RESEARCH POSITION

개인의 AI를 넘어 조직의 업무를 완결하는 Agent

이 과정의 가치는 Agent 사용 능력보다, 고객·시민에게 지킬 수 있는 결과를 만드는 운영 능력으로 검증해야 한다.

1 개인의 효율
검색·초안·정리·추적을 맡겨 본인의 작업시간을 줄일 수 있는가.

2 조직의 효율
부서 사이 대기·인계·중복·오류까지 줄여 최종 결과가 좋아지는가.

3 운영의 지속성
누가 지시하고 판단하며, 문제가 생기면 누가 멈추고 복구하는가.

이번 연구의 잠정 결론

개인에게 유능한 Agent를 지급하는 것과 조직에 유능한 업무체계를 만드는 것은 다른 과제다. 조직에서는 단일 사건 기록, 부서 간 인수 확인, 실행 권한, 공동 완료 기준과 예외 처리 능력이 추가로 필요하다. 이 구조의 효과는 업무·수요·오류 비용에 따라 달라진다.

> 연구 구성: 5개 AI 연구 에이전트가 실증 근거, 기업 사례, 공공·대학 사례, 조직 위험, 시뮬레이션을 분담했다. 주 에이전트가 통합·주요 출처 재열람·모델 실행을 수행했다. 이는 사람 전문가 패널이나 실제 조직 현장조사를 대체하지 않는다.

기준 자료: LDR트랙 설계서 v1.1, 에이전트 지형과 조직 운용 내부 분석. 원본은 수정하지 않았다. 본 문서는 과정 보완을 위한 연구 결과와 제안이다.

01 / 범위와 가설

무엇을 Agent로 보고, 무엇을 검증할까

“진정한 Agent가 이제 처음 나왔다”보다는, 범용·개인용 상시 Agent의 확산과 제한된 업무 Agent의 조직 결합이 겹치는 단계로 정의하는 편이 정확하다. 특정 업무의 이벤트 기반 실행은 이전부터 존재했다.[9][10]

구분	실제 작동	조직 연구에서의 의미
챗봇·코파일럿	사람이 질문·과제를 주면 응답, 사람이 다음 단계 진행	개별 과업의 효과 근거. 무인 운영으로 일반화 불가
직원 전용 업무 Agent	목표·기억·도구·예약을 이용해 담당자의 일을 지속 수행	개인 이득과 타인에게 생긴 검토·인계 부담을 함께 측정
팀 공유 Agent	공용 문의·대장·반복 프로세스를 여러 직원이 이용	요청권·규칙 변경권·자료 접근권을 구분
프로세스 Agent	사건의 상태를 유지하며 여러 부서의 단계와 예외를 연결	공동 완료, 부서 간 인수, 최종 프로세스 책임자가 필요

세 연구 질문을 검증 가능한 가설로 전환

가설	반증 또는 실패 신호
H1. 개인 준비시간이 감소한다	자료정리·지시·검토까지 합치면 시간이 증가
H2. 인계·공동 상태를 재설계하면 조직 성과가 개선된다	통합·정책 운영비, 예외 조정이 이득보다 큼
H3. 권한·위험별 경로가 오류의 영향과 검토 병목을 줄인다	위험 오분류·형식 승인·공통 오류로 사고 증가
H4. 절약 여력이 서비스·학습·새 업무로 이동한다	기존 보고가 남고 생성량·업무 강도만 증가

> 가설은 결론이 아니다. 무작위 실험, 기업 자기보고, 기능 발표, 가상 사례, 수치모형을 같은 근거 수준으로 취급하지 않는다.

02 / 실증 근거 A

개별 과업의 이득은 확인된다. 적용 범위는 제한되어 있다.

근거 / 설계	관찰 결과	이 연구에서의 한계
Noy·Zhang, 2023 [1] 전문직 453명, 사전등록 무작위 글쓰기 실험	평균 시간 40% 감소, 평가 품질 18% 상승	짧은 글쓰기 과업. 조직의 결재·외부 실행·장기 성과 미측정
Brynjolfsson 외, 저자 개정본 [2] 상담직원 5,172명, 순차 도입 관찰	시간당 해결 건수 평균 15% 증가. 숙련도에 따라 차이	최종 응대는 사람. 무작위 도입이 아니며 상시 자율 Agent 아님
Dell'Acqua 외, 2026 게재본 [3] BCG 지식근로자 758명 무작위 실험	역량 범위 안 과제는 12.2% 더 완료, 25.1% 빠름. 범위 밖 과제 정답률 약 19%p 하락	비슷한 업무라도 적합성이 다름. 조직 전체의 업무 완결 실험 아님
METR, 2025 [4] 숙련 개발자 16명·246개 과제 무작위 실험	AI 허용 시 평균 19% 더 오래 걸림	당시 도구·숙련자·저장소 맥락의 결과. 최신 Agent 일반론으로 사용 불가

교육에 반영할 판단

과업 적합성
문서 업무라는 이유만으로 맡기지 않는다. 원문·검증 가능성·예외·오류 손실을 함께 본다.

체감과 실측
빠르다는 느낌 대신, 품질 확인까지의 총 사람시간을 기록한다.

숙련도 차이
신입 지원 이득과 숙련자 검토 부담을 나누어 본다.

