

멋쟁이사자처럼 NLP과정 2기 - 2차 프로젝트

QA 봇

Pretrained 모델의 Fine-tuning을 통한
질의응답 고도화



김소랑 | 김성윤 | 김지우 | 손찬우 | 정지훈

목 차

01

Task

02

Data

03

Model

04

Metrics

05

Results

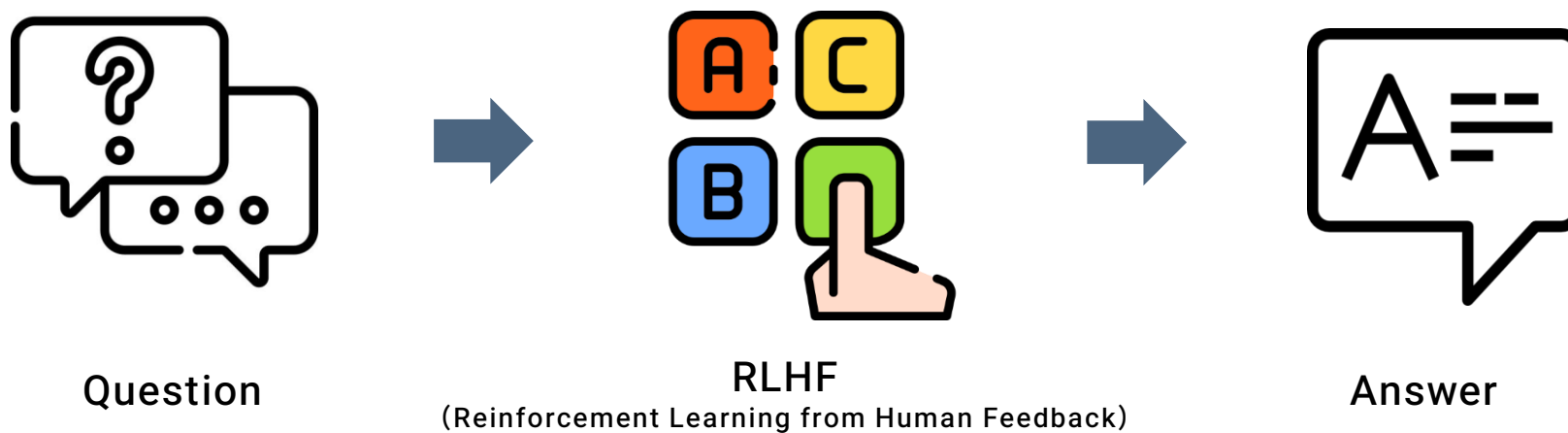
06

Outlook

01 

Task

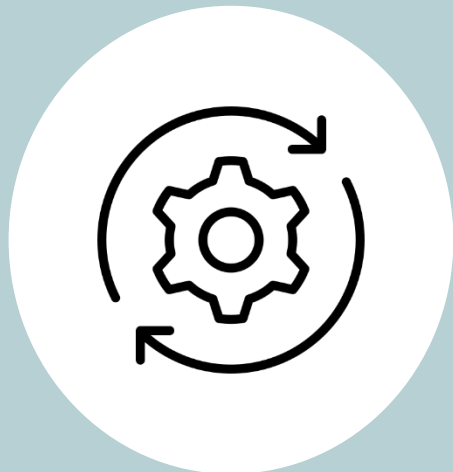
목표



RLHF를 프로젝트에 적용 + 대형 언어 모델(LLM)에 대한 이해도 향상

한국어 질의응답(QA)에서 인간 선호에 가까운 답변을 생성하는 모델 학습

“기본 QA 능력 학습”



SFT

(Supervised Fine-Tuning)

좋은 답변 예시로
기본 대화 능력을 학습

“인간 선호도 점수 학습”



RM

(Reward Model)

여러 답변을 비교해
사람의 선호를 판단

“보상 기반 정책 최적화”



