인사이트

AI가 채운 값은 왜 늘 실제보다 좋게 나올까요

방동걸 (주)웨이스 대표이사 2026년 8월 29일

저희는 공장에서 쓰는 소프트웨어와 **현황 진단 도구**를 만듭니다. 이 도구는 한 공장의 데이터·설비·공정 관리 수준을 항목별로 매기는데, 그 값을 담당자에게 일일이 묻지 않고 **회의 기록과 기존 자료로 자동으로 채웁니다.** 공장 운영체제 VEXPLOR를 만드는 웨이스가 그 도구의 채워진 값을 하나씩 검토하다 확인한 문제입니다.

만들어 둔 진단 결과 **네 건**을 다시 검증했습니다. 이번에는 기존 자료 대신 **실제 회의 녹취와 시스템 실측값**으로 같은 도구에 다시 넣었습니다.

판정이 나빠진 항목이 각각 **6개, 4개, 3개**였고 **좋아진 항목은 네 건을 통틀어 한 개도 없었습니다.** 나머지 한 건만 0개였는데, 그건 처음부터 시스템 실측값으로 채운 것이었습니다.

도구는 손대지 않았습니다. 바뀐 것은 **무엇을 근거로 채웠는가** 하나뿐이었습니다.

이 글은 그 검증에서 나온 판정 기준을 정리한 것입니다. AI에게 현황 조사·설문·진단서를 자동으로 채우게 하는 곳이라면 어디든 같은 자리에 서게 됩니다.

01

같은 도구에 근거만 바꿔 넣었더니 판정이 뒤집혔습니다

처음 채울 때는 손에 있는 문서로 채웠습니다. 우리가 쓴 제안서, 분석 보고서, 업계에서 통하는 표준값 같은 것들입니다. 나중에 실제 회의 기록과 시스템 실측값을 확보하고 나서 **같은 도구에 그것만 다시 넣어** 돌렸습니다.

처음 채운 근거	나빠진 항목	좋아진 항목
업계 표준값	6개	0개
업계 표준값	4개	0개
업계 표준값	3개	0개
시스템 실측	0개	0개

갈림이 근거의 출처를 정확히 따라갑니다. 표준값으로 채운 곳은 예외 없이 무너졌고, 실측으로 채운 곳만 그대로였습니다.

한 방향으로만 틀립니다. 좋아진 항목이 하나도 없다는 것이 이 실험에서 가장 중요한 숫자입니다. 오차가 양쪽으로 흩어지면 평균으로 보정할 수 있지만, 한쪽으로만 쏠리면 그건 오차가 아니라 편향입니다. 문서로 회사를 채우면 **그 회사는 실제보다 잘하는 회사가 됩니다.**

당연한 일이기도 합니다. 우리가 쓴 문서에는 잘 안 되는 이야기가 잘 안 적힙니다. 업계 표준값은 정의상 평균이거나 권장 수준입니다. 그것들을 모아 한 회사를 그리면 실제보다 나은 회사가 나옵니다.

02

틀린 것은 늘 입력이었습니다

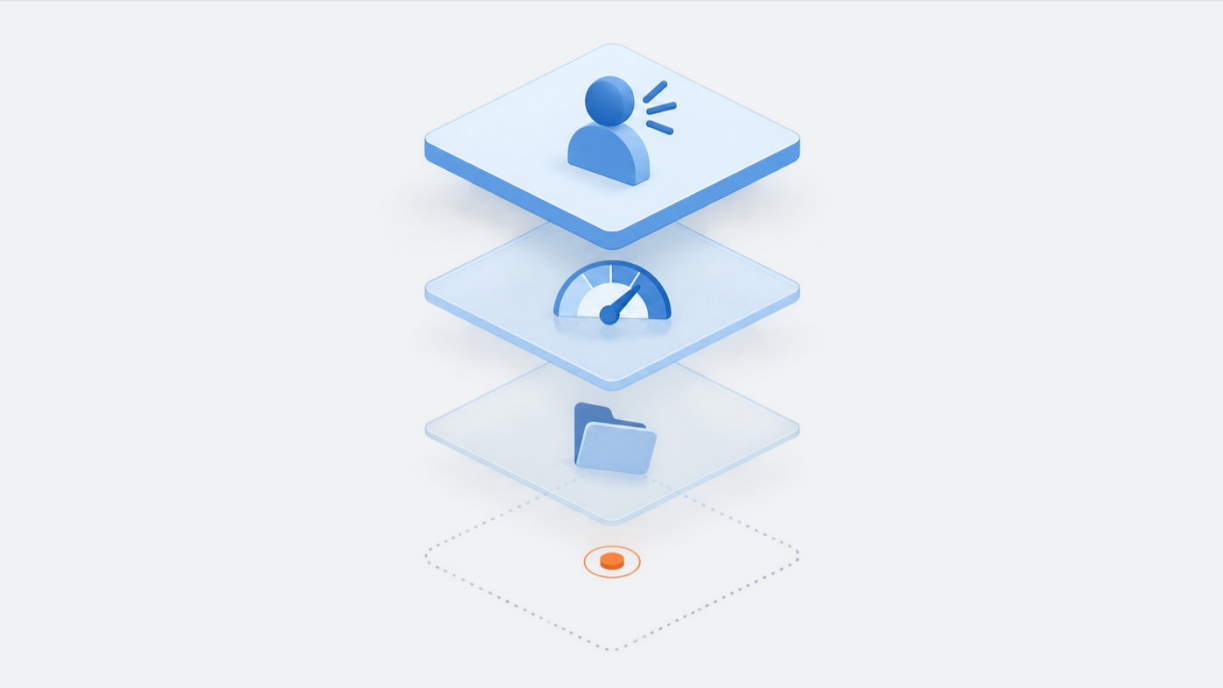
다시 돌리기 전에 한 가지를 먼저 확인했습니다. **처음 넣었던 답을 그대로 다시 넣으면 처음 결과가 그대로 나오는가.** 네 곳 전부 완전히 일치했습니다.