> METR의 2026 후속은 표본 선택과 동시 Agent 사용의 시간 측정 문제를 지적한다. 2025의 역효과도, 후속의 일부 개선도 현재 모든 조직에 적용되는 확정 생산성 비율이 아니다.[5]

03 / 실증 근거 B

개인 시간 절약이 조직 변화로 이어지는가

근거	확인한 내용	읽어야 할 경계
Dillon 외, 최신 저자 원고 [6] 66개 기업·7,137명·6개월 무작위 현장실험	도구를 사용한 집단의 이메일 시간은 주당 약 2시간 감소. 개인 도구 지급만으로 과업 수량·구성의 뚜렷한 변화는 탐지되지 않음	사용자 효과 추정(LATE). 전체 배정 효과와 다름. 실제 업무 내용·완전한 생산성은 관측하지 못함
Vaccaro 외, 2024 [7] 106개 실험·370개 효과크기 메타분석	인간+AI 조합은 평균적으로 인간 단독과 AI 단독 중 더 잘한 쪽에 못 미침	인간 단독보다 항상 나쁘다는 뜻이 아님. 최신 상시 Agent 직접 검증 아님
Anthropic 고용 연구, 2026-10-16 [8] 내부 고용·산출·사용·통신 자료, 17명 인터뷰	직원 지시형 Agent 실행시간은 추정 사람 근무시간의 6배. 대화형 세션 약 80%가 사람 입력 대기. 메타업무 세션 비중 10%→22%	AI 기업 한 곳의 관찰. 성장·자기선택·과업 변화가 섞임. 6배는 생산성 배수가 아니며 80%는 사람의 절약시간이 아님

연구적 해석

개인에게 도구를 제공하는 처치와, 부서 사이 업무·권한·정보·평가 기준을 바꾸는 처치는 구별해야 한다. 현재 근거는 전자의 효과를 더 많이 보여준다. 후자는 제한된 현장 비교와 실패 사례를 통해 확인할 연구 영역이다.

> 효과가 없다는 결론도 성급하다. 측정하지 않은 품질·피로·신규 업무에서 가치가 생길 수 있다. 반대로 사용량·코드량·Agent 실행시간이 증가했다는 사실만으로 사업 성과를 주장할 수도 없다.

04 / 효율의 구조

조직의 이득은 어디에서 생길까

효율 경로	개인에게 생길 수 있는 가치	조직에 추가로 필요한 조건
준비·탐색	자료 수집·초안·요약의 시간 감소	원본 접근과 정보 최신성, 수혜자 관점의 품질 기준
지속 추적	기한·미응답·후속 조치를 잊지 않음	휴가·퇴사에도 유지되는 사건 기록과 소유자
병렬 처리	여러 대안·조회·작업을 동시에 준비	재고·예산·검토·현장 실행의 실제 제약 자원 확인
조정·인계	내가 보낼 자료가 빨리 준비됨	받는 부서가 그대로 쓸 수 있는 인계 조건·공동 상태
조직 기억	이전 경험과 정책 재사용	개인·팀·공식 정책 기억의 구분, 출처·만료·접근권
서비스 확장	혼자 못 하던 분석과 작업 가능	실제 수요·매출·서비스 품질로 연결되는 사용처

절약시간은 세 단계로 분리한다



가정 산식: 순여력 = 기준선 총 사람시간 − 도입 후 총 사람시간. 현금 이익은 실제 지출 감소·추가 이익이 확인된 경우에만 별도 계산한다. 기존 직원의 절약시간 × 시급을 곧바로 현금 절감이라 부르지 않는다.

05 / 조직 운영모델

개인 Agent의 집합이 조직 Agent는 아니다

운영모델	권한·지시	책임·인계	주요 적용
직원 전용 업무 Agent	지정 직원의 요청. 조직이 허용한 자료·도구만	직원과 팀장. 개인 취향·가족 기억은 업무 인계에서 제외	조사, 담당 사건 추적, 내부 문서 초안
팀 공유 Agent	요청자는 여럿, 규칙 변경자는 지정. 요청자 권한 확인	팀 소유자·대체자. 공용 계정 하나로 책임을 대신하지 않음	공용 문의함, 팀 지식, 반복 점검
프로세스 Agent	단계별 실행 범위. 충돌 시 사람에게 조정 요청	전체 프로세스 소유자 + 단계별 업무·정책·보안 책임	주문→납기, 신청→심사준비, 교육요청→정산

인계 카드: “검토해 주세요”를 넘어서

업무 상태
사건 ID / 현재 단계 / 완료·거절 조건 / 미확인 항목

근거·권한
원문 링크 / 정책 버전·기준시각 / 요청자 / 허용 행동

인수·복구
인수자·기한 / 수락 기록 / 실패 시 경로 / 중복 방지 키

리더의 역할 변화: 연구할 가설

지시·독촉·현황 취합의 일부가 줄어들 수 있다. 대신 기준·예외·갈등 조정·운영 품질·직원 육성이 중요해질 수 있다. 관리 계층 감소나 인원 대체를 전제하지 않는다. 부서별 Agent가 서로 다른 KPI를 추구하면 리더의 조정 부담은 오히려 커질 수 있다.

> 에이전트 신원·권한은 진행 중인 표준화 영역이다. NIST 프로젝트는 요구사항 탐색의 근거이며, 모든 제품이 이미 구현한 확정 표준이 아니다.[13]

06 / 기업 사례 1 · 가상

제조·유통: 지킬 수 있는 납기 약속

"제품 300개를 다음 주까지 받을 수 있나요?" 영업 답변의 속도와 실제 납품 가능성은 다르다.

고객 → 영업	요구 품목·수량·희망일 접수, 공통 주문 ID
Agent → ERP	재고·생산·신용·물류 상태와 유효시각 조회
운영 부서	예약 가능 여부와 조건 수락, 충돌은 조정
사람 승인	비표준 가격·우선 생산·위험 납기 판단
고객 · 시스템	승인 조건 회신, 수락 후 주문·예약 상태 일치

운영 형태	달라지는 점
사람만	영업이 각 부서에 문의하고 회신을 모음
개인 AI	메일·견적 초안은 빨라지지만 재고 예약은 별개
부서별 Agent	각자 가능하다고 답해 동일 재고를 중복 약속할 위험
프로세스 Agent	단일 예약 원본·인수 기록·유효기간을 바탕으로 확정

> 사고 주입: 영업은 매출, 생산은 묶음생산, 재무는 신용 위험을 우선한다. AI 간 다수결로 해결하지 않는다. 전체 프로세스 책임자가 가격·납기·마진의 우선순위를 정한다. 재고 확정은 단일 시스템에서 처리한다.

