

소규모 사업자는 어떤 업무부터 AI에 맡겨야 하는가 — 비용·위험·검토 책임 매트릭스

발행 2026-09-22 · 최종 수정 2026-10-09 · 판본 2026-10-09
판 · 웹 본문 <https://gaon-lab.kr/research/ai-task-selection-cost-risk-matrix>

독자의 질문

반복 업무가 많은 소규모 사업자는 어떤 업무부터 AI에 맡겨야 할까요?

핵심 답

시간이 얼마나 줄지 계산하기 전에, 개인정보·기밀이 들어가는지, 되돌릴 수 있는지, 고객에게 검토 없이 나가는지, 전문 판단이 필요한지부터 확인하세요. 해당하지 않을 때만 검토·재작업 시간까지 뺀 ‘순절감시간’을 계산하고, 결론은 점수가 아니라 네 가지 판정(바로 시도 가능 / 제한된 자료로 시범 적용 / AI 보조만 허용 / 현재는 맡기지 않음) 중 하나로 냅니다. 이 판정은 법적 판정이 아닙니다.

발견 3가지

공개 연구

1. 체감과 실측은 다를 수 있습니다. 숙련 개발자 16명이 실제 이슈 246건을 처리한 METR 실험에서 AI를 쓴 쪽이 19% 더 오래 걸렸지만, 참가자들은 끝난 뒤에도 20% 빨라졌다고 믿었습니다. 대상이 좁아 ‘AI는 느리다’는 일반 결론은 아니며, 저자도 일반화를 주장하지 않습니다.

GAON-LAB 계산

2. GAON-LAB 계산식은 월 순 사람시간 절감 = (기존 건당 처리시간 × 월 반복 횟수) - AI 사용 후 월 사람 투입시간입니다. AI 사용 후 시간에는 입력·지시, 사람 검토, 재작업, 준비시간(환산)이 들어가고, AI 실행시간은 사람이 기다려야 할 때만 더합니다.

가상 사례

3. 같은 계산식도 업무에 따라 결과가 갈립니다. 설명용 가정에서 공개자료 상품 설명 초안(월 40건)은 월 620분 절감, 고객 문의 답변 초안(월 200건)은 검토 4분 때문에 월 160분 더 듭니다. 실제 업무의 결과를 뜻하지 않습니다.

대표 사례

같은 계산식, 다른 결과 — 설명용 합성 사례 A와 B

사례 A · 공개자료 상품 설명 초안 (가상)

월 40건, 기존 건당 25분 → 월 1,000분

AI 사용 후 월 사람 투입시간 380분 (= $60 \div 3 + (2+5+2) \times 40$)

월 순 사람시간 절감 +620분 (약 10.3시간)

권장 분류: 바로 시도 가능

사례 B · 고객 문의 답변 초안 (가상)

월 200건, 기존 건당 8분 → 월 1,600분

AI 사용 후 월 사람 투입시간 1,760분 (= $180 \div 3 + (1+2+4+1.5) \times 200$)

월 순 사람시간 절감 -160분 (기존보다 오히려 더 듦)

권장 분류: AI 보조만 허용 — 고객에게 가는 답변이라 확인을 줄이지 않음

자료 기준일

공개 연구는 각 자료의 발행일(본문 표 3)을 따릅니다. METR·NBER·Noy & Zhang·Dell'Acqua 외(2023-09-22 작업본)·NIST·개인정보보호위원회 원문은 2026-10-09에 다시 확인했습니다. Dell'Acqua 외 2026년 학술지 출판본 본문은 확인하지 못했습니다. 법령·가이드라인은 2026-09-22 기준입니다. 합성 사례의 숫자는 설명용 가정입니다. 이 PDF는 웹 본문과 같은 판본(2026-10-09 판)입니다.

핵심 요약

- 반복 빈도가 높다고 자동화 우선순위가 자동으로 높아지는 것은 아닙니다. 검토 시간과 재작업 시간을 빼고 나면 실제로 아낀 시간(순절감 시간)이 음수가 되는 경우도 있습니다.
- AI 효과는 숙련도와 업무가 AI의 능력 범위 안에 있는지에 따라 달라집니다. 범위 밖의 업무에서는 오히려 정확도가 떨어질 수 있고, 틀린 답도 그럴듯해 보여 사람이 알아채기 어려울 수 있습니다.

- 개인정보·기밀이 포함되거나, 오류를 사람이 찾아내기 어렵거나, 되돌릴 수 없는 업무는 절감 시간이 커도 우선순위를 낮춰야 합니다.
- 이 글은 국내 소상공인의 평균 AI 도입률이나 평균 절감시간을 제시하지 않습니다. 신뢰할 수 있는 공식 원자료를 확인하지 못했기 때문입니다. 대신 독자가 자신의 업무를 직접 시범 측정하는 절차를 제공합니다.
- 결론은 자동화 점수가 아니라 네 가지 판정 중 하나이며, 법률·세무·노무처럼 전문판단이 필요한 영역은 이 글의 분류만으로 결정하지 않습니다.

연구 질문과 분석 범위

연구 질문

소규모 사업자와 작은 조직은 업무의 시간 절감 가능성뿐 아니라 오류 비용, 개인정보·기밀 위험, 검토 부담과 되돌릴 수 있는 정도를

함께 고려해 어떤 업무부터 AI에 맡겨야 하는가?

- 분석 단위: 업무 하나(예: '문의 1건 응대')를 기준으로 판단합니다.
- 분석 대상: 소규모 사업자와 작은 조직의 반복 업무.
- 이 글의 결론: 특정 도구 추천이나 단일 자동화 점수가 아니라 네 가지 판정(바로 시도 가능 / 제한된 자료로 시범 적용 / AI 보조만 허용 / 현재는 맡기지 않음) 중 하나입니다.
- 기준 시점: 법령·가이드라인은 2026년 9월 22일 기준 확인한 범위, 실증연구는 각 연구의 발행연도를 표시합니다.
- 기본 모형 밖: 특정 AI 제품의 성능 비교, 시스템 구축·계정 연결·라이선스 구매 대행은 다루지 않습니다.

기존 Research와의 차이

GAON-LAB에는 AI 도입을 다룬 자료가 이미 있습니다. 이 글이 그 자료들과 어떻게 다른지 먼저 밝힙니다.

표 1. 관련 자료와의 역할 분리

자료 Research 「소규모 사업자의 AI 활용 장벽」
다루는 질문 왜 AI 도입이 기대만큼 늘지 않는가(비용 부담, 지식·인력 부족 등 도입 이전 장벽)
이 글과의 관계 도입 여부를 막는 장벽을 다루는 별개의 질문입니다. 이 글은 이미 AI를 써보기로 한 다음, 업무 하나를 어떤 기준으로 고를지를 다룹니다.

자료 「AI가 작은 사업의 업무를 어떻게 바꾸는가」 (기존 프레임워크)
다루는 질문 반복성·되돌릴 수 있음·사람 최종확인이라는 판단 기준을 별도 근거 자료 없이 제시
이 글과의 관계 이 글이 같은 판단기준을 실증연구·공식 가이드라인 근거와 함께 다시 정리했습니다. 기존 URL은 보존하며 이 글로 연결됩니다.

조작적 정의

조작적 정의란 이 글에서 각 용어를 어떤 뜻과 단위로 쓰는지 미리 정해 두는 일입니다. 아래 용어는 모두 GAON-LAB이 이 글의 판단을 위해

설계한 실무 규칙이며, 학술적으로 검증된 척도가 아닙니다.

