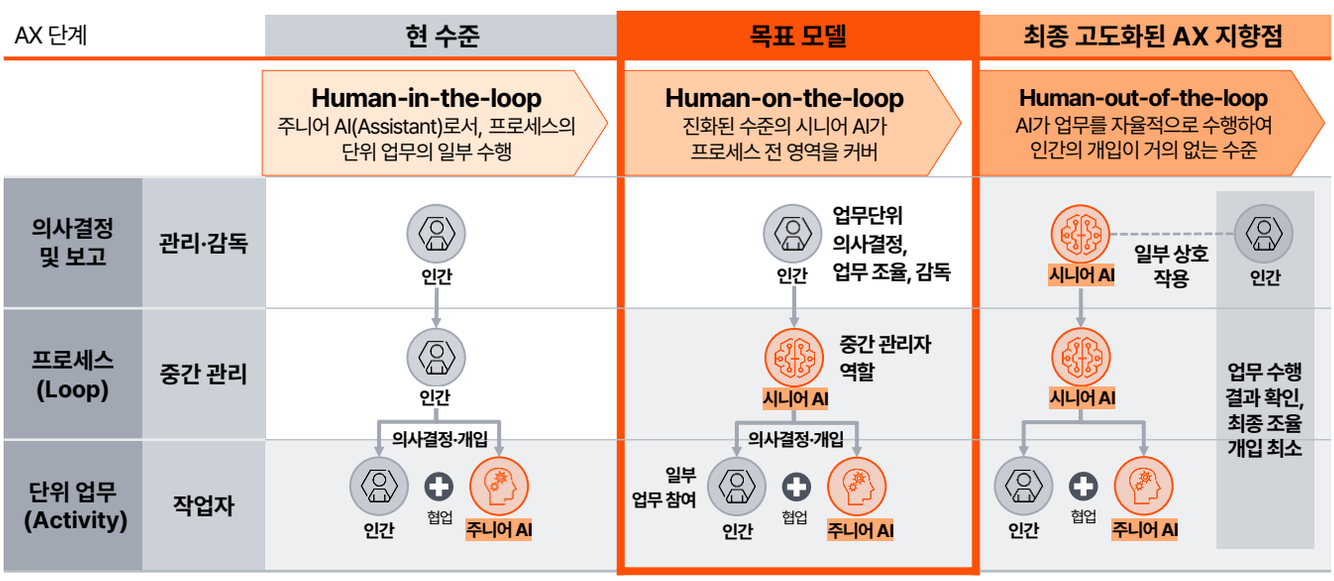
협업 모델의 진화

현재 많은 기업에서 AI는 일부 업무를 보조하는 수준에 머물러 있습니다. 그러나 AI가 고도화됨에 따라, 인간과 주니어 AI가 협업하고 시니어 AI가 중간 관리자 역할을 하는 모델이 가능해질 것이며, 최종적으로는 AI가 인간의 업무를 자율적으로 수행하여 인간의 개입은 최소화되는 모델을 지향하고 있습니다. 이러한 변화는 기술 진보만이 아니라, 의사결정 구조와 책임 배분 방식의 근본적인 변화를 의미합니다.

글로벌 빅테크의 AX 추진 사례

PwC 분석에 따르면 글로벌 빅테크들은 에이전틱 AI를 활용한 협업 체계를 구축하고 있습니다. 아마존은 HITL(Human-in-the-loop) 모델을 보유하여 정형화된 업무를 AI로 자동화하고, 중간 관리 조직을 축소하는 방식으로 조직 슬림화를 추진했습니다. 마이크로소프트는 HOTL(Human-on-the-loop) 모델에서 인간 관리자가 '에이전트 보스'로서 에이전틱 AI와 인력이 결합된 하이브리드 팀을 운영하는 모델을 도입했습니다. 이 사례들은 AX가 기술 도입을 넘어 의사결정 구조, 관리 방식, 조직 역할 전반의 재편을 동반하는 전략적 변화임을 보여줍니다.

인간과 AI 에이전트의 협업 모델



출처: PwC

성공적 AX를 위한 실행 전략

협업 모델의 전환을 성공적으로 추진하기 위해서는 AI-인간 협업 프로세스를 전제로 한 조직 재설계, 운영 모델 재설계, 그리고 변화관리를 병행해야 합니다. 세부적으로 보면, 첫째, 조직 재설계에 있어서는 AI 에이전트 기반으로 운영 효율을 높이기 위해 AX CoE(Center of Excellence)의 역할을 고도화하고, 중복 업무 정리하여 조직을 슬림화하고, 비핵심 단순 반복 업무는 SSC(Shared Service Center)로 전환해야 합니다. 둘째, 조직 운영 모델의 변화가 요구됩니다. 이때 AI와 인간의 직무 체계와 평가 기준을 재정의하고 에이전트 보스를 위한 리더십 역량을 강화해야 합니다. 셋째, 변화가 일회성에 그치지 않도록 지속가능한 변화관리를 위해 업무와 직무의 변화 방향을 공유하고, 인재 역량, 거버넌스 및 성과관리 체계를 구축해야 합니다.

성공적인 AX를 위한 3가지 실행 전략

운영 효율을 높이는 **AI 에이전트 기반 조직 설계**	• AX CoE 조직 강화 • 부문간, 사업부간 중복 업무를 통폐합하여 조직 슬림화 및 최적화 • 비핵심 단순 업무의 SSC 전환 • 유휴 인력의 재배치 및 리스킬링
일하는 방식, 책임, 관리방식 개선을 위한 **조직 운영 모델 정립**	• AI-인간 협업 직무 체계 재설정 • 에이전트 보스를 위한 리더십 역량 재정립(디지털 업스킬링 지원 등) • 조직 KPI를 인간에서 인간 + AI 중심으로 변화 • 인간과 AI 구성원의 평가기준 재정립
성공적이고 지속적인 AX를 위한 **변화관리 실행**	• 업무 방식, 직무 역할 변화 방향성 등 공유 • 전사 공감대 형성 • 전사 AI 변화관리 체계 구축 • AI 기반 인재 역량 강화 • AI 거버넌스 및 성과관리 체계 정립