성과: 확정 가능한 답변까지의 시간, 약속 납기 준수, 마진 훼손, 전체 사람시간, 재확인·재계획 횟수. 실제 병목이 설비라면 AI의 답변 가속만으로 출하량은 늘지 않는다.

07 / 기업 사례 2 · 가상

교육·컨설팅: 제안이 운영 가능한 약속이 되도록

MindCanvas와 유사한 조직을 가정한다. 고객이 “다음 달 교직원 교육을 예산에 맞춰 제안해 달라”고 요청한다.

단계	Agent 업무	사람의 판단·확정
요구 진단	대상·목표·선수요건·예산·납품물의 누락 질문	고객 문제와 교육 목표의 합의
구성안	승인된 과정·가격·실적을 바탕으로 대안 작성	교육 적합성, 맞춤 범위, 실적 사용 동의
가용성	강사·장소·장비·일정 충돌과 원가 확인	강사 배정, 예약, 예외 원가·할인
제안·계약	근거가 붙은 제안서·조건표·계약 체크리스트	대외 발송, 계약 조건, 비용·범위 승인
운영 인수	확정된 범위·일정·산출물·변경 이력 전달	운영 책임자의 인수 수락
교육·정산	준비물·출석·증빙·성과 보고 초안 추적	학습성과·증빙 확인, 결과보고·지급 승인

개인 최적화가 조직 손실이 되는 장면

제안 Agent가 과거 문서의 “2일 교육 + 30일 자문 + 앱 개발”을 그대로 포함한다. 제안은 빨라졌지만 강사·운영·개발팀은 약속하지 않은 일을 떠안는다. 표준 범위를 넘는 항목은 운영 인수 확인 없이는 발송되지 않아야 한다.

첫 실험 후보
요청 접수→필수조건 확인→내부 제안 검토표까지. 외부 계약·발송은 사람 승인.

공동 완료
제안서 파일 완성이 아니라 수주 시 실제 가용성과 운영 인수가 확인된 상태.

함께 볼 지표
응답시간 + 수주 후 범위 변경 + 공헌 이익 + 납품 지연 + 모든 참여자 시간.

> 낮은 도입비로 시작하기 좋은 후보이나 실제 1순위는 업무 빈도·누락 손실·데이터 정비 수준을 확인한 뒤 정한다. 본 사례는 도입 성과가 아니라 검증할 설계다.

08 / 기관 사례 1 · 가상

기업지원사업: 신청부터 심사 준비까지 이어지는 사건

접수	사건번호 생성, 원본 보존, 문서 유형·누락 후보
증빙 확인	증빙별 원문 위치·기준일·공고 버전 연결
보완 추적	담당자 승인 요청 발송, 회신 결합·미응답 추적
심사 준비	미확인 사실·이해충돌·근거를 심사자료로 정리
사람 결정	선정·탈락·금액·예외 판단과 결재, 승인 통지

개인 AI의 가치	조직 Agent의 추가 가치
파일별 요약, 보완 문안	같은 사건의 증빙·기한·정책을 부서가 함께 봄
담당자 자료정리 시간 감소	휴가·전보에도 추적과 인계가 유지됨
보고서 초안	이의제기 때 최초 원문·적용 규칙·수정 이력을 복원

초기 실험 경계: 선정·배제·지원액·불이익을 Agent가 독자 결정하지 않는다. “누락 추정”으로 접수를 막지 않는다. 첨부 PDF 속 지시는 업무 규칙이 아니다. 기관별 승인 환경과 자료 등급부터 확인한다.

성과: 검증된 심사준비 건당 총 사람시간, 처리기간 P90, 재보완률, 잘못된 접수 차단, 신청자의 추가 연락 부담, 문서·언어별 오류 격차.

행정기본법 제20조는 법률 근거와 재량 없는 범위의 자동화 처분을 구별한다. 본 연구는 개별 사업의 법적 적용을 확정하지 않는다.[14]

09 / 기관 사례 2 · 가상

대학 RISE·산학·교육사업: 일정 변경이 모든 단계에 반영되는가

교육운영팀·학과·산학협력단·재무팀이 한 교육 요청을 처리한다. 개인 Agent를 각자 붙이는 것만으로는 공통 일정과 계약 상태가 만들어지지 않는다.

공동 단계	Agent가 준비·추적할 것	사람의 확정
요청·범위	대상·일정·예산·필수 산출물	사업 목적과 교육 범위
일정·인력	강사·강의실·장비 충돌, 대안	강사 적합성·배정·일정
계약·재무	계약·발주·정산 증빙 목록과 대조	계약, 비용 인정·지급
진행·변경	신청·동의·출석·일정 변경 영향 목록	변경 승인과 공지
성과·결과보고	출석·설문·산출물의 증거 연결	수료·학습성과·결과보고 승인

시뮬레이션 사건

교육일이 바뀌었는데 계약·숙박·강사·재무 Agent가 구버전을 사용한다. 또는 성과보고 Agent가 긍정 설문만 골라 높은 만족도를 제시한다.

변경 통제
한 사건의 승인 버전에서만 실행. 변경 전 영향 대상을 표시하고 각 부서가 인수.

성과의 진실성
수료 기준·설문 분모 고정. 원자료·제외 사유·근거를 함께 확인.

검증된 완료
교육 종료뿐 아니라 증빙·정산·결과보고가 일치한 상태.

