정답지를 숨기자
AI가 특허를 배우기 시작했다

AI는 특허 문서를 그럴듯하게 틀립니다. 그래서 **정답지를 숨기고**, AI가 혼자 쓴 문서를 다른 AI가 채점하게 했습니다.

① 정답지는 봉투 속에

눈 가린 시험. 신약 시험처럼 숙련자의 완성본을 숨기고 AI가 혼자 쓴 뒤, 다른 AI가 봉인을 열어 채점

② 짐작은 규칙이 못 된다

측정으로만 만드는 규칙. 숙련자 명세서 2,000여 문단을 28개 항목으로 재고, 규칙·관찰·편차로 등급을 매김

③ 범인은 AI가 아니라 매뉴얼

사고 조사. 비행기 사고 조사처럼. 틀릴 때마다 뿌리 원인을 적고, 같은 실수가 또 나면 "재발"로 따로 조사

④ 모든 규칙에 각주

출처가 달린 규칙. 논문 각주처럼. 규칙 줄마다 누가 언제 왜 정했는지 — 339곳

평일 아침마다 시험이 저절로 돌고, 2년간 틀린 이유가 판단 기준 약 120만 자로 쌓였습니다.

아무도 없는 사무실에서, AI가 먼저 시험을 봤습니다

🌙

05:53

AI가 먼저 일어나
담당 사건 마감 정리

⏰

08:53

예약해 둔 시험 시작
사람은 아무것도 안 시킴

✉️

봉인

숙련자가 반년 전 낸
완성본을 봉투 속에

✍️

29분

처음 자료와 매뉴얼만으로
AI 혼자 작성(~09:22)

🔍

개봉·대조

다른 AI가 두 문서의
다른 곳을 원문 그대로

🗂️

세 갈래

규칙 없음 / 안 지킴
/ 사람만의 판단

📘

매뉴얼에 한 줄

원인은 AI가 아니라
매뉴얼의 빈칸

🧑‍💼

출근

"여기는 결정해 주세요"
남긴 곳만 사람이 판단

그날 짚은 곳은 **대부분 셋째 갈래**였습니다. 숙련자가 처음 자료에 없던 예를 더하거나 권리 범위를 스스로 정한 자리였고, 이런 곳은 **일부러 규칙으로 만들지 않고 기록만** 남겼습니다.

한 번의 시험에서 세 가지가 드러났습니다

범인은 매뉴얼 — 네 번째 재발

AI가 "~를 낮출 수 있다"는 효과를 주장해 놓고, 그 효과가 성립하지 않는 실시예를 함께 썼습니다.

앞선 대책 ① 샘플링 주기 문장 ✓
앞선 대책 ② 판정 규칙 문장 ✓
앞선 대책 ③ 모드 정의 문장 ✓
효과 문장 × — 빠져나감

새 원칙: 단언한 문장은 모든 실시예에 하나씩 대입한다

그리고 하나 더 — 반증

본문을 쓰기 전, AI가 처음 잡은 청구항을 깨뜨릴 문헌부터 다시 찾았습니다.

조사 보고서에 없던 외국 특허 1건

처음 청구항 그대로였다면 신규성이 걸렸을 문헌입니다. 숙련자의 완성본도 범위를 좁혀 이 문헌을 피해 두었습니다.

채점 AI는 반대쪽도 적었다

오탈자처럼 **숙련자 쪽이 틀리고 AI 쪽이 맞은 자리**도 있었습니다.

그래도 정답지는 바꾸지 않는다

기준은 숙련자의 완성본으로 두고, 그 자리를 AI 점수에 유리하게 쓰지 않았습니다. 정답지를 그때그때 바꾸면 시험이 무너집니다.

2026-10-22 채점 기록과 매뉴얼 정정 기록에서 사건·고객·발명 내용을 빼고 옮겼습니다.

특허 문서는 계산이 아니라 세 가지를 한꺼번에 따지는 판단입니다

법
어디까지를 권리로 청구할 수 있나

기술
심사관이 든 다른 특허와 무엇이 다른가

사정
고객은 무엇을 지키고 싶은가

AI가 쓴 답

"동작은 같아도, 쓰는 목적이 다릅니다."