이 확인을 먼저 한 이유가 있습니다. 이걸 안 하면 나중에 나온 차이가 입력 때문인지 도구가 그사이 바뀌어서인지 가릴 수 없습니다. 재현이 확인된 뒤라야 **이후의 모든 차이는 입력에서 왔다**고 말할 수 있습니다.

결과가 이상하면 사람은 대개 도구를 먼저 의심합니다. 로직을 뜯어보고 계산식을 고칩니다. 그런데 실제로는 도구가 멀쩡한 채로 **틀린 값을 정확하게 계산하고 있는** 경우가 훨씬 많습니다. 재현 확인은 그 순서를 바로잡습니다.

03

근거에 등급을 매기지 않으면 무엇이 섞였는지 모릅니다



근거의 종류마다 강도가 다릅니다. 사람이 직접 말한 것이 가장 강하고, 이름에서 뜻을 짐작한 것이 가장 약합니다.

실험 뒤에 채움마다 등급을 붙이도록 규격을 바꿨습니다. 문서 단위가 아니라 **채움 단위**입니다. 같은 문서라도 고객이 말한 대목은 A, 우리가 해석한 대목은 C입니다.

등급	무엇인가	결과
A	상대가 직접 답했거나 실데이터	판정 불변
B	상대 발언이 담긴 녹취	편향 정정 성공
C	우리가 쓴 문서	같은 편향원
D	표 한 행 + 표준값 보간	예외 없이 무너짐

C가 위험한 이유는 틀려서가 아니라 같은 곳에서 나왔기 때문입니다. 우리가 쓴 문서로 상대 현황을 채우면 대조가 아니라 자기 복사입니다. 실제로 **우리 쪽 목표 수치가 상대의 현재값 자리로 들어간 사례**가 나왔습니다. 목표로 적어 둔 숫자가 현황으로 둔갑한 것입니다.

