원어민 감수 없이 두 회사 AI 가 서로 검사 — 판정한 145줄 중 67줄은 틀린 지적(판정 AI 기준)

1인 소프트웨어 사업자(회사원 겸업). 화면 문구 1,016줄(한국어 원문)을 영어·일본어·중국어로 옮긴 번역(초안도 AI)을, 서로 다른 회사 AI 두 곳이 따로 검사하고 판정 AI 가 줄마다 맞음·틀림을 가렸다. 나는 일본어·중국어를 읽지 못한다. **원어민 검수는 하지 않았다.**

흐름 (2026-10-01 저녁 ~ 10-02 오전)

① 두 곳에 따로 검사
OpenAI GPT-5.4 · Anthropic Claude Opus 5.5, 같은 지시문, 3개 언어 × 2곳 = 6회

→

② 판정 대상 145줄
두 곳이 같은 줄을 지적했거나 한 곳이라도 '심각' (지적 합계 251건)

→

③ 줄마다 판정표
맞음 / 일부 맞음 / 틀림 + 이유 한 줄. 78칸 고침

→

④ 지시문 고침
틀린 지적 4유형 → 용어집·언어 방침 추가

→

⑤ 2차 481쌍
두 곳 모두 지적한 31줄만 판정 → 26칸 고침

잰 숫자 (판정표 두 개에서 스크립트로 셈 · 맞음/틀림은 판정 AI 기준 · 줄 = 문구 하나의 한 언어 번역)

항목	1차 (10/1)	2차 (10/2, 규칙 적용 뒤)
판정한 줄	145	31 (481쌍 중 두 곳 모두 지적)
맞음 · 일부 맞음 · 틀림	24 · 54 · **67 (46%)**	10 · 16 · 5
고친 칸	78 (54%)	**26 (84%)**
두 곳 모두 지적한 줄 중 고칠 거리	106줄 중 74 (70%)	31줄 중 26 (84%)
한 곳만 지적한 줄 중 틀린 지적	**39줄 중 35 (90%)**	229줄 — 세기만, 열어 보지 않음
검사 AI 호출 · 비용	6회 · 금액 기록 없음 (오늘 도구로 다시 계산 약 12,415원 [추정] — 1차 때 없던 지시문 포함이라 조금 클 수 있음)	6회 · 약 1,987원 (도구가 글자 수로 어림한 예상치, 청구서 대조 안 함)

사람에게 맡겼다면 — 공개 요금표로 같은 분량 계산 [추정]

분량(1차 직전 번역 파일에서 셈): 일본어 9,397자 · 중국어 6,774자 · 영어 3,723단어 · 한국어 원문 8,735자(공백 포함).

기준	계산	금액 (부가세 별도)
A. 비코스코리아 '원어민 감수' 단가 1페이지 = 영어 200단어·일본어 400자·중국어 300자, 10,000원/페이지, 1페이지 미만도 1페이지	일 24쪽 + 중 23쪽 + 영 19쪽 = 66쪽 × 10,000원	**660,000원**
B. 글로벌번역원 '감수만' = 기본료 60% 한국어→외국어 27,000원/페이지(한글 600자 기준)	8,735자 ÷ 600 = 14.56쪽 × 16,200원 × 3개 언어	**약 707,500원**
우리 쪽 예상 비용 (검사 AI 호출만, 청구서 대조 안 함)	1차 약 12,415원[추정] + 2차 약 1,987원. 판정·수정은 정액 구독(Claude Code) 안에서	**약 14,400원**

같은 일이 아니다. 사람 감수는 원어민이 문장을 읽고 고친다. 우리 쪽은 AI 가 지적하고 AI 가 판정했다. 이 표는 "사람에게 맡겼다면 이만큼"을 보여줄 뿐, 품질이 같다는 뜻이 아니다. 납기도 A 는 "원칙적으로 3일 이상", 우리 1차는 같은 날 저녁 안에 판정·수정까지 끝남(진행 기록 21:10, 검사 실행 시각은 따로 기록 없음). 요금표 열람 2026-10-19, 출처 주소는 4쪽.