출처: PwC

AI 에이전트 기반 조직 설계

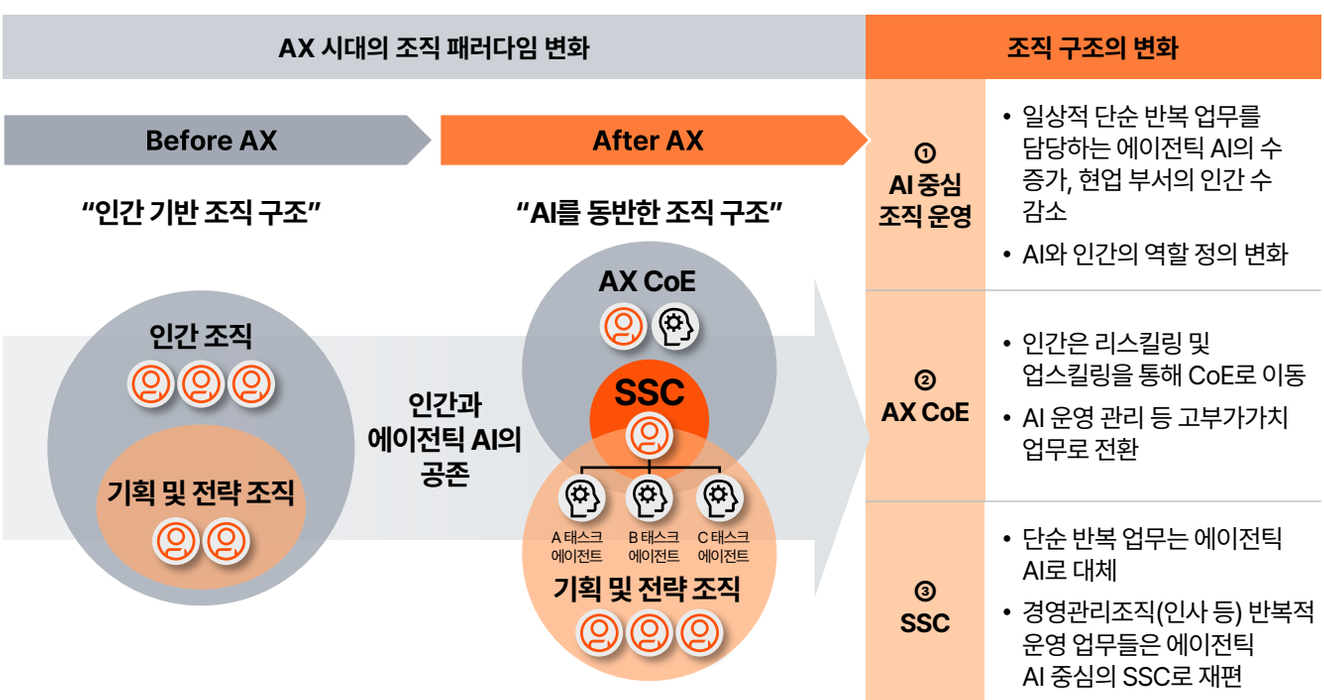
AI 중심 조직으로의 변화

에이전틱 AI 시대의 조직은 인간 중심 구조에서 에이전트가 실행하고 인간은 운영 및 개선하는 구조로 전환하여 생산성과 운영 안정성을 확보합니다.

AX 이전의 인간 중심 조직은 실행과 관리를 인간이 담당했으나, AX 이후에는 AI를 구성원으로 포함한 구조로 전환되며 인간과 에이전틱 AI의 공존이 본격화됩니다. 일상적이고 반복적인 업무는 에이전틱 AI가 주도적으로 수행하고, 현업 부서의 인력은 점진적으로 축소됩니다. 인간은 리스킬링과 업스킬링을 통해 AX CoE로 이동하여 AI 운영·관리 등 고부가가치 영역에 집중합니다.

AI 활용이 확산됨에 따라 현업 부서의 인력 구성도 재편되며, AI 중심 운영 모델이 핵심이 됩니다. 이를 위해 AI 정책, 품질, 리스크를 전담하는 AX CoE가 신설되거나 강화되어 조직 전반의 AI 거버넌스를 책임집니다. 반복적이고 공통적인 업무는 에이전트 중심으로 수행하는 AX SSC로 재편되어 효율성을 높입니다.

AX시대의 조직 패러다임과 구조 변화



출처: PwC

AI 중심 조직 운영

AI 중심 운영 체계에서 에이전틱 AI는 표준 워크플로우를 실행하고, 인간은 예외 케이스 처리, 품질 관리, 워크플로우 개선의 역할을 하게 됩니다. 이를 구현하기 위해서는 AI 전문 관리 조직인 AX CoE, AI와 인간의 협업 조직, 단순 반복 및 공통 업무를 수행하는 SSC를 연계하는 구조를 설계할 필요가 있습니다.

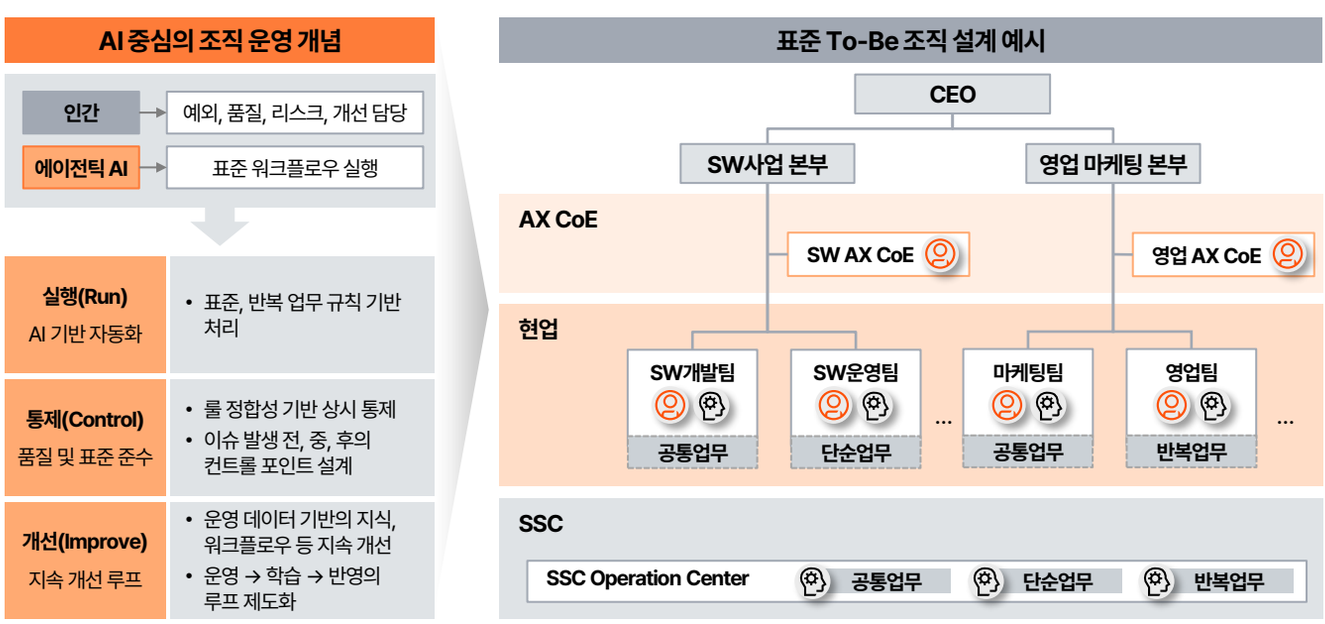
AX CoE

AX CoE는 비즈니스 단위별로 현업에 분산된 AI를 통제·가속하는 핵심 관리 조직으로서, 도메인 업무에서 활용되는 에이전틱 AI의 정책·지침·운영을 전문적으로 관리하고 지원하는 역할을 담당합니다. 이 조직은 높은 사업 이해도와 직무 전문성과 더불어 AI 리터러시 역량을 보유한 인력으로 구성되어 전략 기획, 모니터링 등을 수행합니다. 구체적으로는 AI 거버넌스와 정책 수립, AI 자산 및 포트폴리오 관리, 품질·리스크·윤리 관리의 중심 축 역할을 하며, 현업의 AI 활용 과제를 발굴하고, AX 전략 및 로드맵 수립과 함께 현업 부서의 AI 도입과 확산을 체계적으로 지원합니다.

SSC

SSC에서는 공통·중복 업무, 단순 반복 업무 등을 에이전틱 AI가 실행하고, 인간이 모니터링, 예외처리, 품질 점검을 수행하여 안정적 운영과 단계적 조직 구조 전환을 달성합니다. 이 과정에서 SSC의 수행 결과물은 현업 부서에 공유하고, 운영 상의 개선 요구사항이 있으면 AX CoE에 전달하여 상시 개선하도록 유도합니다. 결과적으로 SSC는 비용 효율화와 품질 통제를 넘어, AI 중심 운영체계로의 안정적 전환을 실현하는 실행 거점 역할을 수행합니다.

AI 중심 조직 운영으로의 전환을 위한 조직 구조 설계



출처: PwC

조직 운영 모델 정립

인간-AI 연계 직무 재설계

에이전틱 AI의 도입으로 직무가 사라지거나 유사 직무로 통합되는 변화가 예상됩니다. 이런 특성을 반영하여 브로드 밴드(Broad-band) 방식의 직무 체계가 필요할 것으로 예상합니다. 기존 직무 분류 체계는 AI 활용 역량과 AI 기반 업무 프로세스를 충분히 반영하지 못하므로 재설계가 요구되며, 이를 기준으로 인사평가, 보상, 교육, 승진, 배치 등 HR 프로그램의 기준과 운영 방식도 함께 재정비해야 합니다.