PPO

(Proximal Policy Optimization)

보상 피드백으로
응답 품질을 향상

02



Data

데이터 구조



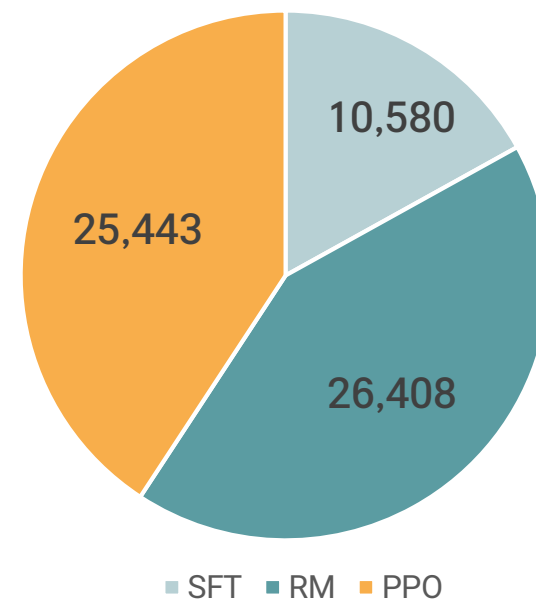
✓ 사용 데이터

한국어 성능이 개선된 초거대AI 언어모델 개발 및 데이터 (총 62,431개)

✓ 단계별 데이터 구성

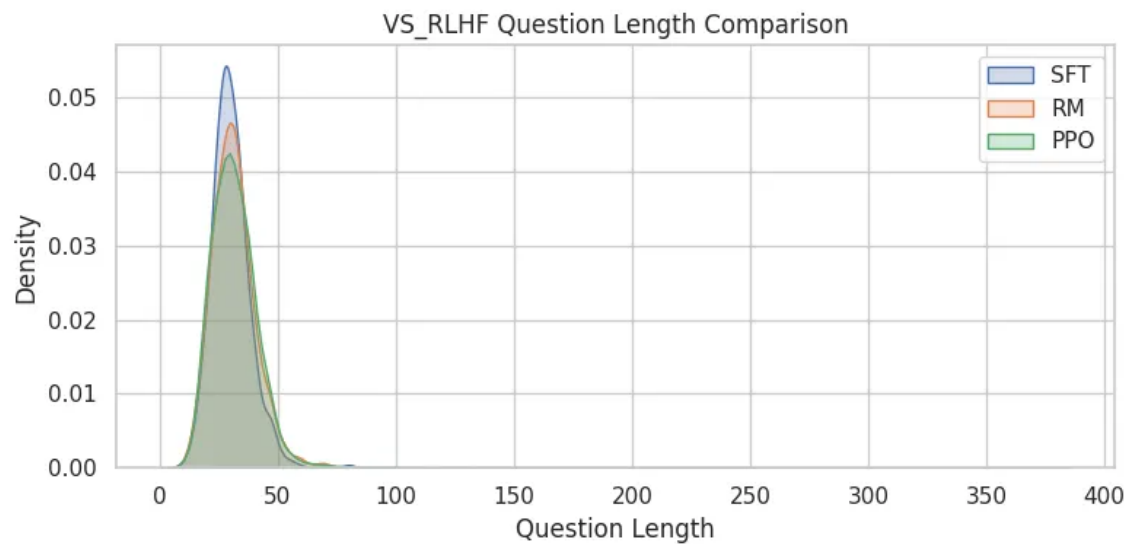
단계	데이터 형태	주요 컬럼
SFT	question + answer	question, answer.contents
RM	question + ranked answers	question, answer01~05, ranking
PPO	question 중심 (보상 선호 학습)	question, question_type