표 2. 조작적 정의

용어 업무 단위
정의 하나의 완결된 산출물을 만드는 반복 가능한 작업
측정 단위 —
누가 입력하는가 독자
판단에서의 역할 분석의 기본 단위
주의점 너무 크면 쪼개서 봐야 함

용어 반복 빈도
정의 같은 업무 단위가 발생하는 횟수
측정 단위 건/월
누가 입력하는가 독자
판단에서의 역할 기대가치 계산의 기준
주의점 계절성이 있으면 별도 표시

용어 기존 처리시간
정의 사람이 처음부터 끝까지 처리할 때 걸리는 시간 (건당)
측정 단위 분/건
누가 입력하는가 독자
판단에서의 역할 비용모형 입력값
주의점 실제보다 낙관적으로 적지 않도록 주의

용어 입력·지시시간
정의 사람이 입력자료를 정리하고 AI에게 지시문(프롬프트)을 작성하는 데 걸리는 시간(건당)
측정 단위 분/건
누가 입력하는가 독자
판단에서의 역할 비용모형 입력값(사람 투입시간)
주의점 항상 사람 투입시간에 포함

용어 AI 실행시간
정의 AI가 결과를 만드는 동안 걸리는 경과시간(건당)
측정 단위 분/건
누가 입력하는가 독자
판단에서의 역할 처리 경과시간 계산에 항상 포함, 사람 투입시간 에는 대기가 필요할 때만 포함
주의점 재시도 포함

용어

AI 실행 중 대기 필요

정의

AI가 실행되는 동안 사람이 다른 업무를 볼 수 없고 기다려야 하는가

측정 단위

예/아니오

누가 입력하는가

독자

판단에서의 역할

사람 투입시간에 AI 실행시간을 더할지 결정

주의점

'예'일 때만 AI 실행시간을 사람 투입시간에 더함

용어 사람 검토시간
정의 AI 결과를 사람이 확인·수정하는 데 걸리는 시간 (건당)
측정 단위 분/건
누가 입력하는가 독자
판단에서의 역할 비용모형 입력값(사람 투입시간)
주의점 가장 과소평가되기 쉬운 값('평균 효과율을 그대로 쓸 수 없는 이유' 절 참고)

용어 재작업시간
정의 결과를 다시 만들어야 할 때 드는 시간(건당 평균)
측정 단위 분/건
누가 입력하는가 독자
판단에서의 역할 비용모형 입력값(사람 투입시간)
주의점 재작업이 없다고 가정하면 0으로 입력

용어 초기 설정·교육시간
정의 도구를 처음 설정하고 익히는 데 드는 시간(1회성)
측정 단위 분
누가 입력하는가 독자
판단에서의 역할 월 환산의 기초
주의점 0 허용

용어

통합·전환시간

정의

기존 업무 방식을 바꾸거나 다른 도구로 옮기는데 드는 시간(1회성)

측정 단위

분

누가 입력하는가

독자

판단에서의 역할

월 환산의 기초

주의점

0 허용

용어 기대 절감시간
정의 도입 전에 예상한 시간 절감
측정 단위 분/월
누가 입력하는가 독자(사전)
판단에서의 역할 실측치와의 대조값
주의점 체감과 실측이 다를 수 있음('평균 효과율을 그대로 쓸 수 없는 이유' 절 참고)

용어

순절감시간(월 순 사람시간 절감)

정의

기존 월 사람 투입시간에서 AI 사용 후 월 사람 투입시간을 뺀 값

측정 단위

분/월

누가 입력하는가

계산값

판단에서의 역할

최종 판단의 핵심 지표

주의점

음수면 '권장하지 않음'으로만 표시, 자동 등급 배정 안 함. 사람이 들이는 시간 기준이며 처리 경과 시간과는 다른 값

용어
처리 경과시간
정의
업무 시작부터 완료까지 걸리는 전체 경과시간(건당)
측정 단위
분/건
누가 입력하는가
계산값
판단에서의 역할
고객 응답속도·납기 판단용 보조 지표
주의점
사람 투입시간이 줄어도 처리 경과시간은 그대로거나 늘 수 있음. 두 지표를 섞어 쓰지 않음

용어 표준화 가능성
정의 업무의 입력·처리 규칙이 매번 비슷하고 설명 가능한가
측정 단위 가능/불가능
누가 입력하는가 독자
판단에서의 역할 기대가치 판단
주의점 변동성 낮음 + 규칙 설명 가능, 둘 다 충족해야 '가능'

용어 업무 변동성
정의 같은 업무 단위의 처리 방식이 매번 얼마나 달라지는가
측정 단위 낮음/중간/높음
누가 입력하는가 독자
판단에서의 역할 표준화 가능성 판단
주의점 과소평가하면 표준화 가능성을 과대판단하게 됨

용어 입력자료 민감도
정의 업무 입력값에 개인정보·기밀이 포함되는가
측정 단위 없음/간접식별/직접식별·기밀
누가 입력하는가 독자
판단에서의 역할 위험 매트릭스 최상위 가중치
주의점 불확실하면 '직접식별·기밀'로 보수적 처리

용어 오류 발생 가능성
정의 AI가 틀린 결과를 낼 가능성
측정 단위 낮음/중간/높음
누가 입력하는가 독자(참고자료 기반)
판단에서의 역할 위험 판단 보조
주의점 공신력 있는 확률 자료가 없어 숫자로 만들지 않 음

용어 오류 발견 가능성
정의 사람이 결과의 오류를 알아챌 수 있는가
측정 단위 쉬움/어려움
누가 입력하는가 독자
판단에서의 역할 위험 매트릭스
주의점 AI 답변은 틀려도 그럴듯해 보일 수 있다는 점 (NIST가 '확신에 찬 오답'이라 부르는 위험)을 전제로 판단

용어 오류 결과의 심각도
정의 오류가 발생했을 때 영향 범위
측정 단위 낮음/중간/높음/법적책임
누가 입력하는가 독자
판단에서의 역할 위험 매트릭스
주의점 고객에게 나가는 결과는 한 단계 상향 권고

용어 되돌릴 수 있는 정도
정의 오류 발생 후 정정·회수가 가능한가
측정 단위 쉬움/어려움/불가능
누가 입력하는가 독자
판단에서의 역할 위험 매트릭스, 강제 하향 규칙
주의점 외부 발송·외부 시스템 실행은 기본적으로 '어려움' 이상

용어
고객 직접 영향
정의
결과물이 검토 없이 고객에게 그대로 전달되는가
측정 단위
내부초안/고객전달
누가 입력하는가
독자
판단에서의 역할
위험 매트릭스, 강제 하향 규칙
주의점
—

용어
외부 실행 권한
정의
AI가 주문·결제·환불·계정변경 등을 직접 실행하는가
측정 단위
예/아니오
누가 입력하는가
독자
판단에서의 역할
강제 하향 규칙
주의점
—

용어 법적·계약상 책임
정의 세무·법률·노무·의료 등 전문판단이 포함되는가
측정 단위 예/아니오
누가 입력하는가 독자
판단에서의 역할 강제 하향 규칙
주의점 포함되면 이 글의 분류만으로 결정하지 않음

용어 자동화 수준
정의 네 가지 판정 중 하나
측정 단위 범주
누가 입력하는가 계산값(규칙기반)
판단에서의 역할 최종 결과
주의점 법적 판정이 아니라 시범운영 권고

용어 시범 중단 기준
정의 시범 적용을 멈춰야 하는 조건
측정 단위 목록
누가 입력하는가 GAON-LAB 설계
판단에서의 역할 운영 규칙
주의점 '중단 기준' 절 참고

사용한 원자료

이 글은 발행기관의 공식 페이지에서 원문을 직접 확인한 자료만 핵심 근거로 사용합니다. 뉴스·블로그·벤더 홍보자료는 핵심 근거로 쓰지 않았습니다.

표 3. 사용한 원자료

기관/저자

개인정보보호위원회

자료명

생성형 인공지능(AI) 개발·활용을 위한 개인정보
처리 안내서

발행일

2025.8.

이 글에서 사용한 항목

개인정보 최소수집·목적제한 원칙, 프롬프트 입력
데이터 취급, 위험관리 방향

원문 링크

privacy.go.kr (안내서 원문 PDF)

기관/저자

NIST

자료명

AI Risk Management Framework 1.0 (NIST
AI 100-1)

발행일

2023.1.