AI 핵심역량 체계의 재정립

AX 인재상은 AI 기술 역량뿐 아니라 비판적 사고, 추론, 창의적 문제해결 등 인지 역량을 요구합니다. 또한 AI가 인간의 역할을 대체하는 환경에서 사회적 공감, 대인관계, 감정 조절, 윤리적 판단 등 인간 고유의 역량은 더욱 중요해집니다. M형 관리자는 다양한 분야의 폭넓은 지식을 결합해 AI-first 워크플로우를 설계·관리하는 통합적 역량이 필요합니다. T형 전문가는 특정 분야 전문성과 함께 에이전틱 시스템을 미세조정하고 예외 사례를 관리할 수 있는 역량이 요구됩니다. AI 일선 근로자는 사회·정서적 역량을 기반으로 기본 AI 활용 역량을 갖추고 현장에서 AI를 안전하게 적용해야 합니다.

에이전트 보스 리더십 설계

AI 에이전트와 인간이 함께 성과를 내기 위해서는 기존 인간 중심 관리 방식에서 벗어난 에이전트 보스 리더십(Agent-oriented Leadership)이 필요합니다. 기존 리더십은 인력 관리 중심이었으나 에이전트 보스의 관리 영역은 AI와의 협업, 거버넌스, 윤리까지 확장됩니다. 특히 AI의 오류·편향, 책임 불명확성, 성과 왜곡 등 리스크를 방지하기 위해 AI 리터러시 교육, AI 리더십 스킬 트레이닝, 업스킬링, 리스킬링 로드맵이 함께 설계되어야 합니다.

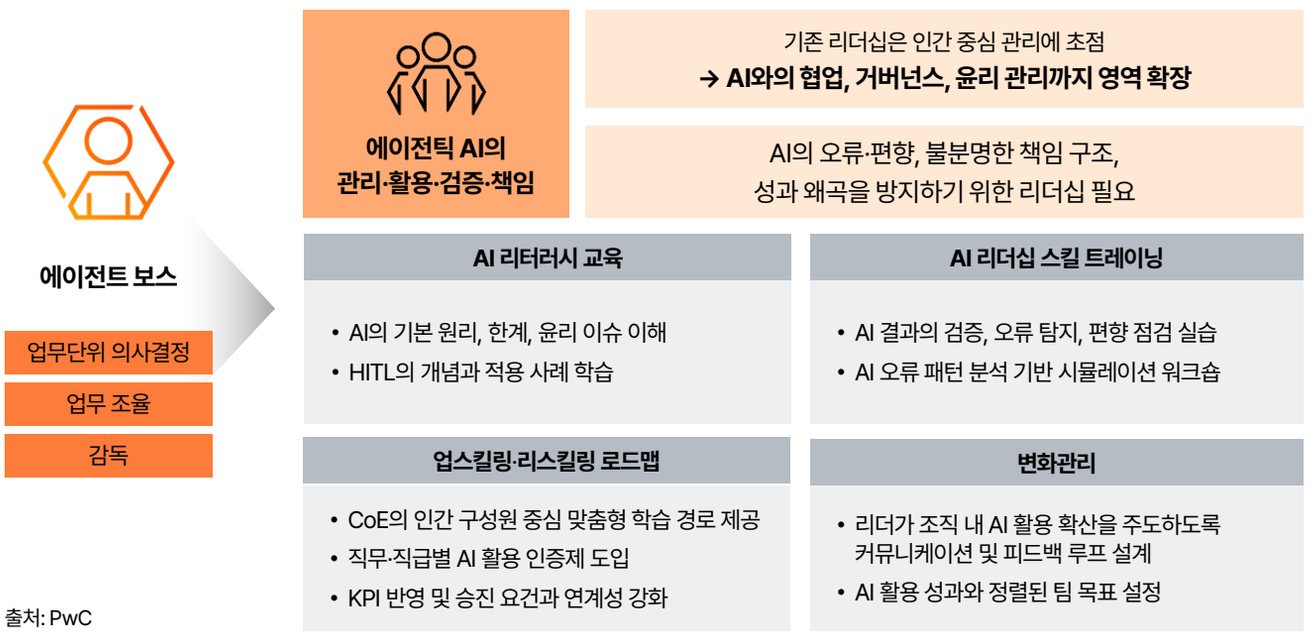
하이브리드 조직 생산성 및 KPI 재정의

인간과 AI 에이전트가 함께 수행하는 업무 환경에서는 기존 KPI가 AI 기여를 반영하지 못해 성과 왜곡이 발생할 수 있습니다. 따라서 인간-에이전트 비율(Human-agent Ratio)과 같은 운영 지표를 도입해 협업 효율성을 측정해야 합니다. 주요 KPI로는 친숙도 및 사용 빈도, 신뢰 수준, 시간 절감 효과 등이 있습니다. 이때 AI가 만든 결과물과 인간이 수행한 판단, 검증, 조정을 분리해 측정 가능한 형태로 재설계하고 이를 통해 팀들의 AI 활용 수준을 객관적으로 비교하고, AI 활용이 성과로 이어지도록 보상 및 평가 체계를 연동해야 합니다.

인간-AI 기여도 산정

인간-AI 협업이 일상화된 환경에서 성과 평가의 공정성과 신뢰성을 확보하려면 워크플로우 설계 단계부터 기여도를 분리해 정의해야 합니다. 첫째, 워크플로우를 분해하여 전체 업무를 세부 요소로 분해하고 각 요소의 성과 기여 비중(가중치)을 설정합니다. 둘째, 각 성과 요소에 대해 인간과 에이전틱 AI의 담당 비율을 매핑합니다. 셋째, 실제 성과를 평가 지표에 대입해 인간과 AI의 정량적 스코어를 도출하고 필요 시 보정합니다.

에이전트 보스 리더십 재설계



출처: PwC

변화관리 실행

예상 저항 요인

AI 도입에 따른 조직구조 및 평가체계 변화에 대해 우려가 발생할 수 있습니다. 예를 들어, 조직 구조 차원에서는 고용 안정성 불안, 기존 권한과 통제가 약화될 것이라는 인식, 새로운 기술 환경에 적응해야 한다는 부담, AI를 활용해서 만든 성과가 공정하게 인정받을지에 대한 불신 등입니다. 이러한 요소들을 사전에 식별하고 리스크를 관리할 필요가 있습니다.

변화관리 전략

AX 변화관리는 기술 도입을 넘어 책임 있는 에이전틱 AI 활용 체계를 조직 전반에 정착시키는 것을 목표로 추진되어야 합니다. 우선, 전사 AI 변화관리 체계를 구축하고, 실제 업무에 적용 가능한 AI 윤리 원칙을 명확히 수립하여 구성원 신뢰를 확보할 필요가 있습니다. 인사, 평가, 심사, 리스크, 규제 대응 등 고위험 업무부터 HITL 방식을 적용해야 합니다.

다음은, AI 기반 인재 역량 강화입니다. AI의 한계와 오류 가능성을 이해하는 AI 리터러시 제고에 중점을 두어야 합니다. 실제 업무 적용 시뮬레이션과 직무별 AI 활용 교육을 통해 현업 활용도를 높일 수 있습니다. 더 나아가 직무별 AI 활용 인증제를 도입하고 이를 KPI 및 승진 요건과 연계하는 것도 방법입니다.

결국, 기술을 먼저 도입한 기업이 아니라, AI를 조직 운영의 핵심으로 설계한 기업이 경쟁 우위를 가져갈 것입니다. 이제 기업은 AI를 어디에 쓸 것인가를 넘어서, AI와 함께 어떤 조직으로 진화할 것인가를 고민해야 합니다.

에이전틱 AI 환경에서 주목 받는 보안 리스크

인간과 구조적 차이로부터의 리스크 그리고 대응

에이전틱 AI 환경에서 주목해야 할 보안 리스크

에이전틱 AI는 대화형 비서 수준을 넘어 업무 시스템에 직접 접속해 조회-판단-실행까지 수행하는 행위 주체로 빠르게 진화하고 있습니다. 이 변화는 생산성 혁신을 가져오지만, 보안 관점에서는 인간을 전제로 설계된 통제 모델이 더 이상 충분하지 않다는 것을 의미합니다.