특허에서는 지는 답

694
2년간 업무 대화(개)

1,200+
"틀렸다" 지적(번)

지적은 쌓이는데 알 수 없었습니다. **내 지적이 옳은 규칙인지, AI가 숙련자에 얼마나 가까워졌는지.** 그래서 질문을 "무엇을 더 가르칠까"에서 **"배우고 있는지 어떻게 잴까"**로 바꿨습니다.

신약 시험처럼, AI는 정답지를 못 본 채 씁니다

1 봉인

숙련자 완성본을 숨긴다

처음 자료만 꺼냄
(발명 설명서·면담 기록)

→

2 작성

AI가 매뉴얼대로 혼자 쓴다

정답지는 볼 수 없음

→

3 개봉·채점

다른 AI가 나란히 놓고 짚는다

쓴 AI ≠ 채점 AI

→

4 분류

다른 곳을 세 갈래로

규칙 없음 / 규칙 안 지킴
/ 사람만의 판단

왜 채점자를 따로 두나

자기 글을 자기가 채점하면 후해집니다. 채점 기준은 AI의 자기 평가가 아니라 사람 숙련자의 실제 완성본입니다.

왜 세 갈래로 나누나

처방이 다르기 때문입니다. 없으면 새로 쓰고, 안 지켰으면 왜 안 걸렸는지 고치고, 사람만의 판단은 **일부러 규칙으로 만들지 않습니다.**

> 회사 시스템에서 숙련자 완성본을 가져오는 일(읽기만)과 시험 실행은 **예약 작업이 평일 아침 사람 없이** 합니다.

"숙련자는 보통 이렇게 쓴다"를 짐작하지 않고 잽니다

숙련자 명세서 2,000여 문단을 28개 항목으로 재는 프로그램을 AI와 만들었습니다. 아래는 실제 측정표의 일부입니다(문서 번호는 가림).

측정 항목	표본1	표본2	표본3	표본4	표본5	표본6	표본7	표본8	등급
"특징으로 하는" 청구항	0항	0항	0항	0항	0항	0항	0항	0항	규칙(0건)
본문 소제목	0/323문단	0/241	0/291	0/267	0/272	0/357	–	0/253	규칙(0건)
방법 청구항 전제부	있음	–	있음	있음	있음	있음	있음	있음	규칙
청구항 끝 쉼표	1/18	20/20	6/20	7/20	6/12	0/20	19/20	1/20	편차(분포)
다중 인용 청구항	0건	0건	0건	0건	0건	0건	1건	0건	편차

규칙 = 기회 있는 표본 3건 이상 전부 일치 **관찰 = 1~2건** **편차 = 숙련자끼리도 갈림**

"–"는 쓸 **기회가 없던 표본**이라 셈에서 뺍니다. 10월 22일 AI가 자기 문서를 "짧다"고 판정했을 때도 글자 수를 재 보니 숙련자와 거의 같았고, AI는 **없는 내용을 지어 넣지 않고 멈췄습니다**.

비행기 사고 조사처럼, 틀린 이유의 뿌리를 적습니다

① 사고

심사관: "이미 있는 것과 동작이 같다."

AI: "동작은 같아도 목적이 다릅니다."

특허에서는 지는 답.

→

② 네 줄 조사

네가 한 말: 목적이 다르다
바른 것: 효과의 근거가 있을 때만
뿌리 원인: 매뉴얼에 조건이 없었다
다음에 막는 법: 조건을 넣는다

→

③ 재발 검사

규칙을 넣기 전에 사고 이력부터 봅니다.

같은 사고가 또 났다면

새 규칙을 얹지 않고, **앞선 대책이 왜 안 먹혔는지**를 새 사고로 조사합니다.

새 규칙에는 **그 규칙을 시험하는 문제도 하나** 붙입니다. 10월 22일의 원칙 한 줄도 이 재발 검사에서 나왔습니다 — **같은 종류의 네 번째 실수**였습니다.

실제 기록(2026-09-07)에서 발명 내용을 빼고 옮겼습니다.