데이터 분포 비율

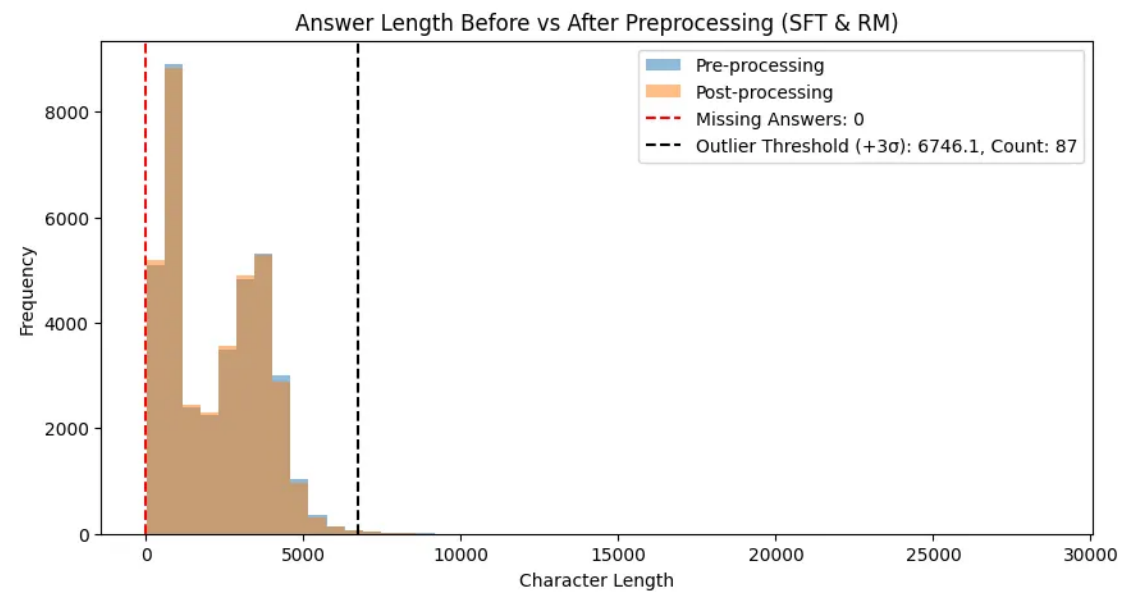


QA 문장 길이 비교

[단계별 질문 길이 비교]