AI 지적은 얼마나 맞았나

1차 언어별 (판정 대상 145줄)

언어	대상	맞음	일부 맞음	틀림	고친 칸
일본어	42	5	18	19	23
중국어	49	13	19	17	32
영어	54	6	17	31	23
합계	**145**	**24**	**54**	**67**	**78**

두 회사 중 누가 더 맞았나 (판정 AI 기준)

	1차: 지적 중 맞음+일부	2차(31줄): 그 회사 지적이 맞음+일부*	특징
OpenAI GPT-5.4	144건 중 75 (52%)	31 중 23 (74%)	많이 지적. 혼자 '심각'으로 올린 38줄 중 34줄이 틀림
Anthropic Claude Opus 5.5	107건 중 74 (69%)	31 중 25 (81%)	적게 지적, 이유가 구체적

* 2차 회사별 칸은 같은 31줄에서 두 회사의 지적(제안 문장)을 따로 채점한 값이라, 줄 판정(맞음 10 · 일부 16 · 틀림 5)과 단위가 다르다.
판정 AI(Claude Code)가 Anthropic 제품이라 이 회사 비교는 편향일 수 있다. 그래서 한 회사만 믿는 대신 "두 회사가 모두 지적한 줄"을 규칙으로 했다.

틀린 지적 4유형 (실제 예)

유형	실제 예
㉠ 내 제품의 용어를 모름	디스크 '저장 공간' = ストレージ 를 두 곳 모두 방 이름으로 오해해 '수납'(収納)·'창고 방'으로 바꾸라고 함 (한 곳은 '심각')
㉡ 일부러 넣은 말장난·별명을 오류로 봄	한국어 원문의 장난스러운 별명·뭉뚱그린 제목을 "원문과 뜻이 다름"으로 지적
㉢ 언어별 방침을 모름	일본어 화면 안내문은 정중체(ます)가 방침인데 "딱딱하다"고 지적 / 자리표시자 변형({이름:lower})을 오류로 봄
㉣ 자연스러운 의역을 직역 기준으로 깎음	물건을 건네며 하는 말 "here you go / 给你 / はい、どうぞ" 를 직역이 아니라고 지적

(그 밖에 1차에 남은 길이 지적 2건도 틀림. 길이는 검사 프로그램이 따로 재므로 원래 판정 대상에서 뺐다.)

2차 언어별 (두 곳 모두 지적한 31줄)

언어	검사한 쌍	OpenAI 지적	Anthropic 지적	둘 다	맞음·일부·틀림	고친 칸
일본어	131	46	20	7	3 · 4 · 0	7
중국어	179	116	28	17	6 · 10 · 1	16
영어	171	62	19	7	1 · 2 · 4	3
합계	**481**	**224**	**67**	**31**	**10 · 16 · 5**	**26**

영어는 2차에서도 두 곳이 같이 의역을 깎았다(7줄 중 4줄 틀림) → 영어는 AI 지적을 더 보수적으로 본다.

규칙 3줄 · 검사 지시문 · 판정표

규칙 3줄

1. 검사 지시문에 **용어집과 언어별 방침**을 넣는다. (우리는 메뉴·제목 같은 고유 용어 32줄을 자동으로 붙임)
2. **두 회사가 모두 지적한 줄만 판정 대상으로 삼고, 그중 맞는 것만 고친다.** (2차 31줄 중 5줄은 기각) 한 곳만 지적한 줄은 세기만 한다.
3. **틀린 지적은 근거 한 줄을 적고 번역을 유지**한다. 같은 유형이 반복되면 지시문에 한 줄 더한다.

검사 지시문 (두 AI 에 똑같이 붙여 넣기 — 우리 지시문을 일반화한 것)

```
너는 [언어] 번역 검수자다. 아래 표는 한국어 원문, 번역, 화면 설명, 글자 수 제한이다.
줄마다 문제가 있으면 "줄 번호 | 문제 | 고칠 문장 | 심각도(높음/중간/낮음)"로 적고, 없으면 적지 마.

[언어 방침] 예) 일본어: 화면 버튼·메뉴는 짧은 명사형, 긴 안내문은 です·ます. 상대를 부르는 말은 ○○로 통일.
              영어: 캐릭터 대사는 소문자로 시작해도 된다.
[용어집] 저장 공간 = ストレージ / 存储空间 (디스크 용량 뜻) · 메뉴 이름 ○○ = ○○ ...
[자리표시자] {이름}, {n} 같은 칸과 {이름:lower} 같은 변형은 정상이다. 지우거나 바꾸지 마.
[주의] 원문의 말장난·별명은 일부러 한 것이다. 자연스러운 의역을 직역이 아니라는 이유로 깎지 마.
```

판정표 양식 (두 결과를 나란히 놓고, 둘 다 지적한 줄만 옮김)

줄	언어	누가 지적(심각도)	판정 맞음/일부/틀림	이유 한 줄 (틀림이어도 적음)	번역 전 → 후 / (유지)
12	ja	A(높음) · B(낮음)	틀림	디스크 저장 공간 뜻이라 ストレージ 가 맞음	ストレージ (유지)

비용 미리 재기

우리는 검사 스크립트가 보내기 전에 "글자 수 → 토큰 수 → 공개 가격표 × 환율 1,400원[가정]"으로 예상 금액을 먼저 보여 주게 했다. 2차 481쌍은 OpenAI 827원 + Anthropic 1,160원 = 약 1,987원(예상치)이었다.