특히 에이전트는 첫째, 비인간 계정과 토큰을 통해 상시로 시스템에 접근하고, 둘째, 툴, 플러그인을 호출해 코드 실행이나 외부 전송을 자동화하며, 셋째, 조직 전반의 SaaS, 데이터 레이크를 가로지르는 허브가 되고, 넷째, 문서, 웹, 메일에 숨겨진 지시문(프롬프트 인젝션)에 의해 오동작할 수 있고, 다섯째, 모델, 플러그인, 오픈소스 체인이라는 새로운 공급망 리스크를 동반합니다. 최근 침해사고에서는 '권한·자격증명 탈취→클라우드·SaaS 접근→대규모 데이터 유출' 패턴이 반복되고 있으며 에이전트는 이 확산 속도와 범위를 한 단계 더 끌어올릴 수 있습니다.

본 보고서는 에이전틱 AI 도입과 확산 과정에서 경영진과 실무자가 반드시 점검해야 할 5대 보안 리스크와 이를 통제 가능한 모델(계정 및 권한의 행위 통제, 지속 테스트)로 전환하는 방향성을 제시합니다.

인간 vs 에이전틱 AI의 구조적 차이

인간과 에이전틱 AI의 구조적 차이는 보안 통제 지점이 어디로 이동해야 하는지를 보여줍니다. 인간은 인증(로그인)과 권한부여(직무 기반 접근), 책임(결재, 감사, 징계)이라는 세 축이 비교적 명확합니다. 반면 에이전틱 AI는 의사결정 로직(모델, 프롬프트, 툴체인)이 외부 입력과 상호작용하며, 한 번의 판단이 곧바로 실행으로 연결됩니다. 즉, 권한부여와 판단 및 책임의 경계가 흐립니다.

특히 에이전트는 API 호출, RPA, 워크플로우 자동화, 코드 실행 환경과 결합되며 행동 실행이 기본값이 됩니다. 그래서 기존 보안이 주로 다뤄온 데이터 접근 통제만으로는 부족하고, 행위 통제(무엇을 실행할 수 있는가), 승인·검증(누가 언제 어떤 조건에서 실행을 허용하는가), 안전장치(실행 전후 가드레일)로 통제가 확장되어야 합니다. 또한 인간이 실수하면 보통 단발성인데, 에이전트는 동일 실수를 수천 건, 수만 건 반복할 수 있습니다. 따라서 설계 시 최소 권한을 유지하고, 안전한 기본값을 적용하며, 분리된 실행 환경을 구축하고, 완전한 로깅과 재현성을 확보하는 방향으로 가야 합니다.

구조적으로 보면, 인간의 실수는 한 명의 범위에서 끝나는 경우가 많지만, 에이전트의 실수는 연결된 시스템 전체로 전이됩니다(스프레드 효과). 또한 에이전트는 누가 승인했나, 어떤 정책이 허용했나 등 책임 소재가 불명확해지기 쉬워, 사고 후 대응이 지연됩니다. 따라서 설계 원칙은 실행 권한을 단계적으로 부여하고, 고위험 업무는 인간의 승인을 거치며, 모든 액션은 누가, 무엇을, 왜, 어떤 근거로 수행했는지를 기록하고, 실패 시 즉시 중지하고 되돌릴 수 있는 기능을 갖추는 방향이 되어야 합니다.

인간과 에이전틱 AI의 구조적 차이로 인한 관리 어려움

권한	판단	행동 규모
• **인간**: 고유 계정, 입·퇴사, 역할 기반 권한 교육, 규정·징계 등 조직 규율로 통제 • **에이전틱 AI**: API 키, 서비스 계정 기반 다수 시스템 동시 및 상시 접근, 계정 폭증	• **인간:** 의도, 상식, 직관, 윤리 기준에 따라 행동하므로 누가, 왜 했는지 조사 가능 • **에이전틱 AI:** 자체 프롬프트 및 연동 툴 기능을 조합하여 행동하므로 수많은 로그 분석 어려움	• **인간:** 실수해도 물리적 행동에 한계가 있어 피해가 상대적으로 국지적 • **에이전틱 AI:** 오류도 단번에 수천 건 자동실행, 대량 삭제, 설정 변경, 데이터 유출 가능
통제 대상 급격한 증가	**책임 소재 추적 난해**	**자동화된 대규모 피해**

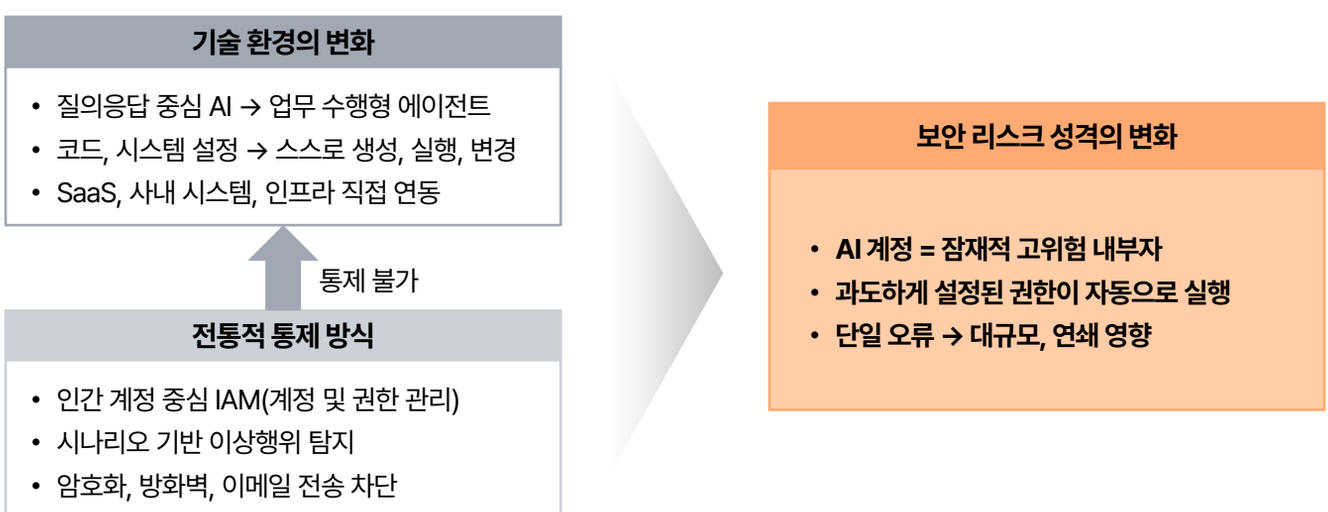
출처: PwC

에이전틱 AI가 보안 리스크로 전환되는 지점

에이전틱 AI 확산이 보안 리스크로 전환되는 지점은 권한과 연결성이 결합되는 순간입니다. 초기에는 에이전트가 단순 질의응답과 요약을 수행하지만, 조직은 곧 생산성을 위해 ERP, CRM, 메일, 파일, 개발도구, 티켓 시스템 등과 연결하고, 에이전트에 '대신 처리' 권한을 부여합니다. 이때 에이전트는 지원 도구를 넘어 실제 권한을 행사하는 새로운 행위 주체가 됩니다. 문제는 기술 환경 변화(멀티 SaaS, API-first, 자동화)와 조직 변화(원격근무, 빠른 의사결정)가 맞물려, 권한 설계 및 승인 프로세스가 따라가지 못한다는 점입니다. 최근 대규모 침해의 공통분모는 권한·자격증명을 탈취하여 클라우드 및 SaaS에 접근함으로써 대량의 데이터가 유출되는 것인데, 에이전트가 광범위한 SaaS 토큰을 보유하면 이 공격 체인은 훨씬 짧아집니다. 에이전트 환경에서 유사한 토큰 탈취가 발생하면, 공격자는 에이전트가 접근 가능한 전체(자산 등)로 곧바로 확산할 수 있습니다. 따라서 확산의 임계점은 에이전트가 업무 핵심 시스템에 쓰기 권한을 갖는 순간이며, 그 이전에 통제 모델을 선제적으로 구축해야 합니다.