논문 각주처럼, 규칙마다 누가 언제 왜가 붙어 있습니다

```
# 실제 매뉴얼의 한 줄 (그대로 옮김)
규칙(사용자 확정, 2026-10-13) — "어떤 특허문헌을 찾을일이 있다면, 그건 그냥 키워트에 연결해서 찾아보자, 구글 패이턴트 보다 이편이 빠르고 간편하니까 그렇게 하자"
# 확인 수준도 표시 — 실측 / ⚖️ 관찰(n=1) / 추측
```

339
출처 꼬리표가 붙은
규칙 줄

1,200+
꼭 지키거나 멈출 자리
경고 표시

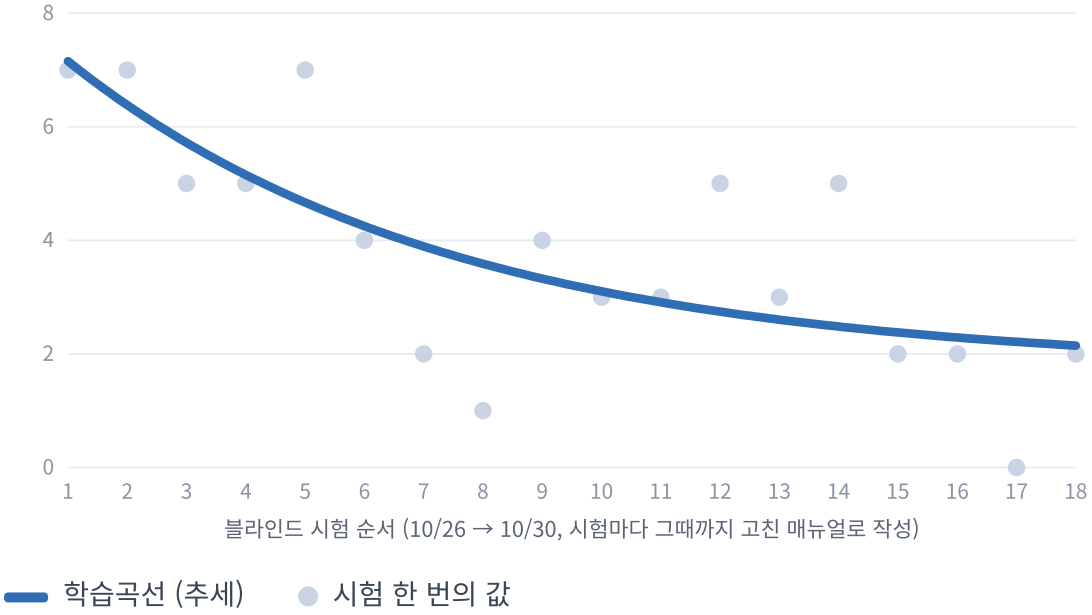
676
매뉴얼을 고친 이력
(대조 기록 149 포함)

120만
판단 기준(자)
업무 매뉴얼 22권

AI가 짐작한 것은 규칙이 되지 못합니다. **"확인 안 됨" 칸**에 따로 두고, 근거가 모일 때까지 기준처럼 쓰지 않습니다.

규칙이 쌓이자, 막을 수 있는 실수가 6.3곳에서 1.3곳으로 줄었습니다

시험 한 번에 나온 "규칙으로 막을 수 있는 실수" (곳)