이 글에서 사용한 항목

위험 식별(MAP)·측정(MEASURE)·관리
(MANAGE)라는 절차 구조, 자발적 프레임워크라
는 성격

원문 링크

nvlpubs.nist.gov (공식 PDF)

기관/저자 NIST
자료명 Generative Artificial Intelligence Profile (NIST AI 600-1)
발행일 2024.7.
이 글에서 사용한 항목 confabulation(확신에 찬 오답)·데이터 프라이버시·정보보안·사람-AI 역할 구성 등 생성형 AI 특유 위험
원문 링크 nvlpubs.nist.gov (공식 PDF)

기관/저자 Noy, S. & Zhang, W. (MIT)
자료명 Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence
발행일 2023.3. (작업본)
이 글에서 사용한 항목 특정 글쓰기 과제의 시간·품질 변화(표준편차 단위)
원문 링크 economics.mit.edu (공식 작업본 PDF)

기관/저자
Brynjolfsson, E., Li, D. & Raymond, L. R.
자료명
Generative AI at Work (NBER Working Paper 31161)
발행일
2023.4.(2023.11. 개정)
이 글에서 사용한 항목
고객지원 업무의 평균 효과와 숙련도별 차이
원문 링크
nber.org (공식 PDF)

기관/저자
Dell'Acqua, F. 외
자료명
Navigating the Jagged Technological Frontier(Organization Science 게재. 이 글의 수치는 2023.9. 작업본 기준)
발행일
작업본 2023.9., 학술지 2026년 37권 2호
이 글에서 사용한 항목
AI 능력 경계 안/밖에서의 성과 차이
원문 링크
SSRN 4573321 작업본 PDF(MIT Sloan 게시본, 2023-09-22판, 2026-10-09 확인). 2026년 학술지 출판본 본문은 접속이 막혀 확인하지 못함

기관/저자

METR

자료명

Measuring the Impact of Early-2025 AI on Experienced Open-Source Developer Productivity

발행일

2025.7.

이 글에서 사용한 항목

숙련 개발자의 체감 속도와 실측 속도의 차이

원문 링크

metr.org (공식 연구 페이지)

확인했지만 이 글에서 쓰지 않는 자료

한국인터넷진흥원(KISA)의 AI 보안 안내서와 ISO/IEC 42001은 존재 자체는 확인했으나, 이번 검증에서 공식 원문에 자동화된 방식으로 접근하지 못했습니다(KISA는 사이트 접속 오류, ISO는 유료 세부조항 비공개). 따라서 이 두 자료의 구체적인 권고나 요구사항은 이 글에서 인용하지 않습니다. 국내 소상공인의 AI 도입률·평균 절감시간에 관한 공식 원자료도 재확인 결과 찾지 못해 제시하지 않습니다.

실증연구에서 확인된 사실

아래 네 연구는 원문(작업본 또는 출판본)을 직접 확인해 정리했습니다. 다만 Dell'Acqua 외 연구는 2023년 9월 작업본의 수치만 확인했고 2026년 학술지 출판본 본문은 확인하지 못했으므로, 출판본에서 달라졌을 수 있는 품질 향상 폭은 심지 않았습니다. 대상·업무·측정방법이 서로 다르므로 수치를 그대로 더하거나 비교하지 않습니다.

표 4. 실증연구 요약

연구

Noy & Zhang(2023, 작업본)

대상

대졸 전문직 444명

수행 업무

직종별 전문 글쓰기 과제(사전등록 실험)

측정한 것

처리시간, 품질(표준편차 단위)

결과

시간 -0.8 표준편차, 품질 $+0.4$ 표준편차(표준편차는 참가자 점수가 평균에서 떨어진 정도를 재는 단위). 저숙련 참가자에게 이득이 더 큼

일반화 범위

중간 수준 전문직 글쓰기 과제에 한정. 모든 사무 업무의 결과가 아님

연구

Brynjolfsson · Li · Raymond(2023)

대상

고객지원 상담원 5,179명(포춘500 소프트웨어 기업)

수행 업무

실제 고객 응대(사내 생성형AI 어시스턴트 단계적 도입)

측정한 것

시간당 해결 건수

결과

평균 +14%, 저숙련·신입 +34%, 숙련자는 거의 무영향

일반화 범위

대기업 콜센터의 특정 사내 도구 결과. 1인 사업자 규모로 직접 환산 불가

연구

Dell'Acqua 외(2023 작업본)

대상

BCG 컨설턴트 758명

수행 업무

18개 실제 컨설팅 과제(경계 안) + 1개 복합 경영
판단 과제(경계 밖)

측정한 것

과제완료율, 속도, 정확도

결과

경계 안: 과제완료 +12.2%, 속도 +25.1%(작업본
기준. 품질 향상 폭은 판본별 서술이 달라 인용하
지 않음). 경계 밖: 통제군 약 84.5% 정답, AI 사
용군 약 60%·70%로 평균 약 19퍼센트포인트 하
락

일반화 범위

BCG 컨설턴트가 쓴 실험 과제에 한정. 컨설팅 외
업종·소규모 사업장에는 직접 적용할 수 없음

연구 METR(2025)
대상 숙련 오픈소스 개발자 16명(해당 프로젝트에 대한 평균 기여 경력 5년)
수행 업무 실제 오픈소스 저장소 이슈 246건(과제 단위 무작위 배정)
측정한 것 실제 구현시간(화면녹화 기반), 사전 기대·사후 체감
결과 AI 사용 시 19% 더 느려짐. 사전엔 24% 빨라질 것으로 예상, 끝난 뒤에도 20% 빨라졌다고 체감
일반화 범위 저자 스스로 'AI가 대부분 개발자를 가속한다는 근거 아님, 결과가 널리 일반화되거나 다른 영역에 적용된다는 근거 아님'이라고 명시

평균 효과율을 그대로 쓸 수 없는 이유

네 연구를 나란히 놓고 보면 공통된 그림이 보입니다. 반복 빈도나 도구 유무만으로는 결과를 예

측할 수 없고, 다음 세 가지가 결과를 갈랐습니다.

- 숙련도 — Brynjolfsson 외에서는 신입·저숙련자에게 이득이 컸고 숙련자에게는 거의 효과가 없었습니다.
- 업무가 AI 능력 범위 안에 있는지 — Dell'Acqua 외에서는 같은 업무 흐름 안에서 도 범위 밖으로 선택된 과제 하나에서 AI 사용군의 정답 비율이 통제군보다 약 19퍼센트포인트 낮았습니다.
- 검토·재작업 시간을 실측했는지 — METR에서는 참가자들이 끝난 뒤에도 20% 빨라졌다고 믿었지만 실제로는 19% 느려져 있었습니다.

이 네 연구는 대기업 사무직·전문직·컨설턴트·숙련 개발자를 대상으로 했습니다. 국내 소상공인을 대상으로 한 신뢰할 만한 공개 원자료는 이번 검증에서 확인하지 못했습니다. 그래서 이 글은 '평균적으로 몇 % 빨라진다'는 수치를 제시하지 않고, 각자의 업무를 직접 시범 측정하는 절차('시범 전후 측정 방법'절)를 중심에 둡니다.

독자가 준비할 업무 목록

판단을 시작하기 전에, 업무 하나를 정해 아래 항목을 적어보세요. 원본 문서를 보낼 필요는 없습니다.

- 업무명과 업무 단계(예: '문의 접수 → 답변 초안 작성 → 검토 → 발송')
- 한 달 반복 횟수
- 현재 사람이 처리하는 데 걸리는 시간(건당)
- 입력 자료에 개인정보·기밀이 포함되는지
- 실패했을 때(오류가 났을 때) 영향
- 결과를 사람이 확인할 수 있는지
- 고객·외부 시스템에 검토 없이 바로 나가는지

비용 모형

AI가 결과를 만드는 동안 사람이 다른 업무를 볼 수 있다면, 그 시간(AI 실행시간)은 사람이 실제로 들인 시간이 아닙니다. 화면 앞에서 기다려야 하거나 다음 단계가 막힐 때만 AI 실행시간을 사람 투입시간에 더합니다. 아래 첫 번째·두 번째 공식은 '사람이 들인 시간'을, 세 번째 공식은 '업무가 끝나기까지 걸

린 전체 경과시간'을 계산합니다 — 두 값은 같은 것이 아닙니다.