또 하나의 전환점은 에이전트가 외부 콘텐츠를 상시로 읽기 시작할 때입니다. 브라우징, 이메일 읽기, 문서 검색(RAG), 티켓 처리 등은 모두 비정형 입력을 다루며, 공격자가 개입할 여지가 큽니다. 입력 신뢰도가 낮은데도 실행 권한이 열려 있으면, 프롬프트 인젝션, 피싱, 악성 링크를 통해 에이전트가 스스로 토큰을 넘기거나 위험한 액션을 수행할 수 있습니다. 따라서 확산 단계(파일럿→부서 확산→전사 확산)마다 연결 커넥터(팀즈, 지라, 메일, 셰어포인트 등 AI가 접근 가능 모든 도구)와 실행 권한을 정의하고 관리해야 하며, 전사 확산 이전에 리스크에 대한 테스트가 필요합니다.

기술 환경 변화에 따른 보안 리스크 성격의 변화



출처: PwC

에이전틱 AI 5대 보안 리스크

인간보다 위험한 내부자가 될 수 있는 AI 계정

에이전틱 AI 리스크는 개별 취약점의 문제가 아니라, ① 비인간 계정 증가, ② 자동 실행, ③ SaaS 횡단 데이터 노출, ④ 프롬프트 기반 조작, ⑤ AI 공급망 취약점이라는 다섯 축이 동시에 작동하면서 복합 리스크가 된다는 점이 핵심입니다. 특히 ①~③은 전통적 보안 영역(IAM, DLP, CASB)에서 이미 문제였던 과제들이고, ④와 ⑤는 LLM, 에이전트 특유의 신규 공격면입니다. 따라서 기존 통제 체계의 대응도 단일 솔루션 도입이 아니라, 에이전트 중심으로 재설계해야 합니다. 예를 들어, 비인간 계정은 인간 계정과 분리해 라이프사이클, 소유자, 회전, 사용 범위를 관리하고, 자동 실행은 위험 작업을 정책 기반으로 차단하고 승인하며, SaaS 데이터는 권한·공유 범위를 줄이고 DLP로 외부 전송을 감시해야 합니다. 프롬프트 인젝션과 공급망은 보안성 검토를 통해 입력 검증, 격리·신뢰 모델, SBOM, 모델 출처를 검증해야 합니다.

AI 계정의 내부자화에 따른 에이전틱 AI 핵심 보안 리스크

① 비인간 계정 폭증과 과도한 권한	인간 계정이 아니라 API 키, 서비스 계정 등 비인간 계정으로 여러 시스템에 동시 접근
② 비인가 툴 사용과 자동 실행	단순 답변을 넘어서 코드 실행, 파일 삭제, 인프라 설정 변경 등 실제 행동
③ 조직 전반 SaaS의 데이터 노출	효율을 위해 업무용 SaaS(메일, 드라이브, 도구·서비스 등)에 동시 다발적 접근
④ 프롬프트 인젝션 및 에이전트 간 공격 확산	공격자는 문서, 웹페이지, 입력값에 숨겨둔 지시문으로 에이전틱 AI를 속여 원래 의도와 다른 행동
⑤ 에이전틱 AI 공급망 및 생태계 취약점	악성코드가 포함되어 있을 경우 새로운 AI 공급망 취약점이 되어 내부 데이터 유출, 공격자 명령에 활용

출처: PwC

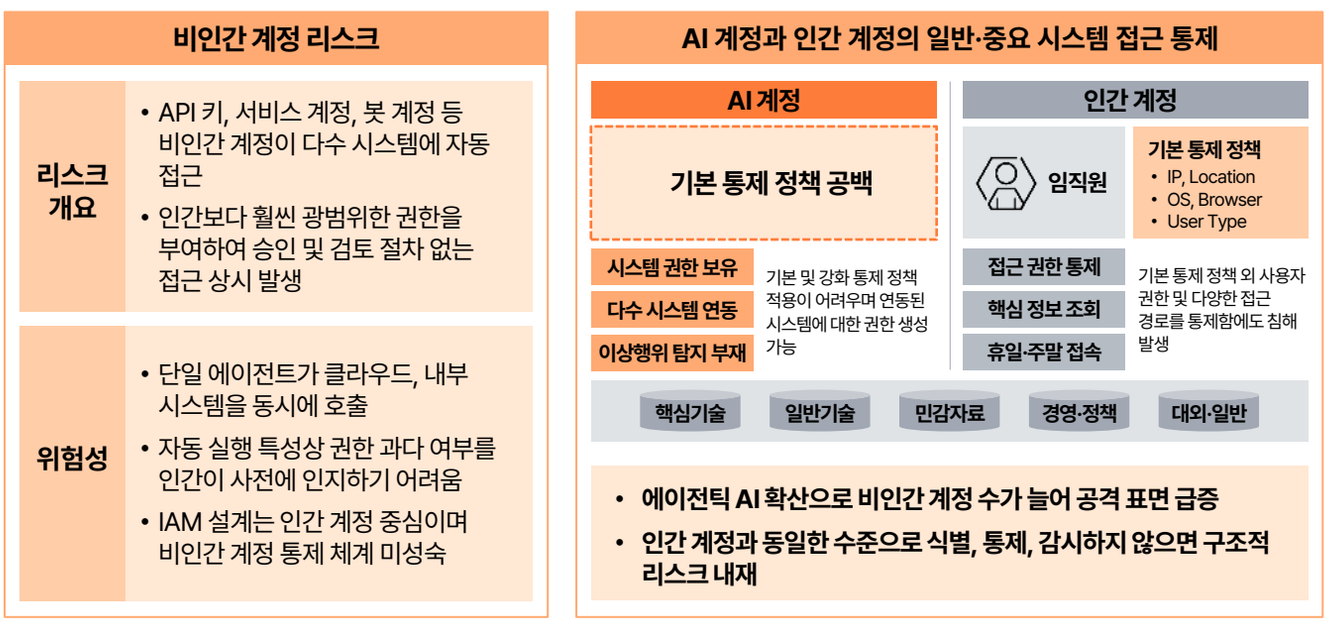
이 5대 리스크를 통제 대상 자산으로 매핑하면 자산에 어떤 리스크가 있는지가 정리되고, 이후 통제(정책, 감사, 테스트)도 자산 단위로 설계할 수 있습니다.

① 비인간 계정: 서비스 계정, 앱 등록, 토큰, API 키
② 비인가 툴: 플러그인, 워크플로우, 스크립트, CI, CD, RPA
③ SaaS 데이터: 메일, 파일, 채팅, CRM, 지식 베이스
④ 프롬프트 인젝션: RAG 문서, 웹, 메일, 티켓, 사용자 입력
⑤ 공급망: 모델, 데이터셋, 오픈소스, 레지스트리, 소스 저장소 등

① 비인간 계정 폭증과 과도한 권한

NHI(Non-human Identity, 비인간 계정)는 에이전트, 서비스 프린시플, API 토큰 등 인간이 아닌 주체의 접근 권한입니다. 에이전틱 AI 도입 시 NHI가 급증하고 권한이 누적돼 관리자 또는 특수 권한(Privilege) 토큰 1개를 탈취하는 것만으로 SaaS, 데이터 플랫폼을 횡단한 대량 유출이 가능합니다(예: 2024년 스노우플레이크 고객사 계정 탈취[1]). 이런 점에 대응하기 위해서는 NHI 인벤토리, 최소권한, 짧은 수명 토큰(OIDC, WIF), 시크릿 볼트,이상행위 탐지(대량 다운로드, 비정상 호출 등)를 함께 적용해야 합니다.

