

사내 **plagia_result** 데이터셋 평가 리포트 (v2.1.0-kosimcse)

- 데이터셋: 표절 페어 500 건 + 비표절 페어 499 건 (총 999 쌍)
- 엔진 버전: o2o-plagiarism-2.1.0-kosimcse
- 운영 가중치: text(KoSimCSE) 0.3 / lemma 0.45 / char 0.15 / motif 0.1

1. 점수 컴포넌트 분포

점수	POS 평균	NEG 평균	분리도	std(POS/NEG)
메타 임베딩 (OpenAI, baseline)	0.8632	0.6665	+0.1967	0.085 / 0.128
KoSimCSE 본문 (자체)	0.9661	0.7346	+0.2315	0.023 / 0.138
Lemma 교집합	0.7807	0.2844	+0.4964	0.100 / 0.179

→ 그래프: [reports/01_score_distributions.png](#)

2. 모델별 최적 성능 (F1 최대화 **threshold**)

모델	Precision	Recall	F1	Threshold
기존 result.json (전임자 1 단계)	0.9520	0.9560	0.9540	0.78
메타 임베딩 단독 (OpenAI)	0.7842	0.8720	0.8258	0.76
KoSimCSE 단독 (자체)	0.9231	0.9600	0.9412	0.91
Lemma 단독	0.9391	0.9560	0.9475	0.59
하이브리드 (운영 가중치)	0.9257	0.9720	0.9483	0.57

→ 그래프: [reports/02_threshold_curves.png](#), [reports/03_model_comparison.png](#)

3. Confusion Matrix (하이브리드, threshold=0.57)

	예측: 표절	예측: 비표절
실제: 표절	TP = 486	FN = 14
실제: 비표절	FP = 39	TN = 460

4. 결론

1. **KoSimCSE** 도입으로 자체 산출물 정합성 확보 — OpenAI 의존 0, 호출 비용 0, 데이터 외부 노출 0
2. **Lemma** 컴포넌트가 단독으로도 강력 — F1 0.9475 (자서전 도메인의 어미 변경 표절을 결정적으로 잡음)
3. 하이브리드가 가장 안정 — recall 0.9720 (실제 표절을 거의 다 잡음)
4. **PDF v1.2** 권장 임계값 **0.85**와의 관계 — 본 평가는 plagia_result 데이터 (출판 콘텐츠) 기준 F1 최적치이며, 실제 자서전 도메인에서는 PDF 권장 0.85 적용을 우선 (정밀도 우선, 재현율 일부 손실 감수)