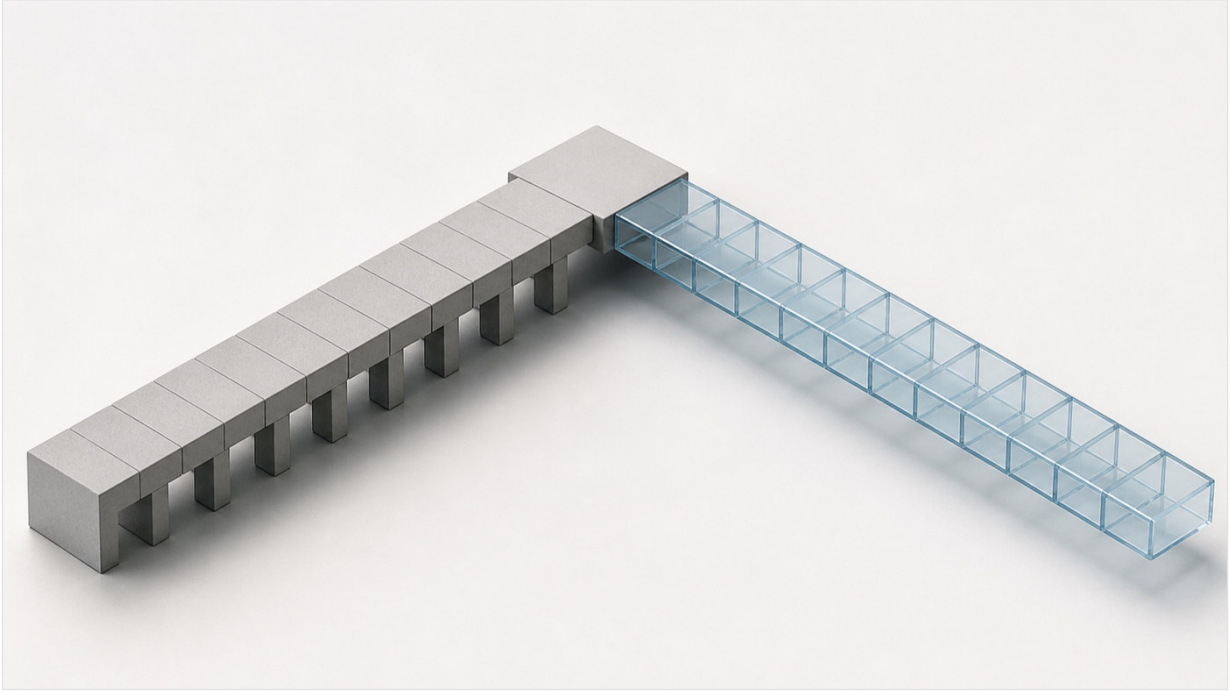
D는 규모부터 어긋납니다. 레퍼런스 표의 한 행에서 **97개 항목**이 나온 경우가 있었습니다. 한 줄로 97개를 채웠다는 것은 96개가 보간이라는 뜻이고, 그 세트는 검증 자체가 불가능했습니다.

그래서 두 가지를 걸었습니다.

- **C·D만으로 채운 항목은 산출물에 [미검증] 표시를 강제합니다.** 어느 문서에 실리든 표시가 따라갑니다.
- **판정과 금액을 가르는 항목은 A·B 없이는 채우지 않습니다.** 어느 항목이 그런지는 위 실험이 알려줬습니다.

04

결론이 같다고 안전한 것이 아닙니다



같은 결론에 닿아도 받쳐 주는 근거가 있는 경로와 없는 경로는 다릅니다.

가장 잘못 읽기 쉬운 지점입니다. 다시 돌렸는데 최종 결론이 같으면 안심하게 되는데, 실험에서는 **결론이 보존된 곳일수록 다른 데서 샜습니다.**

겉보기	실제로 어긋난 곳
성숙해 보이던 곳	판정이 반대로 뒤집힘
초기 단계인 곳	결론 유지, 금액 9.2% 이동
중간 단계인 곳	규모가 4배
결론·금액 모두 불변	서술의 주어가 틀림

두 번째 줄이 특히 무섭습니다. **항목 하나가 금액을 9.2% 움직였습니다.** 사업장이 몇 곳인지를 묻는 항목이었고, 문서로 채울 때는 본사 기준으로 한 곳이 들어가 있었습니다. 실제로는 여러 곳이었습니다.

점수는 그대로였고 결론도 그대로였는데 금액만 조용히 달라졌습니다. **결론을 검사하는 눈으로는 절대 안 보이는 자리입니다.**

마지막 줄이 제일 까다롭습니다. **주어가 틀린 경우**입니다. "43개 화면 중 29개 구현"이라고 적힌 값이 실은 상대 시스템이 아니라 **우리가 만든 쪽의 상태**였습니다. 숫자는 정확하고 점수도 1도 움직이지 않는데, 상대에게 보여 줄 문서의 사실관계가 틀립니다.

점수를 검사하는 게이트는 이걸 못 잡습니다. 점수가 안 변하니까요. 그래서 검사 항목을 하나 더 넣었습니다 — **이 서술의 주어가 상대인가.**

05

편향은 결론만이 아니라 할 일의 범위를 바꿉니다

점수와 금액까지 봤으면 됐다고 생각했는데, 산출물을 나란히 놓고 보니 한 겹이 더 있었습니다.

같은 도구로 요구사항 목록을 다시 뽑아 원본과 대조했습니다. 어떤 곳은 **원본에서 "있으면 좋음"이던 항목이 재작성본에서 "반드시 필요"로 뒤집혔습니다.** 예를 들어 이력을 거슬러 조회하는 기능이 원본에는 "활용 중"으로 적혀 있었는데, 실제 녹취에서는 "안 됩니다"였습니다. 같은 항목이 선택 사항에서 필수 사항으로 자리를 옮깁니다.