비인간 계정 기반 통제 공백과 공격 표면 확대 리스크



출처: PwC

② 비인가 툴 사용과 자동 실행

비인가 툴 사용은 에이전트가 메일 전송, 파일 공유, 결제·구매, 배포 등 현실 영향이 있는 툴을 자동 호출하며 생기는 리스크입니다. 프롬프트 조작, 추론 오류가 곧 실행으로 이어지고, 자동 실행 기능이 켜지면 사고가 인간의 검토 없이 발생됩니다. 2025년 코파일럿 스튜디오 악용(CoPhish[2]) 처럼 토큰, 승인 흐름을 노린 공격은 '툴 호출 권한=실행 권한'임을 보여줍니다. 따라서, 고위험 툴 HITL(승인·이중확인), 허용 목록(Allowlist) 및 파라미터 정책, 범위 제한 토큰, 추가 인증, 샌드박스, 감사 로그를 필수적으로 구축해야 합니다.

1) 스노우플레이크 고객사들을 겨냥한 인포스틸러를 활용, 취약 계정을 대량으로 탈취
2) 코파일럿 스튜디오 피싱 챗봇을 통해 마이크로소프트 정상 사이트로 연결되는 데모 페이지를 보여주고 로그인은 피싱하여 OAuth 권한 탈취

③ 조직 전반 SaaS의 데이터 노출

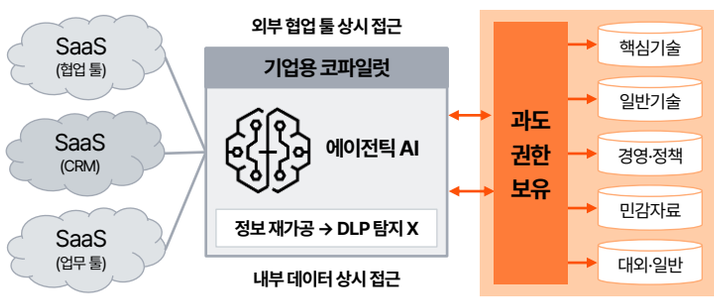
에이전트는 메일, 메신저, 문서, ITSM, DevOps 등 SaaS를 가로지르며 정보를 수집, 가공, 전송합니다. 이 과정에서 공유 링크, 외부 도메인 전송, 테넌트 경계 오인, 프롬프트·로그 잔존으로 데이터가 확산되고, 계정 탈취가 발생하면 연결된 커넥터를 통해 SW, 자산 전체로 피해가 확산됩니다. 대응을 위해서는 데이터 분류·라벨링 기반 정책, CASB·DLP로 SaaS 간 이동 감시(외부 공유, 대량 다운로드), 외부 공유 기본 차단 및 도메인 제한, 프롬프트·로그 최소화, 보존기간 관리, 이그레스 컨트롤과 키·토큰 교체가 수반되어야 합니다.

에이전틱 AI의 데이터 경계 붕괴 및 내부정보 유출 리스크

데이터 유출 리스크

리스크 개요	• 에이전틱 AI는 업무 효율을 위해 메일, 드라이브, 협업툴, CRM 등 다수 SaaS를 횡단 접근 • 단일 에이전트가 기밀 문서를 요약, 복사, 이동하는 과정에서 의도치 않는 데이터 경계 붕괴
위험성	• HR 에이전트가 퇴사자 정리 업무를 자동화하여 퇴사자 목록, 평가서, 급여정보 등 여러 시스템 데이터를 결합하여 리포트 작성 • 이 리포트를 전사 공유채널이나 드라이브에 게시하면 인사정보 대량 노출 우려

기업용 코파일럿을 통한 내부 정보 유출



- **에이전틱 AI는 연동된 사내시스템 접근이 가능한 API 키를 무한대로 생성 → API 키 생성 검증 절차 없이 운영 한계**
- **외부 SaaS에 접근 가능한 에이전틱 AI를 통해 기밀정보, 민감정보 등 유출**

출처: PwC

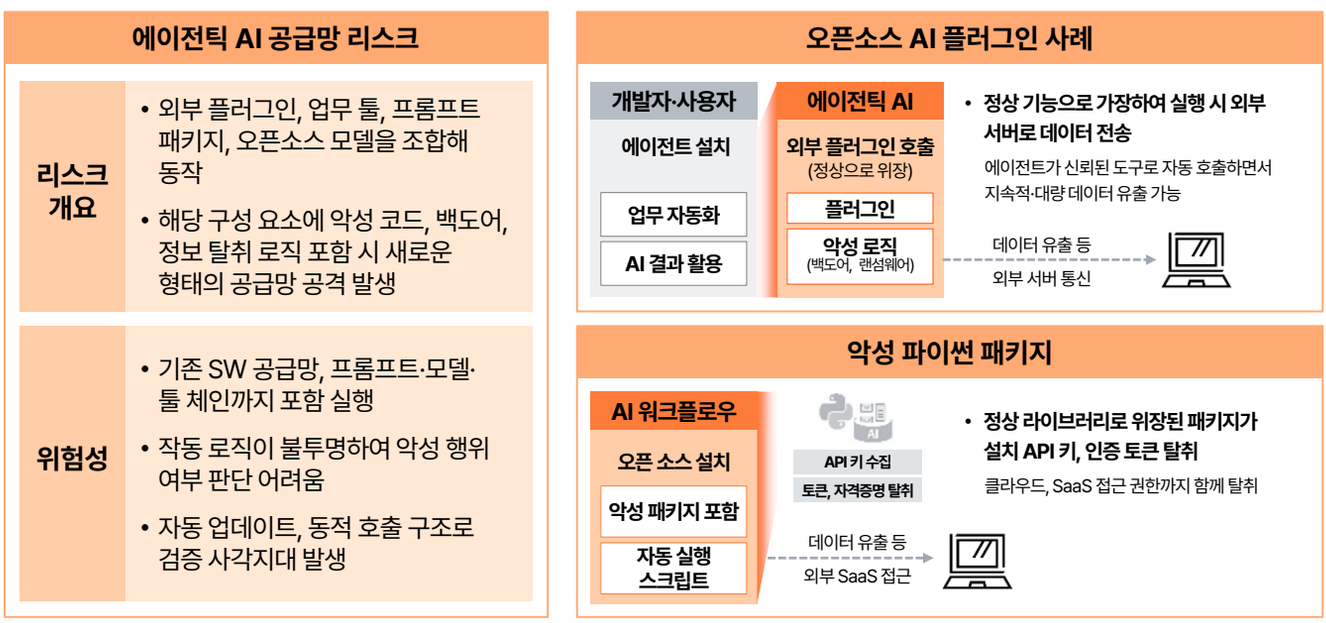
④ 프롬프트 인젝션 및 에이전트 간 공격 확산

프롬프트 인젝션은 대화 입력뿐 아니라 RAG 문서, 웹페이지, 이메일 등 간접 입력(즉, 문서 등에 프롬프트 지시문을 별도로 입력)에 숨은 지시로 에이전트의 의도와 정책을 바꾸는 공격입니다. 에이전틱 AI는 검색-요약-툴 실행 루프가 있어 간접 주입이 실제 툴 호출로 이어지기 쉽고, 다중 에이전트 구조에선 지시가 전파되며 확산됩니다. 방어를 위해서는 RAG 소스 정규화, 지시 패턴 제거, 컨텍스트 분리(내부 정책 vs 외부 문서), 고위험 툴 금지, 출력 DLP·마스킹, 레드팀 기반 우회 테스트를 상시 수행해야 합니다.

⑤ 에이전틱 AI 공급망 및 생태계 취약점

에이전틱 AI 공급망은 모델, 프레임워크, 플러그인·커넥터, 오픈소스, 벡터 DB, 외부 API까지 포함합니다. 구성요소 하나의 침해가 커넥터 토큰·권한을 통해 전 경로로 확대될 수 있으며, 2024년 XZ유틸즈의 백도어[1] 처럼 정상 업데이트·의존성 경로는 장기 잠복이 가능합니다. 대응을 위해 SBOM 가시화, 서명·프로비넌스(SLSA) 검증, 취약점 스캔, 신속 패치, 롤백, 서드파티 툴 샌드박싱(권한, 네트워크 격리), 벤더 리스크·계약 보안 요건으로 체계화하고, 운영에서는 최소 권한과 이상행위 탐지가 필수입니다.