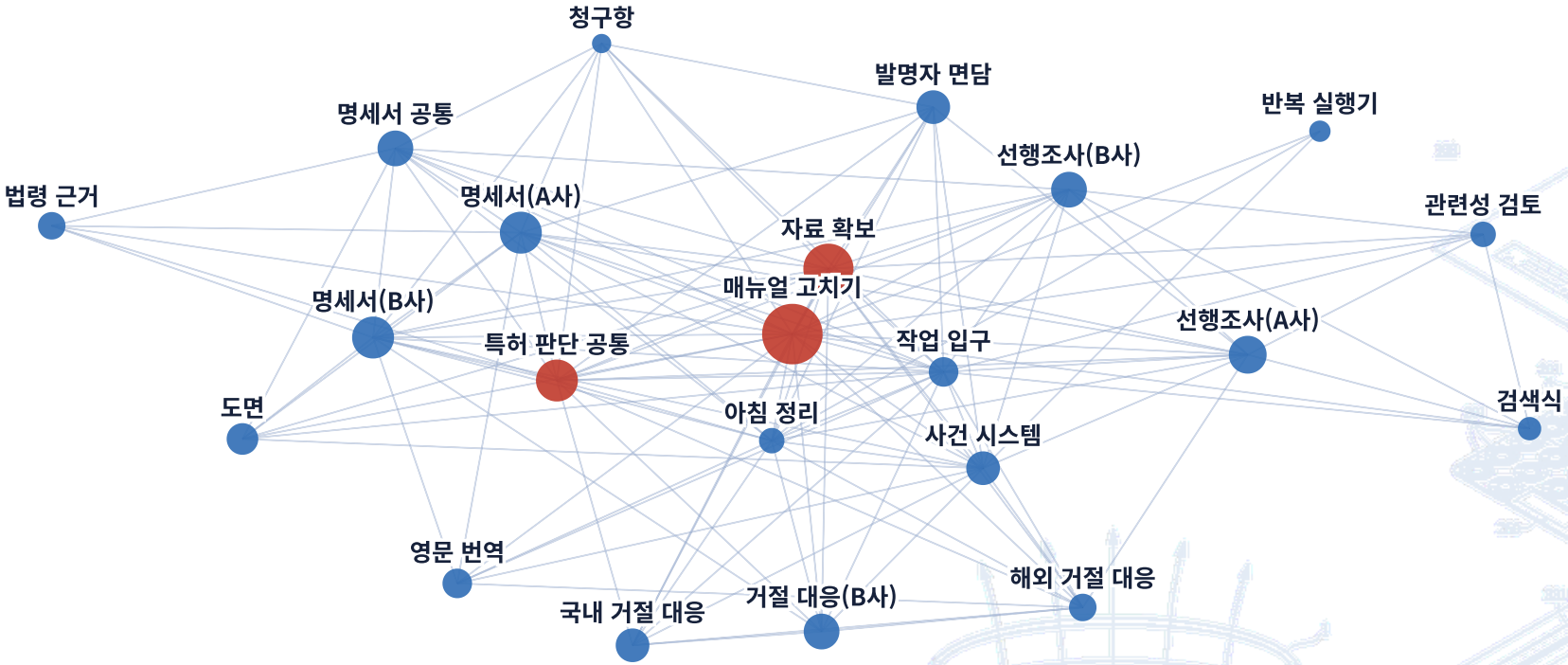
시험 한 번당 평균	처음 세 번	마지막 세 번
규칙으로 막을 수 있는 실수	**6.3**	**1.3**
└ 규칙이 없어서 틀린 곳	**3.3**	**0.7**
└ 규칙을 안 지킨 곳	**3.0**	**0.7**

실제 채점 대장 일부 (사건번호 지움)

시험	사용한 매뉴얼	막을 수 있는 실수
1	10월 23일 판	**7곳**
6	10월 26일 판	**4곳**
13	10월 27일 판	**3곳**
17	10월 30일 판	**0곳**

통계로도 확인했습니다. 시험 순서와 "막을 수 있는 실수" 수의 순위상관은 **−0.69**이고, 순서를 무작위로 2만 번 섞었을 때 이만큼 줄어드는 경우는 **0.1%**였습니다. 사람만 판단할 곳은 일부러 규칙으로 가르치지 않아 이 셈에서 뺐습니다.

22권이 177곳에서 서로를 가리키는 하나의 체계입니다



빨간 점이 가장 많이 참조되는 공통 매뉴얼입니다 — **매뉴얼 고치기 · 자료 확보 · 특허 판단 공통**. 법 조문 근거처럼 한 곳에서만 관리하고 나머지는 그곳을 가리킵니다.

매뉴얼 본문에서 다른 매뉴얼 이름을 부른 곳을 자동으로 센 지도입니다. 고객사 이름은 A사·B사로 바꿨습니다.