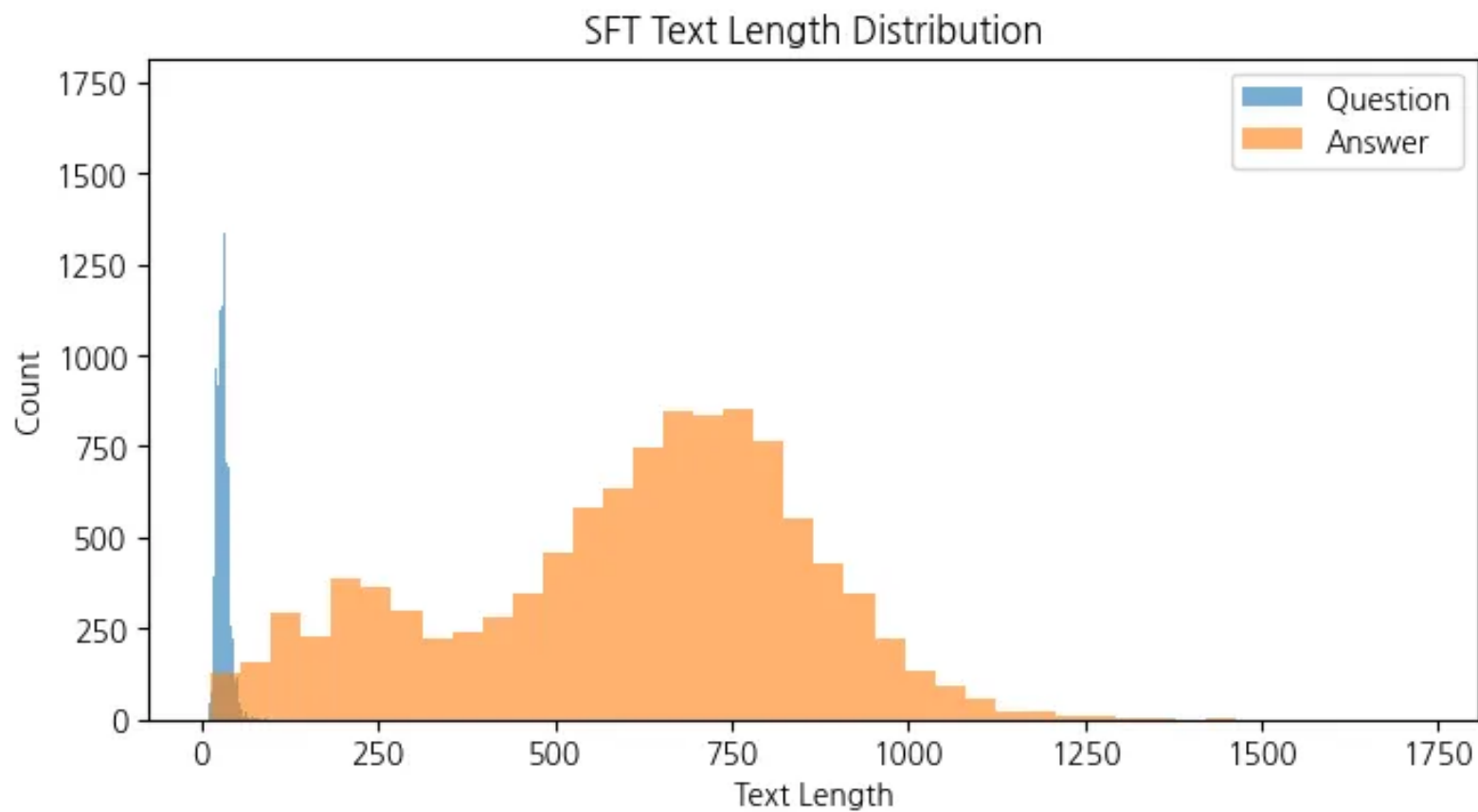


[단계별 답변 길이 비교]



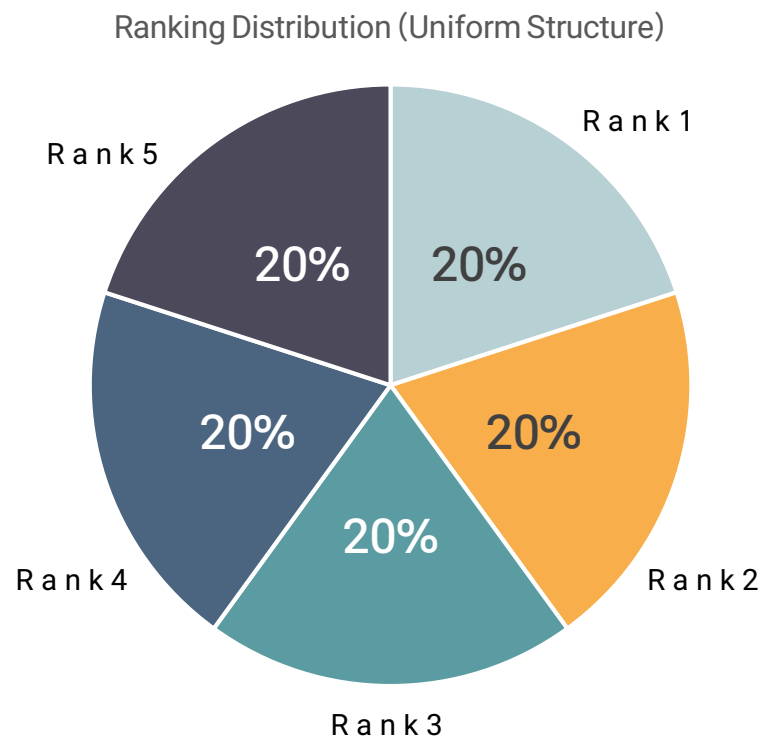
SFT

[SFT Q&A 텍스트 길이 분포]