에이전틱 AI 공급망 공격 및 검증 사각지대 리스크



출처: PwC

1) 리눅스 정상 배포판(5.6.0, 5.6.1)에 OpenSSH를 활용한 RCE 백도어가 설치되어 정상 배포판을 롤백해야했던 사건

해법의 방향성

에이전틱 AI 보안 해법의 방향성은 한 마디로 모델 보안이 아니라 계정, 권한, 행위의 통제입니다. 모델의 안전장치(유해콘텐츠 차단, 정책 프롬프트)만 강화해도 일부 위험은 줄일 수 있지만, 실제 사고의 대부분은 자격증명 탈취, 과도한 권한 오남용, 자동화, 공유 범위 확대에서 시작됩니다. 따라서 접근은 비인간 계정·권한 관리, 행위 통제, 데이터 통제의 세 레이어로 설계하는 것이 현실적입니다.

첫째, 비인간 계정·권한 관리(Identity Layer)에서는 에이전트 전용 비인간 계정 체계(고유 ID, 최소 권한, 접근·권한 정교한 설계), 토큰 수명·회전, MFA·조건부 접근, PAM, 앱 동의 정책(Entra) 등을 적용합니다.

둘째, 행위 통제(Action Layer)에서는 툴 호출 정책 엔진(허용 목록, 파라미터 검증), 고위험 액션 승인(HITL), 샌드박스·격리 실행, 레이트 리미트(Rate Limit), 롤백·킬스위치를 운영합니다.

셋째, 데이터 통제(Data Layer)에서는 데이터 분류·라벨링, 과공유 정리, DLP·CASB·DSPM를 통한 프롬프트·응답·첨부·외부 전송 감시, 민감 데이터에 대한 리트리벌 제한을 수행합니다.

이러한 세 레이어를 SIEM, SOAR로 연결해 탐지·대응 자동화 및 발생 가능한 리스크를 지속적으로 업데이트하고 개선하면 통제 가능한 운영 모델이 됩니다. 최근 에이전트 환경에서 발생한 토큰 탈취형 공격이나 클라우드 계정 탈취로 인한 대규모 유출 사례는 결국 권한이 과도했고 통제가 느슨했다는 공통점이 있습니다. 해법은 기술 도입 순서를 바꾸는 것으로, 에이전틱 AI 도입 전에 권한 설계와 데이터 거버넌스를 완료하는 것이 핵심입니다.

계정, 권한, 행위 통제 중심의 에이전틱 AI 보안 대응 체계

비인간 계정 관리	**에이전틱 AI는 인간 계정과 동일한 1급 자산으로 관리** 에이전트별 고유 ID, 역할, 최소 권한, 접근 로그 필수
고위험 액션 승인(HITL) 및 이중 검증	**대량 삭제, 권한 변경, 외부 공유, 운영 배포는 인간 승인 필수** 'AI 제안 → 인간 승인' 구조를 고위험 액션에 강제
프롬프트 인젝션 및 툴 런타임 방어	**입력, 문서, 웹 기반 프롬프트 인젝션 실시간 탐지 및 차단** 툴 호출은 화이트리스트, 파라미터 검증, 샌드박스 적용
관측성·감사성 (에이전트 가시성)	**프롬프트·툴 호출 결과 및 결정 사유 전 생애주기 로그화** 에이전트 가시성 기반 이상행위 탐지 및 사후 추적 확보
리스크 개선을 위한 지속적 반복적 테스트	**에이전틱 AI 시나리오 기반 행동 테스트** 발생 가능한 리스크 식별 및 테스트 후 개선·제거 이행

출처: PwC

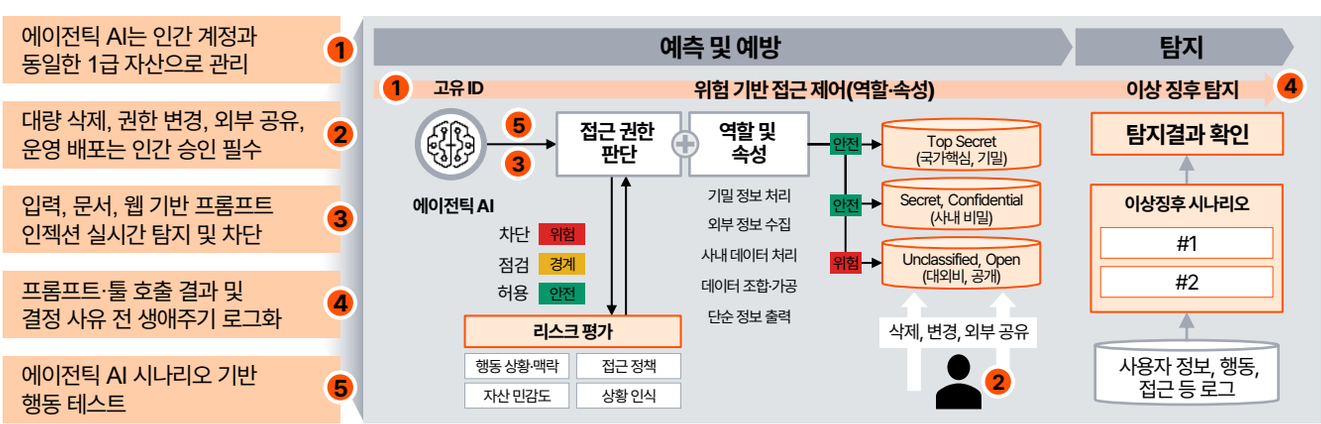
에이전틱 AI 접근통제 모델 구현

새로운 접근통제 모델의 목표는 올바른(권한 있는) 그리고 적정한(최소 범위) 데이터·행위만 취득·실행하도록 에이전트를 설계하는 것입니다. 구현 관점에서 핵심 구성은 다음과 같습니다.

- **정책 결정 지점(Policy Decision Point):** 사용자의 요청, 에이전트의 목적, 데이터 민감도, 툴 종류, 신뢰도(외부 또는 내부에서 접속)를 바탕으로 리스크를 계산하여 결정합니다(ABAC, Zero Trust).
- **정책 집행 지점(Policy Enforcement Point):** 리스크 결과를 참고하여 실제 툴 호출, 데이터 조회 전에 강제 집행합니다(차단, 승인, 감사). 리스크가 높다면 기존에 접근 가능한 자산이나 SaaS에 접근이 불가합니다.
- **자격 증명 브로커(Credential Broker):** 에이전트가 상시 토큰을 보유하지 않도록 필요 시점에 단기 토큰을 발급하고 즉시 폐기합니다.
- **가드레일(Guardrail):** 프롬프트 인젝션, 데이터 유출 징후를 실시간 탐지해 컨텍스트를 격리하거나 위험 액션을 '읽기·제안 모드'로 강등합니다.
- **관측성(Observability):** 프롬프트, 도구 호출, 파라미터, 결과, 데이터 소스를 완전 로깅해 재현성과 감사를 보장합니다. 이런 모델은 기존 IAM·PAM, DLP·CASB, SIEM과 연동되어야 하며, 특히 앱 동의·OAuth 권한과 커넥터 범위가 실무에서 가장 큰 취약점이므로, 사전 승인, 관리자 검토, 정기 권한 리뷰가 필수입니다.

