

AI 영어자료 비교 기록

영어 시험 대비 자료를 ChatGPT로 만들어 보니 겉으로는 꽤 자연스러웠다. 그런데 처음 배우는 내용은 AI가 틀려도 내가 바로 알아채기 어렵다는 생각이 들어, 기존 학습자료를 나중에 열어 직접 비교해 봤다.

비교할 때 한 일

- 기준자료는 먼저 보여주지 않았다.**
영어 원문과 분석 형식만 주고 AI 자료를 먼저 만들었다.
- AI 결과를 먼저 저장했다.**
PDF와 비교용 기록을 남겨, 기준자료를 본 뒤 결과를 고쳐 맞추지 않도록 했다.
- 그다음 기존 학습자료와 대조했다.**
AI가 잡은 내용뿐 아니라 빠뜨린 내용도 같이 확인했다.
- 틀림만 보지 않았다.**
일치, 비슷한 기능, 추가 분석, 잘못된 분석, 누락으로 나눠 기록했다.

왜 이렇게 했나

기준자료를 먼저 보여주면 AI가 실제로 어떤 내용을 스스로 골랐는지 보기 어려웠다. 그래서 결과를 먼저 만들어 둔 뒤 비교하는 순서를 택했다.

남겨 둔 파일

BLIND_A_DECISION_LOG.md
BLIND_A_FROZEN.pdf
BLIND_B_INDEPENDENT_AUDIT.md
FREEZE_MANIFEST_C.md

제3자 학습자료의 원문이나 이미지는 이 첨부자료에 실지 않았다. 여기에는 내가 만든 AI 결과를 비교하면서 남긴 절차와 수치만 정리했다.

비교해 보니 나온 수치

A와 B는 서로 다른 지문을 썼고 세는 단위도 달랐다. 그래서 두 수치를 붙여서 “좋아졌다”거나 “나빠졌다”고 보지는 않았다. 어떤 종류의 문제가 실제로 있었는지 확인하는 데만 사용했다.

항목	BLIND A	BLIND B
기준자료와 대응한 비율(정밀도)	66.04%	62.20%
기준자료 내용 중 AI가 잡아낸 비율(회수율)	41.25%	38.89%
기준자료에 대응되지 않은 추가 분석	28.30%	36.59%
잘못된 분석	5.66%	1.22%
추가로 확인한 누락	중요 누락 11개	수업 완성총 27개 중 25개 미회수

내가 여기서 확인한 것

틀린 설명만 찾으면 충분하지 않았다. 잘못된 분석의 비율 자체는 크지 않아 보여도, 기준자료에 있던 내용을 놓치는 경우가 많았다.

설명이 많다고 해서 바로 공부하기 좋은 자료가 되는 것도 아니었다. 기준자료가 다루지 않은 내용을 AI가 더 붙이는 경우도 꽤 있었다.

AI 자료만 보면 누락은 눈에 잘 안 띈다. 그래서 나중에는 “이 설명이 맞나?”와 함께 “기존 자료에서 다뤘던 내용이 빠지지 않았나?”도 같이 보게 됐다.

이 수치로 말할 수 없는 것
A와 B만으로 영어 성적이 올랐다고 말할 수 없고, A→B에서 AI 자체 성능이 좋아졌다고도 볼 수 없다. 여기서는 내가 만든 공부자료에서 어떤 문제가 실제로 나왔는지만 기록했다.

실제로 걸린 세 가지

BLIND B · B029.a

1. 맞는 표현을 X로 표시한 경우

AI 표시: ignore (X)

다시 확인한 결과: be prone to ignore는 문법적으로 가능한 표현

그대로 외웠다면 맞는 표현을 틀린 것으로 공부할 뻔했다.

BLIND A

2. 기존 자료에는 없던 분석을 더 붙인 경우

예: various = different

이 설명 자체가 항상 틀렸다는 뜻은 아니다. 다만 시험 대비용 기준자료가 다루지 않은 내용을 시가 새로 덧붙였다는 점을 따로 기록했다.

BLIND B

3. 그럴듯한데 필요한 내용이 빠진 경우

본문 영어 재진술: 12개 중 10개 미회수

수업 완성층 필기: 27개 중 25개 미회수

AI 자료만 봤다면 무엇이 빠졌는지 자체를 알아채기 어려운 부분이었다.

지금 공부할 때는

매번 이런 식으로 비율까지 계산하지는 않는다. 시로 자료를 만든 뒤 기존 학습자료나 믿을 만한 해설과 비교하고, 틀린 설명이 있는지와 빠진 내용이 있는지를 먼저 본다. 애매한 내용은 바로 외우지 않고 다시 확인한다.

수치와 사례는 BLIND A/B 비교 기록에서 확인한 내용만 사용했다. 제3자 학습자료의 원문은 재수록하지 않았다.