AI 사용 후 월 사람 투입시간

(초기 설정·교육시간 + 통합·전환시간) ÷ 환산개월수 + (입력·지시시간 × 월 반복 횟수) + (AI 실행 중 대기가 필요할 때만: AI 실행시간 × 월 반복 횟수) + (사람 건당 검토시간 × 월 반복 횟수) + (건당 평균 재작업시간 × 월 반복 횟수)

환산개월수는 사이트가 정하지 않습니다. 독자가 계획한 시범기간이나 사용기간을 직접 입력합니다.

월 순 사람시간 절감(순절감시간)

(기존 건당 처리시간 × 월 반복 횟수) - AI 사용 후 월 사람 투입시간

결과가 음수면 '이 업무는 현재 순절감이 없거나 손해일 수 있습니다'로만 표시합니다.

음수라는 이유만으로 자동으로 '맡기지 않음' 등급을 매기지는 않으며, 위험 매트릭스와 함께 해석합니다.

AI 사용 후 건당 처리 경과시간(참고 지표)

입력·지시시간 + AI 실행시간 + 사람 건당 검토시간 +
건당 평균 재작업시간

대기 여부와 관계없이 AI 실행시간을 항상 더
합니다. 고객 응답속도나 납기를 볼 때는 이
값을 참고하고, 사람이 실제로 들인 시간을
볼 때는 위의 '월 순 사람시간 절감'을 씁니
다.

기대 오류손실은 숫자로 곱해 만들지 않습니 다

같은 업무에서 오류율·건당 손실액·처리건수
를 모두 실제로 관측했을 때만 '관측된 오류
율 × 건당 추정손실 × 처리건수'로 금액을
계산합니다. 그런 관측치가 없으면(이 글이
다루는 대부분의 경우) 오류 발생 가능성·오
류 발생 시 영향·오류 발견 난이도·되돌림 가
능성을 각각 낮음/중간/높음으로만 판단하
고, 표 6(위험 축)의 정성 매트릭스로 조합합
니다. 낮음/중간/높음을 숫자로 바꾸어 곱하
거나 점수를 만들지 않습니다.

금액 환산(예: 절감된 시간 × 시간가치)은 독자가 자신의 시간가치를 직접 입력했을 때만 가능합니다. 이 글은 임금이나 시간가치의 기본값을 제시하지 않습니다. 내부 계산은 반올림하지 않고, 화면에 표시할 때만 반올림합니다.

위험 매트릭스

단일 합산 점수보다, 아래 세 축을 각각 낮음/중간/높음으로 판단한 뒤 '강제 하향 규칙' 절의 규칙을 먼저 적용하는 방식을 권합니다.

표 5. 기대 가치 축

항목 반복 빈도
낮음 월 5건 미만
중간 월 5~30건
높음 월 30건 초과

항목
현재 처리시간
낮음
건당 5분 미만
중간
건당 5~30분
높음
건당 30분 초과

항목
병목 정도
낮음
여유 있게 처리 중
중간
가끔 밀림
높음
항상 밀려 있음

항목
표준화 가능성
낮음
매번 달라짐
중간
일부 규칙 있음
높음
규칙이 설명 가능하고 안정적

항목 결과 재사용 가능성
낮음 1회성
중간 부분 재사용
높음 템플릿화 가능

표 6. 위험 축

항목 개인정보·기밀 민감도
낮음 공개자료·비식별 집계
중간 간접식별 가능 정보 일부
높음 직접식별정보·기밀 원문

항목

오류 결과의 심각도

낮음

내부에서만 확인, 영향 적음

중간

재작업·재발송 필요

높음

금전적·법적 책임, 고객 신뢰 훼손

항목

오류 발견 난이도

낮음

숫자·사실을 바로 대조 가능

중간

전문지식이 있어야 확인 가능

높음

AI 답변이 그럴듯해 보여도 틀릴 수 있어, 범위 밖 과제에서는 정답률이 떨어질 수 있음('실증연구에서 확인된 사실' 절의 Dell'Acqua 외 연구 참고)

항목
되돌릴 수 없는 정도
낮음
발송 전 언제든지 수정
중간
정정 공지가 필요
높음
외부 발송·자동 실행 후 회수 불가

항목
고객 직접 영향
낮음
내부 초안
중간
고객 확인 후 전달
높음
검토 없이 고객에게 자동 전달

표 7. 운영 부담 축

항목
사람 검토시간
낮음
건당 1분 미만
중간
건당 1~5분
높음
건당 5분 초과

항목
재작업 가능성
낮음
거의 없음
중간
가끔 발생
높음
자주 발생

항목
설정·교육 난이도
낮음
바로 사용 가능
중간
간단한 설정 필요
높음
별도 교육·문서화 필요

항목 공급자 종속성
낮음 다른 도구로 쉽게 대체 가능
중간 전환에 시간 필요
높음 전환이 사실상 어려움

강제 하향 규칙

다음 중 하나라도 해당하면 순절감시간이 커도 자동화 우선순위를 강제로 낮춥니다. 이 규칙은 법 위반 여부를 확정하는 기준이 아니라 GAON-LAB의 보수적 시범운영 기준입니다.

- 주민등록번호·계좌·카드·건강정보 등 민감정보가 입력됨
- 고객 또는 직원 개인정보가 포함됨
- 회사 기밀 원문을 외부 서비스에 입력해야 함
- 법률·세무·노무·의료의 최종 판단이 포함됨
- 채용·평가·급여·해고 등 사람의 권리에 직접 영향을 미침
- 오류를 사람이 발견하기 어려움

- 오류 결과를 되돌리기 어려움
- 고객에게 검토 없이 자동으로 전달됨
- 결제·주문·환불·계정 변경 등 외부 시스템을 자동 실행함
- 오류 발생 시 대체 절차가 없음
- 사람이 최종 책임을 질 수 있는 구조가 아님

개인정보·고객권리·인사·계약·세무·법률 판단이 포함되면 이 Research의 분류만으로 결정하지 말고, 해당 업무에 적용되는 법령과 전문 검토를 별도로 확인해야 합니다.

판단 절차

1. 판단할 업무 하나를 정하고 '독자가 준비할 업무 목록' 절의 목록을 적습니다.
2. 표 5(기대 가치 축)로 반복 빈도·처리시간·병목·표준화 가능성·재사용 가능성을 낮음/중간/높음으로 매깁니다.
3. 표 6(위험 축)으로 개인정보 민감도·오류 심각도·오류 발견 난이도·되돌림 가능성·고객 영향을 매깁니다.

- 4. 표 7(운영 부담 축)으로 검토시간·재작업·설정 난이도·공급자 종속성을 매깁니다.
- 5. '강제 하향 규칙' 절의 규칙에 하나라도 해당하는지 확인합니다.
- 6. 해당하면 바로 '현재는 말기지 않음'으로 분류하고 비용 계산은 생략해도 됩니다. 실제 고객·직원 개인정보나 기밀 원문은 이 글의 판정만으로 외부 AI에 입력하지 않으며, 'AI 보조만 허용'은 입력자료에서 개인정보·기밀을 빼거나 합성자료로 바꾼 경우에만 씁니다.
- 7. 해당하지 않으면 '비용 모형' 절의 식으로 월 순절감시간을 계산합니다.
- 8. 위험·부담이 낮고 순절감시간이 양수면 '바로 시도 가능', 일부 위험·부담이 있으면 '제한된 자료로 시범 적용'으로 분류합니다.
- 9. '시범 전후 측정 방법' 절에 따라 실제로 측정합니다.
- 10. '중단 기준' 절의 기준에 해당하면 시범을 멈춥니다.