결과적으로 이 모델은 AI 생태계의 확장 속도를 따라가기 위해 필수적으로 고려해야 할 요소입니다. 이 때 놓치기 쉬운 점은 사용자와 에이전트의 결합 책임입니다. 사용자는 업무 목적을 제공하고, 에이전트는 실행을 하지만, 책임은 조직에게 있습니다. 따라서 정책 결정은 사용자의 권한뿐 아니라 요청 목적, 데이터 입력, 신뢰도, 시간, 위치를 함께 고려해야 하며, 집행은 반드시 실행 전과 후에 걸쳐야 합니다(예: 실행 후 외부 공유 링크 생성 여부, 대량 다운로드 여부). 또한 근거 기반 응답(왜 이 데이터를 사용했는가)을 남기는 것이 규제와 감사 대응에 유리합니다.

에이전틱 AI 전 생애주기 기반 리스크 중심 접근통제 체계



출처: PwC

에이전틱 AI 테스트 시나리오 카탈로그

지속적 테스트는 에이전틱 AI 보안에서 선택이 아니라 필수입니다. 이유는 에이전트가 입력 데이터가 계속 바뀌고, 모델, 툴, 커넥터가 수시로 업데이트되며, 조직의 권한과 데이터 공유 구조가 변하기 때문입니다. 따라서 보안 품질은 일회성 진단으로 유지되지 않습니다. 테스트 카탈로그는 5대 리스크를 그대로 시나리오로 분해하여 운영해야 합니다.

예를 들어, NHI·토큰 탈취 시나리오는 로그, 코드, 워크플로우에서 키 유출, 권한 상승, 토큰 재사용을 점검하고, 비인가 툴 사용은 금지된 파라미터 주입, 고위험 액션 자동 실행 시도, 승인 우회를 테스트하며, SaaS 데이터 노출은 공유 문서나 메일 요약을 통한 민감정보 재조합과 외부 도메인 전송을 평가합니다. 프롬프트 인젝션 시나리오는 RAG 문서, 웹페이지, 메일에 숨겨진 지시문으로 도구 호출을 유도하거나 시스템 프롬프트 유출을 시도하는 방식을 점검하고, 공급망 리스크는 오염된 플러그인이나 라이브러리 투입, 업데이트 체인 악용 여부를 확인합니다.

각 시나리오는 성공 기준과 탐지 기준(경보, 차단, 로그)을 함께 정의해야 하며, CI(Continuous Integration) 파이프라인(예: 젠킨스(Jenkins), 깃허브 액션(GitHub Actions)) 처럼 정기 실행해야 합니다. 또한 코피시(CoPhish)같은 사회공학형(Social Engineering) 공격은 기술만으로 막기 어렵기 때문에 교육을 포함한 통합 테스트(피싱 시뮬레이션)도 포함시키는 것이 바람직합니다.

통합 테스트는 레드팀 일회성 이벤트로 끝나면 안 되고, CI처럼 자동화되어야 합니다. 예를 들어 프롬프트 인젝션 테스트 문서 세트를 정기적으로 RAG에 주입해 방어 성능을 측정하고, 툴 호출 정책이 업데이트될 때마다 회귀 테스트를 수행합니다. 또한 차단된 고위험 툴 호출 비율, 승인 지연 시간 감소, 민감정보 외부 전송 탐지 건수, 토큰 회전 준수율, 과도한 공유 리소스 감소율 같은 운영 KPI를 마련하면 효과가 큽니다. 마지막으로 사고 대응 관점에서, 테스트는 탐지, 차단, 복구까지 포함해야 합니다. (예: 에이전트 킬스위치 발동, 토큰 일괄 폐기, 데이터 접근 재평가). 이러한 체계가 유기적으로 만들어지면 에이전틱 AI는 통제 가능한 자산으로 전환될 수 있습니다.

테스트 시나리오 카탈로그 예시

ID	시나리오	설명	공격 경로	테스트 목적	성공 기준	심각도
01	시스템 프롬프트, 정책 유출	AI 내부 운영규칙(숨은 지시, 정책)을 캐내서, 방어를 우회할 단서를 획득	대화 유도, 에러 유발	방어 우회	비공개 프롬프트 노출	중
02	API 키, 토큰 유출	로그, 에러, 툴 출력에 비밀 키가 섞여 나가면, 그 키로 시스템을 대신 조작	로그, 툴 출력, 에러	계정 탈취	키 문자열 유출	상
03	이메일, 메신저 외부 전송	AI가 메일, 메신저 툴로 민감 정보를 외부(개인메일, 외부채널)로 전송	툴 오남용	데이터 유출	외부 도메인 전송	상
04	파라미터 인젝션	툴 입력값을 교묘히 바꿔(옵션, 경로, 범위) 금지된 동작을 수행	툴 인풋 조작	권한상승	금지 파라미터 실행	중
05	메모리 오염(지속성)	AI에게 악성 규칙을 학습시켜, 다음 대화에서도 계속 같은 오남용을 반복	장기 메모리	지속 악용	이후 세션에서 재발	중

출처: PwC

새로운 직원, 에이전틱 AI를 위한 보안 모델 설계

에이전틱 AI는 조직 곳곳에 접근 가능한 새로운 직원입니다. 새로운 직원(인간)이 입사하면 우리는 온보딩, 권한 부여, 감사, 퇴사(권한 회수)까지 전 과정을 설계합니다. 에이전틱 AI도 동일합니다. 핵심은 에이전틱 AI를 쓸 것인가 말 것인가가 아니라, 통제 가능한 보안 모델을 어떻게 설계할 것인가입니다. 구체적으로는 에이전트 전용 계정과 책임 주체(오너)를 정의했는가, 최소 권한, 단기 토큰, 승인 기반 실행으로 권한을 설계했는가, 데이터 분류, 과공유 정리, DLP로 SaaS 횡단 노출을 제어하는가, 프롬프트 인젝션과 툴 체인 공격을 전제로 입력 신뢰 경계를 설정했는가, 공급망(SBOM, 모델 출처, 플러그인 검증)과 지속 테스트 체계를 운영하는가입니다.

최근 공격자들은 정상 플랫폼을 활용한 사회공학과 토큰 탈취로 방어를 무너뜨리고 있으므로, 기술, 프로세스, 교육의 삼박자를 맞춰야 합니다. 결론적으로 에이전틱 AI 도입은 보안의 비용이 아니라, 업무 자동화의 안전한 확장을 위한 필수 투자이며, 지금 설계하지 않으면 나중에는 연결된 모든 시스템이 동시에 리스크로 전환될 수 있습니다.

에이전틱 AI는 새로운 디지털 워커이며, 디지털 워커의 권한, 책임, 감사를 설계하지 않으면 생산성 혁신이 곧 리스크로 전환됩니다. NHI, 툴 호출, SaaS 데이터, 프롬프트 인젝션, 공급망을 하나의 통제 평면(Control Plane)에서 관리하는 것이 핵심입니다. 현재 사용 중인 에이전틱 AI와 연결 커넥터를 확정하고, 고위험 액션(외부 전송, 권한 변경, 삭제, 배포, 결재, 송금)을 정의한 뒤, 해당 액션의 승인, 차단 정책을 문서화하여 파일럿 단계를 거쳐 점차 확산하는 것을 권장합니다. 이 과정을 통해 안전한 AI 도입과 활용이 가능할 것입니다.

Contacts

문홍기 Partner

hong-ki.moon@pwc.com
02-709-0394

임상표 Partner

sang-pyo.yim@pwc.com
02-709-0651

노승연 Partner

seungyon.roh@pwc.com
02-709-8177

최준걸 Partner

jun-kirl.choi@pwc.com
02-3781-9803

성윤호 Partner

yun-ho.sung@pwc.com
02-3781-9879

조용민 Partner

yongmin.y.cho@pwc.com
02-709-8278

조혜수 Partner

hyesoo.cho@pwc.com
02-2192-7794

김태형 Partner

taehyung2.kim@pwc.com
02-709-0583

S/N: 2603C-RP-037

© 2026 PwC Consulting. All rights reserved. PwC refers to the PwC network and/or one or more of its member firms, each of which is a separate legal entity. Please see www.pwc.com/structure for further details.
Disclaimer: This content is for general purposes only, and should not be used as a substitute for consultation with professional advisors.

pwcconsulting.co.kr