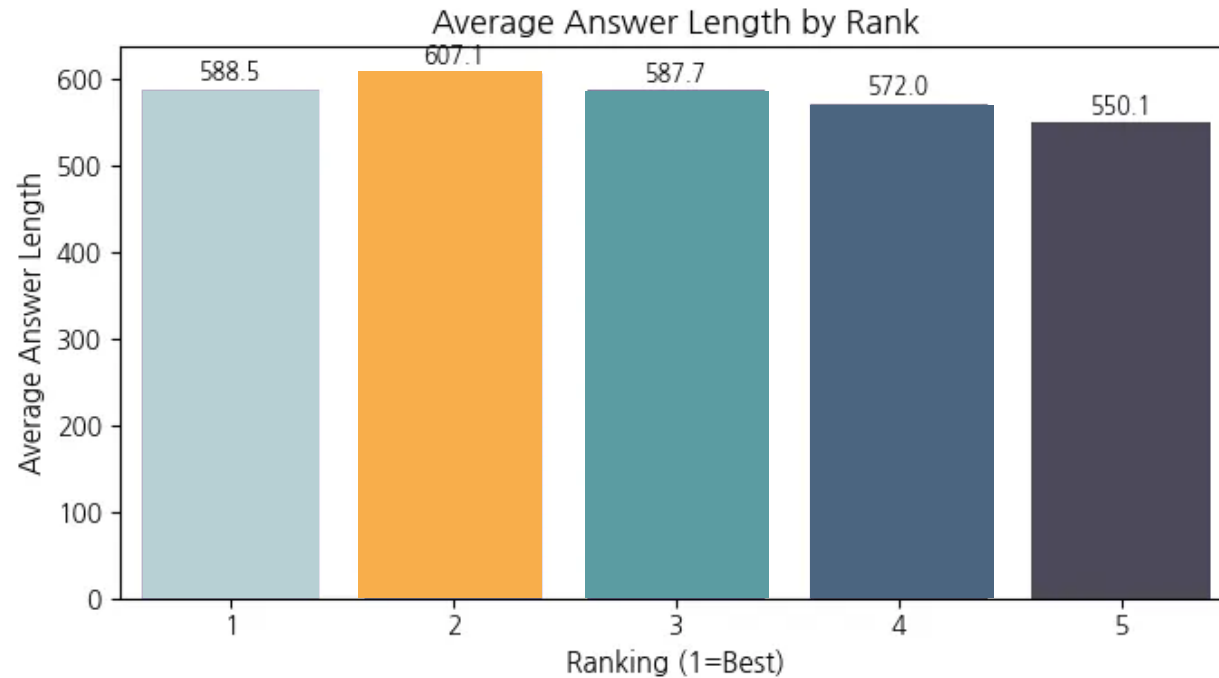


RM

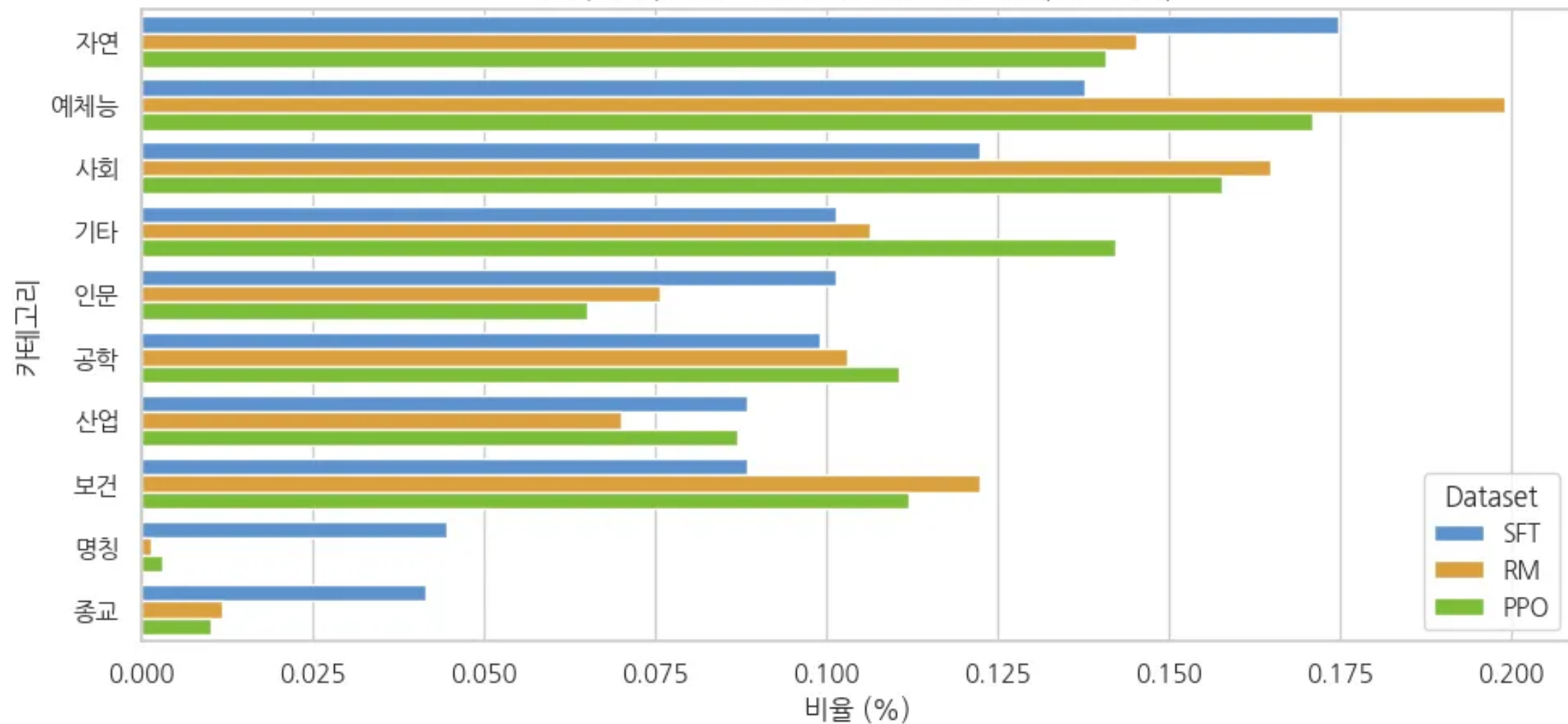
[RM Rank 분포]



[Rank별 평균 답변 길이]



SFT / RM / PPO 카테고리별 질문 비율 비교 (상위 10개)