네 가지 판정

표 8. 네 가지 판정 기준

판정

바로 시도 가능

조건

공개자료·비민감 합성자료·고객 정보 없는 업무
템플릿 · 오류가 쉽게 발견됨 · 되돌리기 쉬움 · 외
부 자동 실행 없음 · 사람이 결과를 확인함 · 순절
감시간이 양수

판정

제한된 자료로 시범 적용

조건

일부 변동성이 있음 · 오류 영향이 제한적 · 검토
절차가 있음 · 비식별·최소 자료로 시험 가능 · 실
제 시범 전후 시간을 측정해야 함

판정

AI 보조만 허용

조건

판단 요소가 포함됨 · 고객에게 전달될 가능성이
있음 · 사람 검토가 반드시 필요 · AI는 초안·정리·
후보 생성까지만 담당

판정

현재는 말기지 않음

조건

민감정보·기밀 원문 · 고영향·전문판단 · 오류 발
견이 어려움 · 되돌릴 수 없음 · 고객·직원 권리에
직접 영향 · 무검토 자동 실행

업무별 적용 예시

표 9는 설명용 분류이며 개별 사업장마다 달라질 수 있습니다. 각 행의 분류는 입력자료가 그 행에 적은 조건(공개자료, 합성자료, 이름·연락처·기밀을 지우거나 가린 자료 등)을 충족할 때에만 적용합니다. 실제 고객·직원 개인정보나 기밀 원문이 들어가는 업무는 이 글의 표만으로 외부 SI 입력을 허용하지 않으며, 개인정보 처리의 적법성은 이 글이 판단하지 않습니다. 개인정보 관련 법령, 내부 방침, 전문 검토로 먼저 확인하세요. 그래서 '고객 명단 분석'은 방법 시험과 실제 고객 정보 두 행으로 나뉘었습니다.

표 9. 업무별 적용 예시(설명용)

업무 공개자료 요약
입력자료 공개 문서
민감도 낮음
오류 발견 쉬움
되돌림 쉬움
고객·직원 영향 없음(내부용)
사람 검토 필요
권장 분류 바로 시도 가능
시범 중단 조건 요약이 원문 의미를 반복적으로 왜곡함

업무 반복 이메일 초안
입력자료 고객 정보가 없는 업무 템플릿
민감도 낮음
오류 발견 쉬움
되돌림 쉬움(발송 전)
고객·직원 영향 발송 전 내부
사람 검토 필요
권장 분류 바로 시도 가능
시범 중단 조건 톤·사실관계 오류가 반복됨. 실제 고객 이름·문의 내용이 입력되면 이 분류를 쓰지 않음

업무 회의 메모 정리
입력자료 참석자 이름·연락처·고객 정보·기밀을 지우거나 가린 회의 내용
민감도 중간
오류 발견 쉬움
되돌림 쉬움
고객·직원 영향 내부
사람 검토 권장
권장 분류 제한된 자료로 시범 적용
시범 중단 조건 핵심 결정사항 누락이 반복됨. 이름·고객 정보·기 밀이 가려지지 않은 채 입력되면 즉시 중단하고 이 분류를 쓰지 않음

업무 상품 설명 초안
입력자료 공개 가능한 제품 정보
민감도 낮음
오류 발견 쉬움
되돌림 쉬움(게시 전)
고객·직원 영향 게시 전 내부
사람 검토 필요
권장 분류 바로 시도 가능
시범 중단 조건 과장·허위 표현이 반복됨

업무 고객 문의 답변 초안
입력자료 이름·연락처·주문 정보를 뺀 가상 문의 또는 일반 화한 문의 유형
민감도 중간~높음
오류 발견 중간
되돌림 쉬움(발송 전)
고객·직원 영향 발송 전 고객용
사람 검토 필수
권장 분류 AI 보조만 허용
시범 중단 조건 확인해도 오류를 놓치는 빈도가 높음. 실제 고객 의 이름·연락처·문의 원문이 들어가면 이 분류를 쓰지 않음

업무 매출 집계표 정리
입력자료 고객·거래처를 알아볼 수 없게 합산한 거래 집계
민감도 중간
오류 발견 쉬움(숫자 대조 가능)
되돌림 쉬움
고객·직원 영향 내부
사람 검토 필요
권장 분류 제한된 자료로 시범 적용
시범 중단 조건 집계 오류가 반복 발견됨. 고객·거래처가 드러나는 원본 거래내역이 입력되면 중단

업무 고객 명단 분석 — 방법 시험
입력자료 합성자료, 또는 특정인을 알아볼 수 없도록 처리 한 집계값
민감도 낮음~중간
오류 발견 중간
되돌림 쉬움
고객·직원 영향 내부
사람 검토 필수
권장 분류 제한된 자료로 시범 적용
시범 중단 조건 처리가 충분한지 확신할 수 없음. 원본 명단이 섞 여 들어감

업무 고객 명단 분석 — 실제 고객 정보
입력자료 실제 고객의 이름·연락처·구매 이력 등
민감도 높음
오류 발견 어려움
되돌림 중간
고객·직원 영향 고객 권리에 영향 가능
사람 검토 필수
권장 분류 현재는 말기지 않음
시범 중단 조건 이 표로는 외부 AI 입력을 허용하지 않음. 법령·내부 방침·전문 검토를 먼저 확인

업무 급여·인사 평가
입력자료 직원 평가·급여 정보
민감도 높음
오류 발견 어려움
되돌림 어려움
고객·직원 영향 직원 권리에 직접 영향
사람 검토 필수(그래도 권장 안 함)
권장 분류 현재는 말기지 않음
시범 중단 조건 해당 없음(애초에 시작하지 않음)

업무 세금 신고 판단
입력자료 재무·과세 정보
민감도 높음
오류 발견 어려움
되돌림 어려움
고객·직원 영향 법적 책임
사람 검토 필수(그래도 권장 안 함)
권장 분류 현재는 말기지 않음
시범 중단 조건 해당 없음

업무 계약서 최종 판단
입력자료 계약 조건
민감도 높음
오류 발견 어려움
되돌림 어려움
고객·직원 영향 법적 책임
사람 검토 필수(그래도 권장 안 함)
권장 분류 현재는 말기지 않음
시범 중단 조건 해당 없음

업무 고객에게 자동 발송
입력자료 발송 내용
민감도 가변
오류 발견 어려움(발송 후)
되돌림 어려움
고객·직원 영향 고객에게 직접
사람 검토 필수(발송 전 게이트)
권장 분류 AI 보조만 허용
시범 중단 조건 검토 없이 발송되는 경로가 하나라도 발견됨

업무 주문·결제·환불 자동 실행
입력자료 거래 실행 정보
민감도 가변
오류 발견 어려움
되돌림 불가능~어려움
고객·직원 영향 고객·자금에 직접
사람 검토 필수
권장 분류 현재는 말기지 않음
시범 중단 조건 해당 없음

표 8의 '바로 시도 가능'은 공개자료, 비민감 합성자료, 고객 정보가 없는 업무 템플릿을 쓸 때의 분류입니다. 같은 업무라도 실제 참석자 이름이나 고객 정보가 들어가면 민감도가 올라가므로 '제한된 자료로 시범 적용' 이상으로 다시 분류합니다. 반복 이메일 초안은

고객 정보가 없는 템플릿일 때만 '바로 시도 가능'이며, 실제 고객 문의에 답하는 초안은 위의 '고객 문의 답변 초안'으로 분류하며, 이는 실제 고객의 이름·연락처·문의 원문을 외부 AI에 넣어도 된다는 뜻이 아닙니다.

합성 사례

아래 세 사례는 계산 방법을 보여주기 위해 만든 설명용 합성 사례이며, 실존 인물·실제 고객 사례가 아닙니다. 숫자는 공식 통계가 아니라 계산을 설명하기 위한 가정이며, 독립적으로 다시 계산해 원고와 일치함을 확인했습니다.