더 큰 것은 **원본에 아예 없던 영역이 통째로 생겼다는 점**입니다. 재고와 경영 분석 쪽 요구가 원본 목록에는 한 줄도 없었는데 재작성본에서 새로 나타났습니다. 문서로 채울 때 그 영역이 "양호"로 들어가 있었기 때문에 할 일 목록에서 통째로 빠져 있었던 것입니다.

이게 실무에서 어떻게 나타나느냐면 — 견적과 일정이 그 목록을 기준으로 잡히므로, 빠진 영역은 계약 뒤에 발견됩니다. 그때는 범위 밖이라 추가 비용 이야기가 되고, 상대는 "처음부터 필요하다고 말했다"고 하고 우리는 "목록에 없었다"고 합니다. **양쪽 다 맞습니다.** 목록을 우리가 우리 문서로 만들었기 때문입니다.

06

녹취를 그대로 믿으면 안 되는 이유 여섯 가지

녹취는 B등급이라 문서보다 낫습니다. 그렇다고 그대로 넣으면 안 됩니다. 실험에서 밟은 사고를 그대로 검사 항목으로 만들었습니다.

검사	무엇을 막는가
시각 표기 없으면 반려	근거 위치 인용 불가
화자 번호를 근거로 금지	재실행마다 배정이 바뀜

검사	무엇을 막는가
숫자는 사람 확인 후	받아쓰기가 숫자를 깨뜨림
다른 회의 혼입 검사	문장 단위 교차 오염
길이 절단 검사	끝이 잘린 줄 모름
회의 종류 분류	안 다룬 주제를 공백으로

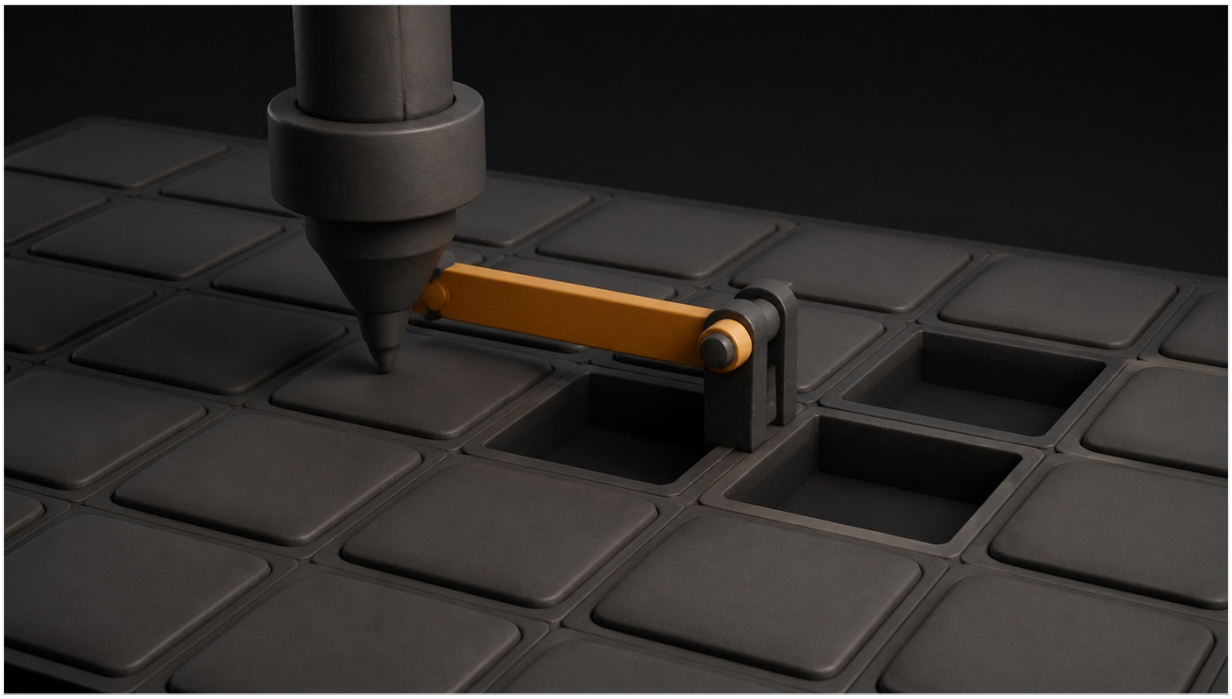
두 번째가 특히 의외였습니다. **같은 녹음을 다시 받아쓰게 했더니 화자 수가 6명에서 9명으로 바뀌었고, 같은 문장의 화자 배정도 달라졌습니다.** "3번 화자가 말했으니 상대편이다" 같은 판단은 성립하지 않습니다. 근거는 시각과 원문 문자열로만 잡아야 합니다.