성과: 요청→교육 확정 시간, 일정·계약 불일치, 증빙 재수집, 정산 반려, 참가자 재문의, 최종 보고까지의 총 사람시간. 실제 개인정보·학생 평가에 대한 자동 결정은 초기 범위에서 제외한다.

10 / 추가 적용과 공개 사례

도입 사례는 운영의 단서를 준다. 효과 입증의 수준은 다르다.

가상 적용 후보 두 가지

조직	Agent가 맡을 흐름	성과와 함정
고객지원	인증된 문의→상태 조회→허용 조치→실패 시 사람 인계	확인된 해결률·재문의·잘못된 환불. 쉬운 건이 줄면 상담원의 남은 업무 난도가 상승
정책 의견 분석	개인정보 최소화→주제 후보→전문가 확정→분류→소수·경계 의견 검토	분류 품질·소수 의견 보존·원문 추적. 많은 의견을 요약했다고 정책 판단이 끝나는 것은 아님

웹 원문으로 확인한 공개 사례

사례	확인 사실	증거 수준 / 사용 경계
Dow / Microsoft [9]	청구서 수집·구조화·이상 탐지 자율 Agent와 사람이 묻는 Freight Agent 병행	공급사 사례. 당시 자율 Agent 1개. 다중 Agent 연결은 미래 예시, 절감액은 예상
Pamarco / Microsoft [10]	결산 작업의 소유자·기한·선행조건을 두고 실행 가능 작업을 주기적으로 확인	공급사·고객 자기보고. 약 40%는 결산 작업 자동화 추정이지 시간·인원 절감률 아님
영국 Consult / DSIT [11]	5만+ 의견 분류 약 2시간, 전문가 검토 22시간과 상세 수동 검토 별도	정부 공식 사례. 한정된 분류 파이프라인이며 기관 전체 생산성·범용 Agent 실험 아님

> 이번 연구에서는 실제 고객 시스템에 Agent를 연결하거나 행동을 실행하지 않았다. 위 사례의 기능·운영 방식은 확인했지만 공급사 자기보고를 독립 인과 추정으로 바꾸지 않았다.

11 / 위험 A

실행이 연결되면 한 번의 오류도 연결된다

가상 사고	예방·탐지	복구 / 운영 책임
그럴듯한 오판 폐지된 조건으로 보완 요청	원문·시행일·적용 조건. 독립 표본·재문의 확인	정정·불이익 방지 / 업무 소유자
외부 문서의 악성 지시 첨부파일이 자료 전송 요구	자료와 지시 분리, 수신자·도구 범위 통제	격리·자격 회수·영향 조사 / 보안 관리자
권한 세탁 다른 Agent에게 대신 조회 요청	원 요청자·목적·인계 권한 유지, 하위 권한 축소	하위 토큰 회수·노출 조사 / 플랫폼 관리자
공동 기억 오염 임시 할인이 상시 정책으로 저장	개인·팀·정책 기억 분리, 출처·만료·승인	파생 요약·결과도 재검토 / 지식 소유자
중복·경쟁 실행 같은 물량을 각각 예약	공통 사건·단일 상태, 중복 방지 키·확정 단계	예약·약속 정합성 복구 / 프로세스 소유자
공통 원인 오류 여러 Agent가 같은 오판에 동의	모델 다수결 대신 원문·규칙·별도 검증	관련 흐름 동시 제한 / 품질 책임자

운영 원칙

Agent끼리 동의했다는 사실은 독립 검증이 아니다. 같은 모델·원본·기억을 공유하면 오류도 함께 움직인다. 실제 행동의 권한은 자연어 지시만으로 보장하지 않고 도구·계정·정책 계층에서도 제한해야 한다.

> OWASP의 Agent 위험 프레임워크는 위협 점검의 근거다. 그 목록 자체가 도입 성공률이나 위험 발생 확률을 알려주지는 않는다.[12]

12 / 위험 B

승인과 책임을 둔다고 운영이 저절로 안전해지지는 않는다

가상 사고	예방·탐지	복구 / 운영 책임
승인 과부하·자동화 편향	유입·동시작업 상한, 근거와 예외 표시. 승인시간+오류·반려 품질	투입 축소·우선순위 조정 / 승인체계 책임자
부서 KPI 충돌	공동 완료·마진·서비스 기준, 조정권 지정	사람의 갈등 조정 / 프로세스 소유자
정책 드리프트·소유자 공백	버전·만료·인사발령 연동, 예약·미완료 작업 점검	중단·인계·재승인 / 팀장·정책 소유자
주니어 판단 경험 감소	AI 없이 근거 설명·예외 판단 과제. 기회수와 역량은 구분	교육·업무배분 수정 / 팀장
감시 불안·오류 은폐	로그 목적·열람자·보존 제한, 신고·회고 절차	평가·로그 운영 수정 / 인사·업무 책임자
장애·호출 폭주·업데이트	비용·시간·호출 상한, 회귀 점검, 수작업 대체	차단·정정·단계 재개 / 서비스 운영자

책임을 원인별로 설계한다

업무 소유자는 결과·복구, 규칙 승인자는 허용 범위, 플랫폼·보안 관리자는 접근·격리·기록, 운영자는 예외·대기열, 프로세스 소유자는 부서 충돌을 맡는다. “마지막 승인자 한 명”에게 시스템·정책 실패까지 모두 몰아주면 학습과 개선이 어렵다. 이 구분은 운영 설계이며 법적 책임의 확정이 아니다.

> 사고 대응: 중단 → 확산 차단 → 증거 보존 → 영향 사건 식별 → 정정·보상 → 재개 판단. 현재 대화뿐 아니라 하위·예약·재시도 작업과 토큰 회수까지 시험한다. 이미 발송·노출한 정보는 버튼으로 완전히 되돌릴 수 없다.