사례 A — 공개자료 기반 상품 설명 초안(바로 시도 가능 후보)

표 10. 사례 A 입력값과 결과

항목
월 반복 횟수
값
40건(설명용 가정)

항목

기존 건당 처리시간

값

25분(설명용 가정)

항목

입력·지시시간

값

2분(설명용 가정)

항목

AI 실행시간

값

3분(설명용 가정)

항목

AI 실행 중 대기 필요

값

아니오 — 여러 건의 상품 정보를 먼저 입력해 두고, 결과가 나온 것부터 모아서 검토하는 방식으로 가정(설명용 가정)

항목

사람 건당 검토시간

값

5분(설명용 가정)

항목

건당 평균 재작업시간

값

2분(설명용 가정)

항목

초기 설정시간 + 통합·전환시간

값

60분(설명용 가정)

항목

환산개월수(독자 입력)

값

3개월(설명용 가정)

항목

AI 사용 후 월 사람 투입시간(계산값)

값

380분($=60 \div 3 + (2+5+2) \times 40$ — 대기가 필요
없어 AI 실행시간은 포함하지 않음)

항목

월 순 사람시간 절감(계산값)

값

620분(약 10.3시간) — 기존 1,000분에서 380분
으로

항목

AI 사용 후 건당 처리 경과시간(참고)

값

12분(=2+3+5+2) — 기존 25분보다 짧아짐. 사람 투입시간과 다른 지표라 별도로 표시

항목

민감도·되돌림·고객영향

값

낮음 · 쉬움 · 게시 전 내부

항목

권장 분류

값

바로 시도 가능

사례 B — 고객 문의 답변 초안(AI 보조만 허용 후보)

표 11. 사례 B 입력값과 결과

항목

월 반복 횟수

값

200건(설명용 가정)

항목

기존 건당 처리시간

값

8분(설명용 가정)

항목

입력·지시시간

값

1분(설명용 가정)

항목

AI 실행시간

값

2분(설명용 가정)

항목

AI 실행 중 대기 필요

값

예 — 고객이 답변을 기다리고 있어 결과가 나올 때까지 대기한 뒤 바로 검토·발송하는 방식으로
가정(설명용 가정)

항목

사람 건당 검토시간

값

4분(설명용 가정, 고객 발송 전 확인 포함)

항목

건당 평균 재작업시간

값

1.5분(설명용 가정)

항목

초기 설정시간 + 통합·전환시간

값

180분(설명용 가정)

항목

환산개월수(독자 입력)

값

3개월(설명용 가정)

항목

AI 사용 후 월 사람 투입시간(계산값)

값

1,760분(=180÷3 + (1+2+4+1.5)×200 — 대기
가 필요해 AI 실행시간도 포함)

항목

월 순 사람시간 절감(계산값)

값

-160분 — 기존 1,600분보다 오히려 160분 더
들

항목

AI 사용 후 건당 처리 경과시간(참고)

값

8.5분(=1+2+4+1.5) — 기존 8분과 거의 같음. 처리 속도는 크게 달라지지 않았지만 사람이 들이는 시간은 오히려 늘었음

항목

결과를 바꾸는 질문

값

검토시간을 4분에서 더 줄일 수 있는가? 초안 품질이 낮아 검토가 오래 걸리는가, 아니면 발송 전 확인 자체가 원래 필요한 절차인가?

항목

입력자료 조건

값

이름·연락처·주문 정보를 뺀 문의 또는 가상 문의만 사용(설명용 가정)

항목

민감도·되돌림·고객영향

값

중간~높음 · 발송 전엔 쉬움 · 고객에게 직접 전달

항목

권장 분류

값

AI 보조만 허용(순 사람시간 절감이 음수이므로
시간 절감을 목적으로 도입하지 말고, 검토 품질
이나 응대 일관성 등 다른 목적이 있을 때만 검토)

사례 C — 급여·인사평가(현재는 말기지 않음)

표 12. 사례 C — 강제 하향 규칙 적용 예시

항목

업무

값

직원 급여 조정안 또는 인사평가 초안 작성(설명
용 가정)

항목

해당하는 강제 하향 규칙

값

채용·평가·급여·해고 등 사람의 권리에 직접 영향
을 미침 / 되돌리기 어려움 / 오류를 발견하기 어
려움

항목

비용 계산 진행 여부

값

진행하지 않음 — 강제 하향 규칙에 해당하면 순
절감시간이 아무리 커도 판정에 영향을 주지 않으
므로 계산을 생략합니다.

항목

권장 분류

값

현재는 말기지 않음

항목

대안

값

직원 개인정보를 제거한 일반적인 평가 기준 문서
작성 보조 정도까지만 검토하고, 개별 직원에 대
한 최종 평가·급여 결정에는 사용하지 않습니다.

사례 A·B의 시간을 두 가지 방식으로 읽기

표 10과 표 11의 '월 사람 투입시간'(사례 A 380
분, 사례 B 1,760분)은 1회성 준비시간(초기 설
정·통합·전환)을 환산개월수 3개월로 나눠 매달
같은 크기로 반영한 월 환산값입니다. 첫 달에
실제로 들어가는 시간이 아닙니다. 준비시간을
첫 달에 전부 반영하면 사례 A는 60 +

$(2+5+2) \times 40 = 420$ 분, 사례 B는 $180 + (1+2+4+1.5) \times 200 = 1,880$ 분이 됩니다. 아래 표는 두 방식을 같은 설명용 가정으로 나란히 놓은 것입니다.

표 13. 월 환산값과 첫 달 값 비교

사례 A
기존 월 사람 투입시간 1,000분
AI 사용 후, 준비시간 월 환산(표 10·11) 380분
월 환산 순절감 +620분
AI 사용 후, 준비시간 첫 달 전부 반영 420분
첫 달 순절감 +580분

사례 B
기존 월 사람 투입시간 1,600분
AI 사용 후, 준비시간 월 환산(표 10·11) 1,760분
월 환산 순절감 -160분
AI 사용 후, 준비시간 첫 달 전부 반영 1,880분
첫 달 순절감 -280분

환산개월수 3개월도 독자가 정하는 계획 가정입니다. 기간을 바꿔도 결론이 유지되는지 확인하고, 실제 준비시간과 첫 달 합계, 환산값을 함께 기록하세요. 사례 B는 준비시간을 아예 0으로 두어도 매달 $(1+2+4+1.5) \times 200 = 1,700$ 분이 들어 기존 1,600분보다 100분 많으므로, 환산개월수를 늘려도 부호가 바뀌지 않습니다. 또 사람 검토시간과 재작업시간은 겹쳐 세지 않습니다. 계획한 확인·수정은 검토시간, 오류 때문에 다시 요청하거나 고쳐 만든 시간은 재작업시간으로 나누어 기록합니다.

검토시간 손익경계(다른 입력값을 고정했을 때)

손익경계 검토시간 = 기존 건당 처리시간 - 입력·지시시간
- (대기가 필요할 때만 AI 실행시간) - 건당 평균 재작업
시간 - 초기 설정·전환시간 ÷ (환산개월수 × 월 반복 횟
수)

손익경계는 AI를 써도 사람 시간이 더 들지 않는 최대 검토시간을 뜻합니다. 사람 건당 검토시간이 이 값보다 작으면 월 순 사람시간 절감이 양수이고, 같으면 0, 크면 음수입니다. 검토시간 하나만 바꾸고 나머지 입력값은 그대로 둔 계산입니다.

3.2분은 검토를 줄이라는 목표가 아닙니다

사례 A의 손익경계는 $25 - 2 - 0 - 2 - 60 \div (3 \times 40) = 20.5$ 분입니다. 가정한 검토시간 5분은 이보다 작아 순절감이 양수입니다. 사례 B는 $8 - 1 - 2 - 1.5 - 180 \div (3 \times 200) = 3.2$ 분이어서, 가정한 검토시간 4분은 이 경계를 넘어 순절감이 음수입니다. 3.2분은 이 설명용 가정에서 순절감이 0이 되는 손익경계일 뿐이며, 필수 검토를 생략하거나 검토시간을 그 수치 이하로 줄이라는 목표가 아닙니다. 고객에게 나가는 답변의

확인을 줄여 3.2분 아래로 맞추면 사례 B가 전제한 품질 기준이 깨집니다. 필요한 확인은 유지한 채 입력자료 정리나 업무 범위를 바꾸면 시간이 달라지는지 시험하고, 그래도 경계를 넘지 못하면 이 범위에서는 기존 방식을 유지하세요.