제가 자는 동안 시험이 돌고, 저는 멈춘 곳에서만 결정합니다

1

새벽 05:53

담당 사건의 기한을 읽어
오늘 할 일·우선순위 정리

2

평일 08:52 예약

눈 가린 시험 한 건
결과는 채점 대장에

3

수요일

매뉴얼끼리 어긋난 곳 검사
1주일 지적을 규칙 후보로

✓

사람의 몫

멈춘 곳에서 결정
보내기·제출은 사람만

680건 → 36건

**제 이름으로 붙은 업무 중
고객이 고쳐 달라고 돌려보낸 것**

2023년 11월부터, 회사 업무 시스템 기록

화상 면담에서 한 고객 담당자

"너무 잘 쓰셔서 얼굴 보고 싶었어요."

AI가 배우되, 판단과 책임은 사람이 집니다

모르면 멈춥니다

10월 22일 AI는 분량이 모자라 보여도 처음 자료에 없는 내용을 채우지 않고 멈췄습니다. 지어낸 내용은 고객의 권리를 흔듭니다.

AI가 이겨도 정답지는 바꾸지 않습니다

쓴 AI와 채점 AI를 나누고 기준은 숙련자의 완성본. 숙련자 쪽 오탈자가 보여도 AI 점수에 유리하게 쓰지 않았습니다.

사람의 판단은 AI에게 가르치지 않습니다

고객 사정을 보고 내린 숙련자의 선택은 규칙으로 만들지 않고 기록만. 맥락 없이 흉내 내면 다른 고객에게 잘못 쓰입니다.

보내는 버튼은 사람만 누릅니다

시험 문서는 채점에만 쓰고 고객에게 나가지 않습니다. 고객 문서는 사람 검토 뒤에만, 보내기·제출·삭제는 사람만.

고객 이름은 매뉴얼도 모릅니다

AI 계정은 대화가 모델 학습에 쓰이지 않게 설정. 매뉴얼에는 "어떻게"만, 고객·사건 번호·발명 내용은 없습니다.

답안지는 사무소 밖으로 나가지 않습니다

회사 시스템에서는 읽기만. 답안지로 쓰는 숙련자 문서는 업무 개선에만 쓰고, 이 응모 자료와 영상에도 없습니다.

AI에게 스스로 배우게 하는 방법이라, **AI가 사람의 판단과 책임까지 가져가지 않게** 막는 것이 핵심이었습니다.

규칙은 여섯 줄, 어느 AI에서나 따라 할 수 있습니다

어느 AI에서나 — 프로젝트 지침 칸에 붙여 넣을 여섯 줄

1. 일을 시작하기 전에 매뉴얼을 읽고, 이번 작업에 해당하는 줄을 먼저 밝힌다.
2. 내가 지적하면 "네가 한 말 / 바른 것 / 뿌리 원인 / 다음에 막는 법" 네 줄로 정리하고 매뉴얼에 붙일 문장을 써 준다.
3. 매뉴얼에는 내가 정한 것만. 네 짐작은 [확인 안 됨]에 두고, 근거 3건 전에는 기준처럼 쓰지 않는다.
4. 같은 실수가 다시 나오면 줄을 늘리지 말고, 있던 줄이 왜 안 걸렸는지 말한 뒤 그 줄을 고친다.
5. 잘한 사람의 완성본과 비교할 때는 "규칙 없음 / 규칙 안 지킴 / 이 건만의 사정"으로 나누고, 마지막은 규칙으로 만들지 않는다.
6. 되돌리기 어려운 결정은 멈추고 묻는다. 매뉴얼에는 "어떻게"만, 고객·사건 내용은 적지 않는다.

Claude에서는 사람 없이 돈다

스킬 — AI가 매뉴얼 파일을 직접 고친다

크롬 플러그인 — 답안지(숙련자 문서)를 가져온다. 읽기만

예약 작업 — 매일 아침 시험 한 건, 매주 점검

보조 AI — 쓴 AI 대신 채점한다

다른 AI에서는 네 줄을 사람이 매뉴얼에 옮겨 붙이면 된다

이 별첨과 신청서, 소개 영상은 **모두 Claude와 대화하며 만들었습니다**(영상의 실사풍 장면은 Google Flow의 Veo로 만든 AI 영상이고, 내레이션은 AI 음성(Google Gemini)입니다). 이 응모 자료 자체가 이 사례의 결과물입니다.