13 / 수치 시뮬레이션 설계

같은 수요·인원에서 운영 방식만 바꿔 비교한다

모든 입력은 교육용 합성 가정이다. 실제 조직의 생산성 예측, Dots·Muse 성능 평가, 제품 간 실험이 아니다. 실행 결과의 구간은 모형 난수 변동폭이며 실제 효과의 신뢰구간이 아니다.

조건	사람 준비 / 마무리·조정	검토 경로 / 초기 정비
H 사람만	50분 / 12분	모든 건 리더 12분 / 0시간
P 개인 AI	20분 / 12분	모든 건 리더 12분 / 16시간
U 부서별 Agent	8분 / 14분	모든 건 리더 36분(중복 경로 3회) / 40시간
G 운영 규칙 적용	8분 / 10분	저위험 현장 12분+10% 표본 8분, 고위험 리더 18분 / 80시간
H+ 사람+흐름 개선	50분 / 12분	G와 동일한 라우팅·동시작업 상한 / 40시간

30 유효영업일, 담당자 4명×하루 300분, 리더 하루 180분. 하루 평균 8건과 20건의 포아송 수요를 분리한다. 40%는 고위험이며 업무 복잡도는 0.75~1.25배다. G·H+ 동시작업 상한은 8건. 막힌 요청도 외부 대기열에 남는다.

사람의 잠재 오류는 저위험 1%·고위험 5%, AI는 4%·16%로 가정했다. P·U·G의 AI 기본 성능은 같다. 검토시간에 따라 탐지 확률이 증가하며 탐지 오류는 한 번의 25분 수정으로 해결한다고 단순화했다. 조건마다 동일 사건·난수(seed 1000~1199)를 재사용했다.

총 사람시간은 관측기간에 소비한 작업·검토·수정과 초기 정비를 합산한다. 초기 정비는 비용·시간에 더하지만 대기열에는 배치하지 않았다. 실시간 달력 대신 유효영업일을 사용하므로 야간·휴일 효과를 분석할 수 없다.

14 / 실행 결과 · 합성 업무 대기열 모형

처리량·잔량·품질을 함께 보면 결론이 달라진다

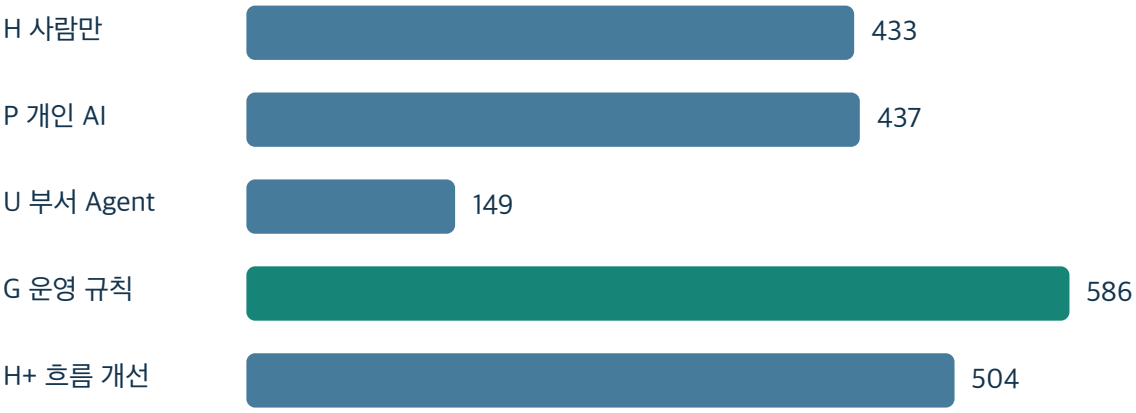
200회 평균. 완료는 독립 정답 기준을 통과한 건수, 잔량은 미완료 전체. 오류율은 보고상 완료 중 놓친 오류. P90은 완료된 건만의 경과 유효영업일이다.

소량: 하루 8건 · 30일 약 241건 접수

조건	검증 완료	잔량	사람시간*	P90 일	놓친 오류
H 사람만	237	2	298	0.40	0.53%
P 개인 AI	234	2	199	0.34	1.96%
U 부서 Agent	148	93	201	10.32	0.09%
G 운영 규칙	237	1	219	0.27	1.13%
H+ 흐름 개선	238	2	349	0.39	0.30%

증가: 하루 20건 · 30일 약 600건 접수

조건	검증 완료	잔량	사람시간*	P90 일	놓친 오류
H 사람만	433	165	668	7.44	0.54%
P 개인 AI	437	155	407	6.95	1.99%
U 부서 Agent	149	452	250	20.26	0.09%
G 운영 규칙	586	7	425	0.68	1.16%
H+ 흐름 개선	504	94	698	4.29	0.31%



해석: G는 이 가정에서 고수요 처리량·대기를 개선하지만 오류율은 H보다 높다. 따라서 이를 품질을 보장한 조직 효율 개선으로 확정할 수 없다. U는 낮은 오류율을 얻는 대신 검토 병목과 큰 잔량을 남긴다. 소량에서는 P가 G보다 사람시간을 적게 쓴다.

* 정비 완료 후 운영 용량에 초기정비 시간을 비용으로 추가한 모형. 완료만의 P90은 잔량이 큰 조건을 낙관적으로 보일 수 있다. 잔량·최장 미완료 대기는 동봉 CSV에 제공한다.

14B / 품질 게이트와 구성 차이

더 많이 끝냈다는 사실만으로 효율 개선을 인정할 수 없다

고수요 조건의 200회 평균을 위험도별로 분리했다. 각 표의 오류율은 평균 놓친 오류 건수 / 평균 보고상 완료 건수다. 기본 결과표의 반복별 비율 평균과는 계산 방식이 다르다.