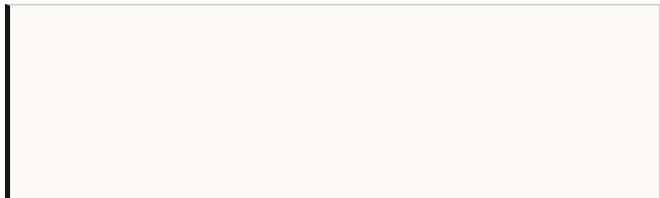
준비시간이 언제 회수되는지도 따로 볼 수 있습니다. 사례 A에서 준비시간을 뺀 건당 절감은 $25 - (2+5+2) = 16$ 분이므로 60분의 준비시간을 상쇄하려면 $60 \div 16 = 3.75$ 건, 온전한 건수로는 최소 4건이 필요합니다. 이는 시간만 계산한 값이며, 유료 도구의 비용 회수나 실제 성공 여부를 보여주지 않습니다.

시범 전후 측정 방법

고정된 '2주'나 '10건' 같은 숫자를 운영 기준으로 만들지 않습니다. 대신 아래 절차를 따릅니다.

1. 같은 종류의 업무를 사람이 처리한 기준시간을 기록합니다.
2. AI를 사용했을 때 처리시간을 기록합니다.

3. 사람 검토시간을 별도로 기록합니다.
4. 재작업시간을 별도로 기록합니다.
5. 오류 유형과 발견 시점을 기록합니다.
6. 시험 전에 정상 건뿐 아니라 입력 누락·예외·보류 건을 어떻게 포함할지 정해 둡니다.
7. 같은 문서를 사람이 먼저 끝낸 뒤 AI로 다시 하면 이미 아는 내용이 섞이므로, 비슷한 과제를 번갈아 맡기거나 나누어 배정합니다.
8. 고객에게 전달되기 전 사람이 반드시 확인합니다.
9. 순절감시간을 계산합니다('비용 모형' 절의 공식).
10. '중단 기준' 절의 기준에 해당하는지 확인합니다.
11. 반복 횟수와 측정 기간을 결과와 함께 남깁니다.
12. 업무 조건이 달라지면 별도 표본으로 다시 측정합니다.



평균값만 남기지 말고 범위·예외·실패 사례도 함께 기록하세요. METR 연구에서 보듯 체감과 실측이 다를 수 있습니다.

효율 개선의 조건

순 사람시간 절감이 양수라도 품질이 나빠졌다면 효율 개선으로 판단하지 않습니다. 측정을 시작하기 전에 완료 기준과 허용 가능한 오류 수준을 먼저 정하고, 그 기준을 지켰을 때만 '효율 개선'으로 표시합니다. 측정 건수가 적거나 업무 난이도가 섞여 있으면 결론을 확정하지 말고 표본을 늘려 다시 측정합니다. 몇 건이면 충분하다는 고정 기준은 두지 않으며, 같은 결과가 반복되는지, 예외 사례가 포함됐는지, 건수가 적어 우연이 크게 작용하지 않았는지를 기록으로 판단합니다.

중단 기준

시범 적용 중 아래 중 하나라도 발생하면 즉시 멈추고 재검토합니다. 중단은 실패가 아니라 시범 설계에 원래 포함된 안전장치입니다.

- 예상하지 못한 개인정보·기밀이 결과나 로그에 남는 것을 발견했다.
- 허용하지 않은 외부 저장·전송이 발생했다.
- 오류를 검토 과정에서 놓친 사례가 반복해서 확인됐다.
- 사람이 오류를 안정적으로 발견하지 못한다(검토를 거치고도 오류가 계속 새어 나간다).
- 고객에게 검토 없이 결과가 나간 경로가 하나라도 발견됐다.
- 검토·재작업시간이 절감시간보다 커졌다(순 사람시간 절감이 여러 차례 측정에서 계속 음수로 나온다).
- 재작업이 예상보다 잦아 검토 부담이 오히려 커졌다.
- 외부 시스템에서 의도하지 않은 실행(주문·결제·환불·계정 변경 등)이 발생했다.

- 품질 기준(완료 기준·오류 허용 수준)을 지속적으로 충족하지 못한다.
- 도구를 바꾸면서 기존에 있던 확인·통제 절차가 더 이상 작동하지 않게 됐다.
- 사고가 발생했을 때 되돌아갈 수 있는 대체 절차가 없다.
- '강제 하향 규칙' 절의 규칙에 새로 해당하게 됐다(예: 업무 범위가 넓어져 고객 정보가 포함됨).

한 건의 결과를 보류하는 것과 시험 전체를 멈추는 것은 다릅니다

검토 중 틀린 가격이 들어간 초안 하나를 발견했다면 그 결과는 즉시 사용을 보류합니다. 여러 번 틀려야 보류하는 것이 아닙니다. 원자료로 바로잡고 무엇이 틀렸는지 기록해 두세요. 위 목록은 개별 오류를 고치는 단계가 아니라 시험 전체를 멈추고 설계를 다시 확인해야 하는 경우입니다. 허용하지 않은 자료가 입력됐거나, 사람 확인 없이 외부로 발송·실행되는 경로가 발견됐거나, 책임자가 확인할 수 없는 중요한 판단이 필요해졌다면 한 건만으로도 시험 전체를 멈춥니다. 반면 개별 오류는 고칠 수 있어도 같은 종류의 오류를 검

토에서 놓치는 일이 반복되거나 재작업이 늘어 순절감시간이 계속 음수라면, 입력을 바꿔 다시 시험할지, 범위를 좁힐지, 기존 방식을 유지할지 기록을 근거로 결정합니다.

대안 설명

순절감시간이 기대에 못 미치거나 음수로 나오는 원인을 'AI가 별로다' 하나로 단정하지 않습니다. 다음처럼 다른 원인일 수 있습니다.

- 검토시간을 처음부터 과소평가했다 — METR 연구처럼 체감과 실측 사이에 큰 차이가 있을 수 있습니다.
- 업무가 AI 능력 범위 밖에 있다 — Dell'Acqua 외 연구처럼 겉보기엔 비슷한 난이도라도 범위 밖에서는 성과가 떨어질 수 있습니다.
- 숙련자가 이미 효율적으로 일하고 있다 — Brynjolfsson 외 연구처럼 숙련자에게는 효과가 작을 수 있습니다.
- 재작업 기준이 너무 엄격하거나 반대로 느슨하다.

- 초기 설정·교육 비용을 시범기간에 과도하게 몰아 계산했다 — 환산개월수를 늘려 다시 계산해봅니다.
- 학습 효과로 설정 부담이 줄 수 있지만, 측정하지 않은 감소를 미래 비용에 미리 반영하지 않았다 — 환산개월수를 늘려 다시 계산한 값과 실제로 관측한 부담 감소는 다릅니다.
- 업무 변동성이 생각보다 높아 표준화가 되지 않은 상태에서 시작했다.

결과 해석

점수나 단일 합격선 대신 아래 질문으로 해석합니다.

- 강제 하향 규칙에 해당하는 항목이 있는가?
- 순절감시간(월 순 사람시간 절감)이 양수인가, 음수인가?
- 시범 전후 품질(누락·사실오류·형식오류, 재작업 발생 여부)이 미리 정한 기준을 유지했는가?
- 검토시간을 실제로 측정했는가, 아니면 짐작만 했는가?

- 오류가 발생했을 때 사람이 발송 전에 반드시 확인하는 절차가 있는가?
- 한 번의 예외 상황(범위 밖 업무)이 결과를 크게 바꾸는가?
- 시범 기간과 반복 횟수를 늘려도 같은 결론이 나오는가?