내가 한 일 / AI 가 한 일

나	AI
AI 가 제안한 '다른 회사 AI 함께 쓰기' 안을 채택하고 결제 키를 넣음 · 2차를 다음 날 오전으로 앞당김 · 원어민 검수(일본어·중국어 약 13만 원 예산안)를 하지 않기로 결정(10/12) · '사람이 검토함' 표기 금지 유지	검사(두 회사 API) · 검사 스크립트 작성 · 판정표 작성(줄마다 이유) · 번역 파일 수정 · 고친 뒤 자동 시험 실행 · 틀린 지적 유형 정리와 지시문 개선

해 본 경로: Claude Code 가 스크립트로 두 API 를 부르고 판정표를 만듦. 웹 채팅에 직접 붙여 넣는 경로는 해 보지 않았다.

한계와 출처

한계 (숨기지 않음)

1	**원어민 검수를 하지 않았다.** 일본어·중국어 원어민 검수 예산안(약 13만 원)을 10월 12일 하지 않기로 했다. 그래서 '사람이 검토함'이라고 쓰지 않기로 했고, 판매 페이지에 넣을 AI 사용 공개 문구(아직 공개 전)에 "글 작성과 번역에 AI 를 썼다"고 적어 두었다.
2	**판정도 AI 가 했다.** 그리고 판정 AI 는 검사 AI 한 곳과 같은 회사(Anthropic) 제품이다. "맞음 · 틀림" 숫자는 판정 AI 의 판단이다.
3	**한 곳만 지적한 줄은 보지 않았다.** 2차에서 한 곳만 지적한 229줄(OpenAI 193 · Anthropic 36)은 규칙대로 고치지도, 열어 보지도 않았다. 그 안에 진짜 오류가 있을 수 있다.
4	**1차 비용은 기록이 없다.** 12,415원은 오늘 같은 도구로 다시 계산한 추정이다(1차 때 없던 언어 방침·자리표시자 지시문이 포함돼 실제보다 조금 클 수 있다). 2차 1,987원도 도구 예상치이며 청구서와 대조하지 않았다.
5	문화적 뉘앙스·유행어·실제 화면 어울림은 AI 검사로 못 본다(검사 결과 파일 머리말에도 적혀 있다).

근거 기록 (모두 작업 저장소의 실제 파일)

기록	내용
검사 결과 파일 6개 (10/1)	두 회사 × 3개 언어, 1,016줄씩. 그림 2 가 이 파일 머리말과 줄을 그대로 옮긴 것
1차 판정표 (10/1)	145줄 전부: 언어·판정·누가 지적·이유·전→후. 그림 1·3 의 숫자 원본
2차 판정표 (10/2)	481쌍 집계 + 두 곳 모두 지적한 31줄 판정, 비용 예상치
진행 기록 (10/1 20:20 · 21:10 · 10/2 10:55)	검수 도구 완성 → 1차 판정 145줄·78칸 → 2차 481쌍·6회·약 1,987원·26칸
결정 기록 (10/12)	원어민 검수 하지 않음, '사람이 검토함' 표기 금지

사람 감수 단가 출처 (열람 2026-10-19, 부가세 별도)

비코스코리아 요금표	원어민 감수 10,000원/1페이지(영어 200단어·일본어 400자·중국어 300자), 납기 원칙 3일 이상 http://www.b-cause.kr/price/index.shtml
글로벌번역원 번역비용	한국어→외국어 27,000원/페이지(한글 600자), 감수만 원하면 기본료의 60% http://www.glo.co.kr/mobile/page?sub=sub02_03
(참고) 한국국제회의통번역센터 번역요율	원어민 감수는 번역과 함께만, 번역료의 최저 50% 이상 추가(감수만 별도 접수 안 함) — 계산에는 쓰지 않음 http://kcbctrans.co.kr/sub/sub04_02.php

AI 활용 범위: 이 증거 문서와 그림은 AI(Claude Code)가 기록 파일에서 숫자를 스크립트로 다시 세어 만들었고, 신청자가 읽고 확인했다. 그림의 회색 '가림'은 제품 이름이 드러나는 줄 이름·장면 설명을 가린 곳이다.