조건	저위험 완료 / 잔량	고위험 완료 / 잔량	저위험 오류	고위험 오류
H 사람만	261 / 99	174 / 66	0.18%	1.08%
P 개인 AI	267 / 93	178 / 62	0.83%	3.73%
U 부서 Agent	90 / 270	59 / 181	0.03%	0.18%
G 운영 규칙	357 / 3	236 / 4	0.78%	1.73%
H+ 흐름 개선	304 / 56	202 / 38	0.17%	0.53%

이번 설정에 적용한 판정

조건	속도·시간의 신호	사람 기준 품질 게이트
P 개인 AI	총 사람시간 감소, 리더 병목은 지속	오류율 상승 → 효율 개선 합격으로 판정하지 않음
U 부서 Agent	중복 검토 가정으로 처리량·대기 악화	오류율은 낮으나 잔량이 커 서비스 개선 아님
G 운영 규칙	검토 분산으로 처리량·대기 개선	오류율 상승 → 조건 수정·추가 검증 필요
H+ 흐름 개선	AI 없이도 리더 병목 일부 완화	품질 신호 양호. 초기정비·더 많은 노동 투입 함께 평가

> 이 결과는 좋은 모형 결과를 만들기 위해 정확도 가정을 조정하지 않았다는 점에서 유용하다. 다음 실험에서는 사람과 AI의 원래 오류율·검토 탐지율·위험 분류 정확도를 실제 표본으로 측정해야 한다.

독립 에이전트가 모델을 읽고 재현성·건수 보존·자원 상한을 점검했다. 주 에이전트는 위험도별 보존, 공통 접수, 동일 seed 재실행, 오류 0 조건과 극단 입력을 추가 검증했다. 이는 구현 검증이며 현실 적합성 검증은 아니다.

15 / 민감도와 반증

운영 규칙의 가치는 조건부다

민감도는 기준 시나리오에서 한 변수씩 바꾸었다. 각 조건 80회, 고수요 하루 20건이다. 조건 간 정확한 효과는 대응 seed로 비교할 수 있다. 여러 요인이 동시에 바뀌는 상호작용은 이번 실행에 포함하지 않았다.

변경 가정	검증 완료	잔량	놓친 오류	사람시간*
U 검토 경로 1회	438	157	1.94%	327
U 검토 경로 3회	149	455	0.08%	250
G 검토시간 2배	357	246	0.14%	379
G 리더 용량 90분/일	354	245	1.17%	289
G 일별 공통 오류 15%	575	8	3.46%	456
G 동시작업 상한 2건	331	268	1.18%	275

일부러 확인해야 할 역전 조건

조건	효과가 줄거나 역전될 이유
중복 검토가 실제로 없다	U의 불리함은 작아진다. 3회 중복은 현장 사실이 아니라 시험 가정
수요가 적고 개인 AI로 충분	G의 초기 정비·운영비를 회수할 여지가 작음
위험 분류가 틀린다	G가 잘못된 건을 현장 경로로 보내 품질·권리 위험 증가
공통 정책이 틀린다	많은 Agent가 동시에 틀릴 수 있음. 수가 독립성은 아님
실제 병목이 물리적 실행	검토를 분산해도 제조·현장 출동·전문가 판단은 그대로

모형의 미구현: 위험 오분류, 공통 원인에 따른 검토 실패 상관, 외부 시스템 장애·보안 공격, 재수정 반복, 고객 반응, 사람 학습은 수치화하지 않았다. 공통 오류 변수는 잠재 오류를 함께 높일 뿐 공통 검토 실패까지 재현하지 않는다. 이 영역은 사건 역할극과 현장 검증으로 보완한다.

* 초기정비 포함. 동봉 결과에는 정비비 0·0.5·1·2배도 포함한다. 금전 계산은 시간당 4만원의 인력용량 가치, 도구비 P 9만원·U/G 18만원, 실행당 150원이라는 임의 가정이다. 실제 예산절감이 아니다.

16 / 교실 시뮬레이션

85분 동안 조직 운영의 차이를 경험한다

시간	참가자 활동	남길 증거
0~10분	담당자·Agent 운영자·검토 리더·감사/고객 역할. 완료 기준 합의	동일 난도의 사례 12건, 역할·권한
10~25분	라운드 1: 기존 또는 개인 AI 흐름 처리	건별 상태·사람시간·결정·잔량
25~35분	독립 품질 확인과 병목 진단	누구의 시간이 줄고 늘었는가
35~45분	인계·공동 상태·위험 경로·WIP·대리 승인 설계	인계 카드·운영 규칙 v1
45~65분	라운드 2: 평행 사례 12건, 사건 3~4개 주입	충돌·중복·멈춤·복구 기록
65~75분	품질·완료·잔량·부담 비교	검증된 결과와 놓친 문제
75~85분	효과·한계 판정과 30일 검증 질문 작성	계속·수정·중단·보류와 이유

사건 카드 6종

상충 약속
영업과 운영 Agent가 다른 납기를 제시

공통 원본 오류
정책·가격표 한 줄이 틀림

승인자 부재
리더가 자리를 비워 요청 누적

중복 실행
다른 채널에서 같은 신청·주문 접수

소유자 변경
담당자 전보 후 예약작업이 남음

목표·지표 왜곡
처리율을 높이려고 미완료를 제외

연습 효과를 줄이기 위해 팀 절반은 조건 순서를 반대로 한다. 동일 난도·사건 구조의 평행 자료를 사용한다. 역할극은 운영 판단의 학습 증거이며 실제 조직 생산성의 실증 자료가 아니다. 외부 시스템 쓰기·결제·개인정보는 사용하지 않는다.