데이터 처리 경계

초기 문의나 시범 설계에는 다음 요약만 있으면 됩니다: 업무명, 업무 단계, 월 반복 빈도, 현재 소요시간, 실패했을 때의 영향, 사람이 확인할 수 있는지, 사용하는 자료의 종류, 원하는 결과물.

보내지 말아야 할 것

고객 이름·전화번호·이메일, 직원 개인정보, 주민등록번호, 계좌·카드·결제정보, 건강정보, 회사 기밀 원문, 계약서 원문, 로그인 정보·비밀번호, API 키, 내부 시스템 전체 화면·데이터베이스는 어디에도 입력하거나 보내지 마세요. 초기 문의에는 내부 문서 원문이나 고객·직원 데이터를 첨부하지 마세요. 업무명

·단계·빈도·실패 위험을 비식별 요약으로 정리한 뒤, 진행 가능 여부와 필요한 자료 범위를 먼저 협의합니다.

이 Research의 한계

- 인용한 실증연구 4건은 모두 대기업 사무직·전문직·컨설턴트·숙련 개발자를 대상으로 했으며, 국내 소상공인을 대상으로 한 공식 원자료는 확인하지 못했습니다.
- 오류 발생 확률과 손실액에 관한 공신력 있는 자료가 없어 비용모형의 '기대 오류손실'은 금액으로 계산하지 않고 정성 분류로만 다룹니다.
- KISA AI 보안 안내서와 ISO/IEC 42001은 이번 검증에서 공식 원문에 접근하지 못해 이 글의 근거로 쓰지 않았습니다.
- 개인정보 보호법의 자동화된 결정 관련 조항과 AI기본법의 투명성 확보 의무 조항은 국가법령정보센터 원문을 직접 열람하지 못해 신뢰할 수 있는 법령 미러로만 방향을 확인했으며, 구체적인 조문 번호나 의무 내용을 이 글에서 단정하지 않습니다.

- 위험 매트릭스의 낮음/중간/높음 기준은 GAON-LAB이 설계한 실무 규칙이며 학술적으로 검증된 척도가 아닙니다.
 - 합성 사례는 설명을 위한 가정이며 실제 업무 조건과 다를 수 있습니다.
 - 이 글은 특정 AI 제품의 성능을 비교하거나 추천하지 않습니다.
 - 이 글의 판정은 법적 판정이 아니며, 법률·세무·노무·의료 판단이 포함된 업무는 별도 전문 검토가 필요합니다.
-

재현 방법

이 글의 코드를 공개하지 않아도 스프레드시트에서 재현할 수 있습니다.

1. 표 2(조작적 정의)의 입력값을 각각 하나의 셀로 만듭니다.
2. '비용 모형' 절의 두 공식을 그대로 셀 수식으로 옮깁니다.
3. 표 5~7의 낮음/중간/높음 기준을 참고해 업무를 채점합니다.
4. '강제 하향 규칙' 절의 규칙에 하나라도 해당하는지 먼저 확인합니다.

5. 해당 사항이 없으면 비용 모형으로 순절감시간을 계산합니다.
6. 표 8의 네 가지 판정 기준과 대조합니다.
7. '시범 전후 측정 방법' 절의 절차로 실제 시범을 진행하고 같은 항목을 다시 측정합니다.
8. 이 글이 사용한 출처의 발행일(표 3 참고)을 함께 기록해두면 나중에 갱신 여부를 확인하기 쉽습니다.

이 글로 알 수 있는 것과 알 수 없는 것

알 수 있는 것

업무 하나를 판단하는 절차와 기준, 검토시간을 포함한 순절감시간 계산법, 개인정보·오류·되돌림 위험을 함께 보는 방법, 시범 적용과 중단 기준.

알 수 없는 것

특정 AI 도구의 성능, 국내 소상공인의 평균 도입률·절감시간, 특정 업종에 적용했을 때의 정확한 효과 크기, 개인정보 처리의 최종 적

법성 판단, 이 글의 분류만으로 내리는 법률·세무·노무 결정.

다음 행동과 서비스 연결

아래는 광고가 아니라 이 Research 혼자서 답할 수 있는 것과 개인별 확인이 필요한 것 사이의 경계를 보여주는 순서입니다.

1. 먼저 '독자가 준비할 업무 목록' 절의 업무 목록을 자신의 업무로 채워봅니다.
2. 무료 AI 활용도 진단으로 스스로 점검합니다.
3. 업무 구조와 우선순위를 함께 정리하고 싶다면 AI·업무자동화 프로그램을 검토합니다. 이 프로그램은 어디에 무엇을 적용할지 판단하는 진단과 우선순위 설계를 맡으며, 실제 시스템 구축·계정 연결·라이선스 구매 대행은 제공하지 않습니다.
4. 진행 가능 여부를 상담하려면 문의를 남겨주세요. 초기 문의에는 해결하려는 문제, 준비된 자료의 종류, 원하는 결과물, 희망 일정만 알려주시면 됩니다. 내부 문서 원문이나 고객·직원 데이터는 첨부하지 마시고, 업무명·단계·빈도·실패 위험을 비식별 요약으로 정리해 주세요.

분석의 한계와 참고자료

이 글은 GAON-LAB이 공식 발행기관 페이지에서 직접 확인한 원자료(개인정보보호위원회, NIST, 그리고 동료평가를 마쳤거나 작업본으로 공개된 실증연구 4건)만을 핵심 근거로 사용했습니다. 검색 결과나 언론 보도로만 확인된 수치, 그리고 이번에 원문에 접근하지 못한 자료(KISA 보안 안내서, ISO/IEC 42001 세부조항)는 핵심 근거로 쓰지 않았습니다.

국내 소상공인의 AI 도입률이나 평균 절감시간처럼 공식 원자료가 없는 값은 만들어 제시하지 않았습니다. 이 글의 비용 모형과 위험 매트릭스는 GAON-LAB이 설계한 실무 판단 규칙이며, 법률·세무·노무·의료 판단을 대신하지 않습니다.

출처

개인정보보호위원회, 「생성형 인공지능(AI) 개발·활용을 위한 개인정보 처리 안내서」 (2025.8.)

https://www.privacy.go.kr/front/bbs/bbsView.do?bbsNo=BBSMSTR_000000000049&bbscttNo=20836

NIST, 「Artificial Intelligence Risk Management Framework (AI RMF 1.0)」, NIST AI 100-1(2023.1.)

<https://doi.org/10.6028/NIST.AI.100-1>

NIST, 「Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile」, NIST AI 600-1(2024.7.)

<https://doi.org/10.6028/NIST.AI.600-1>

Noy, S. & Zhang, W.(2023), 「Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence」, MIT Working Paper

https://economics.mit.edu/sites/default/files/inline-files/Noy_Zhang_1.pdf

Brynjolfsson, E., Li, D. & Raymond, L. R.(2023), 「Generative AI at Work」, NBER Working Paper 31161

<https://www.nber.org/papers/w31161>

Dell'Acqua, F. 외, 「Navigating the Jagged Technological Frontier」, Organization Science 37(2) (2026). 이 글의 수치는 2023.9. 작업본(SSRN 4573321) 기준

<https://doi.org/10.1287/orsc.2025.21838>

Dell'Acqua, F. 외(2023), 「Navigating the Jagged Technological Frontier」, SSRN 4573321, 2023-09-22 작업본(MIT Sloan 게시 PDF) — 초록의 758명

·12.2%·25.1%·19%포인트와 본문의 84.5%·60%·70% 확인, 확인일 2026-10-09

<https://mitsloan.mit.edu/sites/default/files/2023-10/SSRN-id4573321.pdf>

METR(2025), 「Measuring the Impact of Early-2025 AI on Experienced Open-Source Developer Productivity」

<https://metr.org/blog/2025-07-10-early-2025-ai-experienced-os-dev-study/>

METR(2025), 「Measuring the Impact of Early-2025 AI on Experienced Open-Source Developer Productivity」, arXiv:2507.09089 (논문 본문 — 개발자의 해당 저장소 평균 경력 5년 확인용), 확인일 2026-10-09
<https://arxiv.org/abs/2507.09089>