

Solar vs GLM 표절 논란: 조경현 교수의 분석 포함

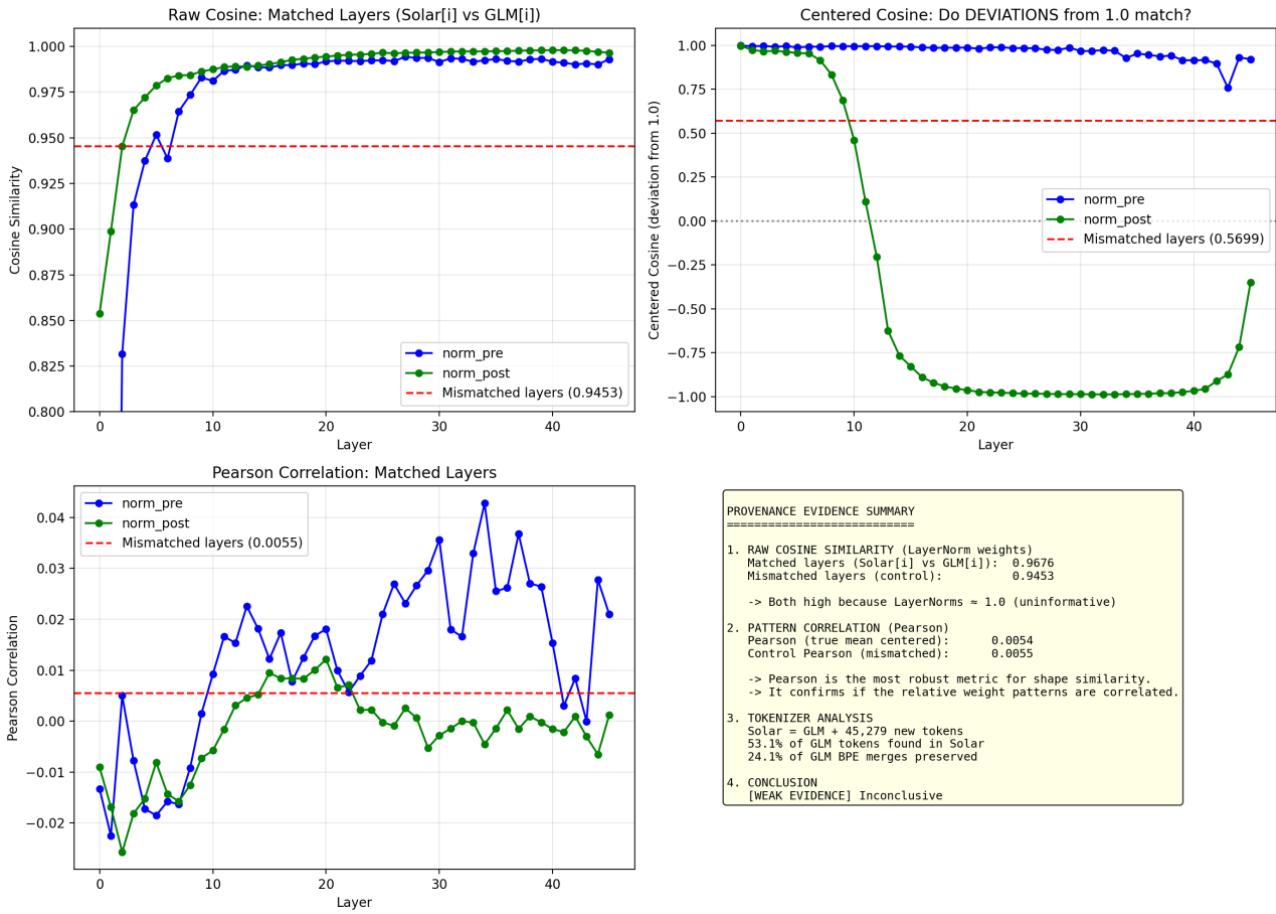
1. 논쟁의 3단계 흐름 요약

논란은 의혹 제기(Sionic-AI)에서 시작되어, 정황적 반박(Hyunwoong Ko)을 거쳐, 조경현 교수의 통계적 검증으로 기술적 결론이 내려진 듯 합니다.