17 / 30일 현장 연구

작은 실험으로 다음 결정을 위한 근거를 만든다

기간	실행	판정에 필요한 기록
1~7일	업무 하나의 기준선. 난도·오류·모든 참여자 시간 측정	접수·완료·재작업·대기·현재 예외 처리
8~21일	같은 건의 섀도 비교. AI 결과는 대외 실행하지 않음	원문·규칙 버전, 블라인드 품질평가, 총시간
14일	승인자 부재·규칙 변경·긴급 중단 모의시험	하위·예약·토큰까지 멈추는지와 복구시간
22~27일	지연 오류·재문의 확인, AI 없는 판단 과제	추가부담·놓친 오류·독립 판단·구성원 의견
28~30일	계속·수정·중단·보류 중 선택	효과·불확실성·제약·다음 시험 범위

현장 연구의 비교 설계

가능하면 사건·팀을 난도별로 묶어 배정하거나 단계적으로 도입한다. 같은 건을 두 번 처리하는 섀도 실험에는 이중 수행 비용이 생긴다. 연구 수행비와 향후 정상운영 건당 비용을 나누어 보고한다. 한 팀의 전후 비교만 가능하면 수요·숙련·계절성·동시 제도변경을 기록한다.

> 표본의 한계: 기준선 5건은 착수 확인용이다. 독립·동일 오류확률 가정에서 30건 모두 오류가 없어도 오류확률의 95% 단측 상한은 1−0.05^(1/30) ≈ 9.5%다. 오류 0건은 안전 증명이 아니다. 30일로 드문 중대 사고·장기 학습·전사 ROI를 확정하지 않는다.

90일 선택 단계: 승인 환경이 준비된 기관만 제한된 실제 실행과 부서 간 인계를 확대한다. 오류·복구·비용이 허용범위인지 재확인하고, 결과가 불충분하면 보류한다.

18 / 측정과 의사결정

교육 평가와 조직 성과 측정을 분리한다

관점	핵심 지표	판정 오류 방지
주성과	품질 통과 완료 / 모든 참여자의 총시간	문서·하위작업·사용량을 완료로 세지 않음
서비스	접수→완료 P50·P90, SLA, 미완료 잔량·최장 대기	완료된 건만 빨라진 효과와 전체 대기를 구분
품질	놓친 오류·재작업·재문의·심각도·집단별 차이	속도 개선으로 품질 저하를 상쇄했다고 선언하지 않음
관리·조정	승인·재지시·인계·중단 횟수, 리더·타부서 시간	절약시간의 타인 전가와 기존 보고 존속 확인
경제성	초기정비·연동·실행·운영·감사·복구 비용	인력용량 가치와 실현 현금효과 분리
사람·학습	부담·신뢰·오류 신고, AI 없는 판단 근거 설명	로그·사용량을 개인 서열·인사평가로 전환하지 않음

착수 전에 적을 판정표

가설, 주지표 하나, 품질의 허용범위, 실패·중단 조건, 제외 기준, 지연 오류 관찰기간, 책임자·대체자, 비용 한도를 먼저 정한다. 결과를 본 뒤 분모나 성공 기준을 바꾸면 변경 사유와 원래 지표를 함께 공개한다.

계속
품질 경계와 업무 가치를 충족하고 운영 부담이 감당 가능

수정
효과 신호는 있으나 병목·오류 원인이 특정되어 교정 필요

중단·보류
금지행동·중대 오류 또는 환경·표본 부족. 보류를 실패로 몰지 않음

교육에서는 판단 근거·운영 설계·새 사례 대응을 평가한다. 현장에서는 품질 통과 완료량·처리기간·총시간을 연구한다. 완료량 측정은 감원 규모 산정이나 개인 속도 경쟁과 구별할 수 있다.

19 / LDR v1.1 보완안

위임 교육을 업무 완결의 연구로 연결한다

현재 설계의 강점·제약	보완 제안	연결 위치
위임 의도문·헌장·책임·멈춤이 강함	앞에 고객·시민 가치와 현재 실제 병목 진단 추가	1.3, LDR-301-M01
개인의 미뤄 둔 일에서 시작	부서 사이 자주 멈추는 사건 하나를 병행 선택	6.1, 6.9
팀 사이 인계를 90일로 미룸	교실에서는 다부서 흐름 필수. 현장 확대는 준비도에 따라	6.4~6.9, 8.1
기획을 승인 용량 중심으로 설명	검토뿐 아니라 생산·현장·데이터·전문가 병목도 진단	1.3, LDR-302-M02
비교를 주의·결정 지점으로 제한	현장 연구에는 품질 통과 완료량·P90·잔량 추가	6.9, 7.4, 8.3
규칙·건별 승인자의 책임 강조	업무·정책·시스템·운영·프로세스 책임 구분	6.7~6.8
소유자·버전·권한 원칙 보유	구체 인계 카드, 공통 상태, 중복 방지·재개 훈련	6.3, 9.3
30일 판정·심리적 안전 보유	희귀 사고 표본 한계, 이중수행 비용, 학습 기회 측정	8.1~8.4, 9.5

추가 시간을 늘리기보다 사례를 관통시킨다

기업은 주문·교육 제안, 기관은 지원사업·대학 교육사업 중 하나를 골라 개인 위임→팀 운영→다부서 인계→사고·복구→30일 검증으로 반복한다. 제품 화면 사용은 짧게 유지하되 합성 자료의 실제 도구 연결 시험을 별도 단계로 둔다.

> 기존 문서의 “담당자 칸만 AI로 바꾸면 아직 설계하지 않은 상태”는 완화할 필요가 있다. 단순 대체만으로도 충분한 가치가 있을 수 있다. 재설계는 필요성이 확인된 흐름에 적용하고, 그 비용까지 평가한다.

20 / 연구의 다음 단계

가장 먼저 확인할 것은 업무가 어디서 멈추는가다

이번에 완료한 것

5개 연구 에이전트의 문헌·공개 사례 검토, 6개 가상 적용 시나리오, 12종 사고와 대응, 5조건·2수요의 수치 실험, 8개 변수의 민감도, 교실·현장 실험 및 LDR 보완안을 구성했다.