다섯 번째도 조용한 사고입니다. 어떤 녹취가 **50,002자에서 끝나 있었습니다.** 딱 떨어지는 상한 근처에서 끝난다면 회의가 거기서 끝난 게 아니라 파일이 잘린 것입니다. 잘린 뒷부분의 내용은 "언급 없음"으로 처리되고, 그게 곧 "그 회사에는 그게 없다"가 됩니다.

마지막 줄도 그냥 넣은 게 아닙니다. **어떤 회의의 녹취냐에 따라 자동으로 채워지는 비율이 58.5%에서 87%까지 갈렸습니다.** 30% 가까운 차이를 만든 것은 도구가 아니라 그날 무슨 회의였는가입니다. 착수 보고 자리에서는 현황이 넓게 나오고, 특정 안건만 다룬 회의에서는 그 영역만 채워집니다. **그래서 회의 종류를 먼저 분류하고, 그 회의가 다루지 않은 영역을 공백으로 명시**합니다. 분류하지 않으면 87%와 58.5%가 같은 얼굴로 산출물에 실립니다.

07

그래서 채우지 않고 비웁니다



근거가 기준에 못 미치면 채우는 동작 자체를 막습니다. 빈칸은 아직 확인하지 않았다는 표시로 남습니다.

규격의 기본 태도는 하나로 정리됩니다. **통과 근거가 없으면 채우지 않습니다.**

빈칸을 남기면 산출물이 허술해 보이니 무언가로 메우고 싶어집니다. 그런데 **"녹취에 없다"는 "그 회사에 없다"가 아닙니다.** 그날 그 자리에서 그 이야기가 안 나왔을 뿐입니다. 그 둘을 같게 취급하는 순간, 안 물어본 것이 없는 것으로 확정됩니다.

그래서 못 채운 항목은 추측으로 메우지 않고 **미측정으로 표시해 확인 목록으로 보냅니다.** 실제 재실행에서 미측정 6건과 미검증 22건이 이 원칙으로 처리됐고, 임의값을 넣은 건은 한 건도 없었습니다.

빈칸은 결함이 아니라 정보입니다. **어디를 아직 안 물어봤는지**를 알려 주니까요. 그럴듯하게 채운 값은 그 정보를 지웁니다.

08

정리 — 다섯 줄로 재 보기

자동으로 채우는 구조를 이미 쓰고 있다면 아래 다섯 개를 자기 것에 대 보시면 됩니다.

- **한 방향으로만 틀리는지 봤습니까** — 좋아진 항목과 나빠진 항목의 수가 비슷하지 않다면 오차가 아니라 편향입니다

- **처음 입력을 다시 넣으면 처음 결과가 나옵니까** — 이걸 확인하기 전에는 도구를 의심할 근거가 없습니다
- **채움마다 근거 등급이 붙어 있습니까** — 문서 단위가 아니라 채움 단위여야 합니다
- **우리가 쓴 문서로 상대 현황을 채우고 있지 않습니까** — 대조가 아니라 자기 복사입니다
- **못 채운 칸을 비워 둘 수 있습니까** — 비울 수 없는 구조라면 어딘가는 반드시 지어냅니다

검증이 시스템 밖의 증거로 이뤄져야 한다는 이야기는 전에 한 번 쓴 적이 있습니다. 이 글은 그 다음 이야기입니다 — **밖에서 가져온 증거에도 등급이 있고, 우리가 쓴 문서는 그중 낮은 쪽입니다.**

여러분 조직에서 AI가 채워 준 값에는 지금 어떤 근거가 붙어 있습니까?

WACE(웨이스)는 공장이 스스로 판단하고 멈추지 않게 하는 제조 AI 운영체제 VEXPLOR를 만드는 회사다.

원문 blog.vexplor.com/post/why-auto-filled-values-look-too-good
발행 2026년 8월 29일 · 방동걸 (주)웨이스 대표이사

회사 주식회사 웨이스 (WACE Inc.) · 대표이사 방동걸 · 사업자등록번호 439-87-02852
주소 (18462) 경기 화성시 동탄구 동탄영천로 150, 에이동(A동) 1702-1708호 (영천동, 실리콘앨리동탄)
연락처 전화 070-4730-0629 · contact@wace.me
홈페이지 wace.me · linkedin.com/company/wace-inc

원문 홈페이지 링크드인