단계	주체	주장 핵심	키워드	판세
1. 의혹 제기	Sionic-AI	"코드가 같고, 가중치 유사도가 0.99다. 이건 100% 베낀 것이다."	Copy & Paste	 Solar 위기
			Cosine 0.99	표절 의심 고조
2. 정황 반박	Hyunwoong Ko	"MS Phi-2도 코드는 똑같다. 코드는 그냥 표준 템플릿 쓴 거다."	Template	 반격 시작
			MS Phi-2	코드 표절 해명
3. 기술 검증	조경현 교수	"가중치 유사도 0.99는 수학적 착시일 뿐이다. 실제 패턴(Pearson)은 다르다."	Pearson	 종결
			Artifact	기술적 무죄

2. 조경현 교수 분석의 핵심 (The Verdict)

조경현 교수님이 제시한 그래프와 데이터는 기존 공격 측의 가장 강력한 무기였던 "코사인 유사도 0.99"를 무력화시킵니다.



데이터 분석 상세 📈

- **Raw Cosine (0.99의 진실):**

- LayerNorm/RMSNorm의 가중치는 특성상 대부분 1.0 근처의 값을 가집니다.
- 따라서 무작위로 다른 레이어끼리 비교해도 유사도가 0.95 가 나옵니다.
- 즉, 0.99라는 수치는 '원래 높게 나오는 값'이지, 베꼈다는 증거가 될 수 없음 (**Uninformative**) 을 증명했습니다.

- **Pearson Correlation (0.00의 증거):**

- 값의 절대적인 크기가 아니라, 가중치의 '모양(Pattern)' 이 얼마나 비슷한지 보는 피어슨 상 관계수를 측정했습니다.
- 그 결과값은 0.0054 였습니다.
- 이는 두 모델의 가중치 패턴이 전혀 상관관계가 없음(**독립적임**) 을 의미합니다.

💡 결론 메시지

"Weak Evidence, Inconclusive"

- Sionic-AI의 주장은 증거 불충분/결정적이지 않음.

- "0.99라는 숫자에 속지 마라. 통계적으로 뜯어보니 두 모델의 가중치는 서로 독립적으로 학습된 것으로 보인다"는 의미.

3. 최종 결론 요약

🔥 Solar 표절 논란, 최종 종결 (feat. 조경현 교수)

Upstage Solar 모델이 중국 GLM을 베꼈다는 논란, 세계적인 AI 석학 조경현 교수(NYU)의 등판으로 기술적 검증은 의외로 간단히 끝난 듯 합니다.

쟁점	공격 측 (베꼈다!)	방어 및 검증 (아니다!)	최종 판정
코드	"GLM이랑 코드가 토씨 하나까지 똑같다!"	"MS Phi-2도 똑같다. 그건 업계 표준 템플릿이라 그렇다."	무죄 (관행)
가중치	"벡터 코사인 유사도가 0.99다! 빼박이다!"	[조경현 교수] "LayerNorm 특성상 원래 높게 나온다. 패턴(Pearson)을 보니 0.00이다. 전혀 안 베꼈다."	무죄 (통계적 입증)



3줄 요약

1. **Sionic-ai:** "숫자(0.99)는 거짓말 안 한다! 베꼈다!"
2. **Hyunwoongko:** "코드는 MS도 똑같다. 억까 하지 마라."
3. 조경현 교수: "0.99는 착시(Artifact)다. 통계적으로 뜯어보니 서로 다른 모델 맞다. 땅땅땅."

이제 이 논란은 "오픈소스 템플릿 활용에 대한 오해" 와 "RMSNorm의 수학적 특성에 대한 이해 부족"에서 비롯된 해프닝으로 정리할 수 있습니다.

Pebblous

Pebblous Makes Data Tangible

contact@pebblous.ai