실제 조직을 만나 확인할 질문

확인 대상	인터뷰·관찰 질문
업무·고객	무엇이 끝나야 완료인가. 최근 완료·재작업·실패 사례는 무엇인가
흐름·병목	어느 부서에서 얼마나 기다리는가. 승인·실행·정보 중 실제 제약은 무엇인가
자료·도구	정책·가격·일정·사건 상태의 원본과 소유자는 누구인가
권한·예외	정상 건과 예외를 누가 판정하는가. 담당자 부재·퇴사 시 누가 인수하는가
구성원	자동화 뒤 어떤 일이 남고 늘어나는가. 여력·학습·보상 약속은 무엇인가
효과·중단	어떤 변화가 나와야 계속할 것인가. 오류·부담·비용의 중단선은 무엇인가

실행 제안

가장 접근하기 쉬운 후보는 MindCanvas형 교육 제안·운영 인수, 대학 교육사업의 일정·증빙 추적이다. 각 후보의 사건 10~20건은 우선 흐름·표본 설계에 사용하고 효과 입증으로 부르지 않는다. 실제 비교 표본은 업무 변동성과 검출할 차이에 따라 별도로 정한다.

> 다음 현장조사 전에는 제품의 조직 신원·자료처리·로그·예약작업 중단·처리 지역·계정 인계의 실제 구현을 다시 확인해야 한다. 본 보고서는 현재 제공 가능성을 기관별로 확정하지 않았고, 누구에게도 조사 요청을 발송하지 않았다.

동봉 자료: 표준 Python만 사용하는 실행 코드, 가정 JSON, 200회 기준 결과·80회 민감도 CSV/JSON, 읽는 법. API 키·개인정보·외부 계정 연결 없이 재현할 수 있다.

REFERENCES / 1

근거와 출처

접속·검증 기준일: 2026-10-20. 연구 논문·기관 원문과 공식 발표를 우선했다. 링크는 클릭 가능하다. 기능 발표는 실현 효과의 증거가 아니다.

[1] Noy & Zhang (2023), Science. 전문직 글쓰기 무작위 실험.
https://scale.stanford.edu/publications/experimental-evidence-productivity-effects-generative-artificial-intelligence

[2] Brynjolfsson, Li & Raymond. Generative AI at Work, 저자 개정본 v2 (2024).
https://arxiv.org/abs/2304.11771

[3] Dell'Acqua 외 (2026 게재본). Navigating the Jagged Technological Frontier. Organization Science.
https://doi.org/10.1287/orsc.2025.21838

[4] METR (2025-07-10). Experienced Open-Source Developer Productivity. 연구기관 원문.
https://metr.org/blog/2025-07-10-early-2025-ai-experienced-os-dev-study/

[5] METR (2026-02-24). We are Changing our Developer Productivity Experiment Design.
https://metr.org/blog/2026-02-24-uplift-update/

[6] Dillon 외. Shifting Work Patterns with Generative AI, v4 HTML 열람본. 문서 본문 날짜 2026-08-24와 arXiv 표시 날짜가 달라 버전 URL로 특정.
https://arxiv.org/html/2504.11436v4

[7] Vaccaro, Almaatouq & Malone (2024). When combinations of humans and AI are useful.
https://arxiv.org/abs/2405.06087

[8] Wilmers, Catipon, Massenkoff & McCrory (2026-10-16). How Employment is Changing at Anthropic.
https://www.anthropic.com/research/employment-at-anthropic

REFERENCES / 2

근거와 출처

접속·검증 기준일: 2026-10-20. 연구 논문·기관 원문과 공식 발표를 우선했다. 링크는 클릭 가능하다. 기능 발표는 실현 효과의 증거가 아니다.

[9] Microsoft WorkLab. AI impact at Dow. 공급사 공식 고객 사례.
https://www.microsoft.com/en-us/worklab/ai-impact-at-dow-copilot-identifies-millions-in-cost-savings

[10] Microsoft Customer Stories (2026-09). Pamarco finance operations. 공급사·고객 자기보고.
https://www.microsoft.com/en/customers/story/27289-pamarco-technologies-dynamics-365-finance

[11] UK DSIT (2025-10-16). Consult: public consultation analysis. 정부 공식 사례.
https://www.gov.uk/government/news/ground-breaking-use-of-ai-saves-taxpayers-money-and-delivers-greater-government-efficiency

[12] OWASP. Top 10 for Agentic Applications for 2026. 위험관리 프레임워크.
https://genai.owasp.org/resource/owasp-top-10-for-agentic-applications-for-2026/

[13] NIST NCCoE. Software and AI Agent Identity and Authorization. 진행 중인 프로젝트.
https://www.nccoe.nist.gov/projects/software-and-ai-agent-identity-and-authorization

[14] 국가법령정보센터. 행정기본법 제20조. 법률에 따른 자동적 처분과 재량 구분.
https://www.law.go.kr/LSW/lsLinkCommonInfo.do?chrClsCd=010202&lsJoLnkSeq=1029334663

[15] OpenAI (2026-09-29). Introducing dots. 기업 specialist dots는 초기 시범. 기능·계획 근거.
https://openai.com/index/introducing-dots/

[16] Meta (2026-09-08). Introducing Muse. 개인 상시 Agent 발표. 기능 근거.
https://about.fb.com/news/2026/09/introducing-muse-personal-ai-agent/

> 원자료: MindCanvas 「LDR트랙_설계서_v1.1.md」와 「에이전트지형과조직운용_내부분석_2026-10.md」. 제안과 가상 시나리오는 본 공동 연구의 합성 결과다. 제품의 출시가 효율을 입증하지 않으며, 본 연구는 어느 제품도 직접 비교 시험하지 않았다.