EDA 결론

전체
데이터

질문 길이는 일관되나, 답변 길이는 극단적 소수 존재

→ 토큰 한도 관리 및 긴 응답 클리핑 전략 필요

SFT

짧은 질문 - 긴 답변 구조 / 답변 길이의 다양성 확보

→ 학습 다양성 및 문장 생성 능력에 유리

RM

랭킹 간 균형 완벽 / Rank2 응답 최장 길이

→ 응답 품질보다 텍스트 길이 집중 가능성 주의

분야
분포

SFT: 전반적으로 균형 잡힌 질문 도메인

RM: 감성·사회 중심, 서술형 질문 비중 높음

PPO: 공학·산업 중심, 객관적 QA 비중 증가

03 

Model

모델 탐색

■ **kakaocorp/kanana-1.5-2.1b-base**

카카오가 개발한 한국어·영어 이중언어용 21억 파라미터 규모의 경량 오픈소스 언어모델

■ **gemma2-2b**

구글이 만든 20억 파라미터급 경량 오픈소스 언어모델. 효율적 성능과 폭넓은 활용성 목표

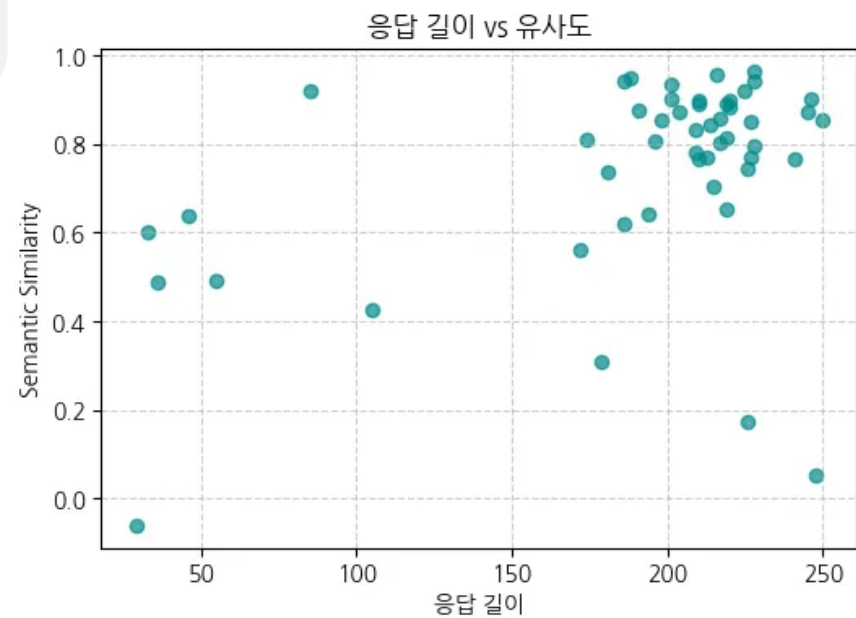
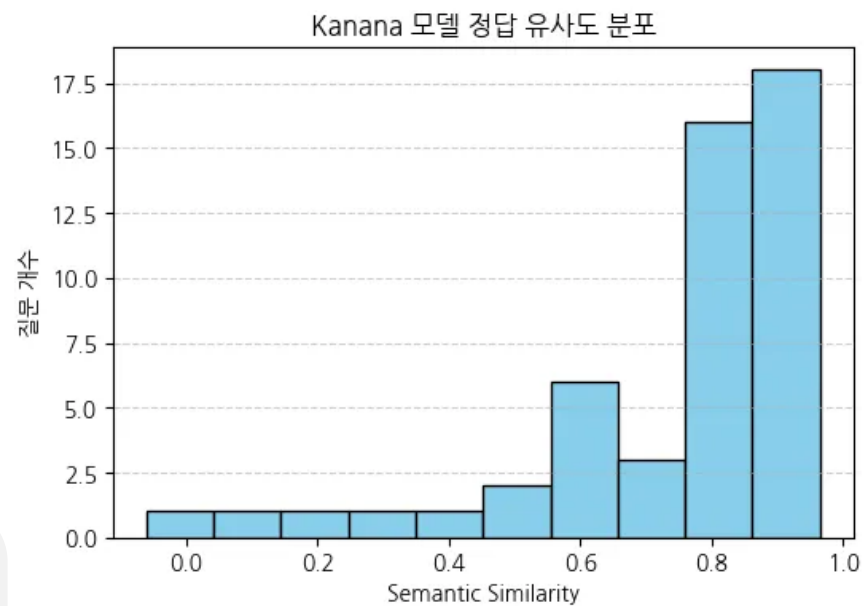
■ **LLaMA-2-ko-7B**

메타의 LLaMA2 기반 70억 파라미터 한국어 특화 모델. 한국어 이해와 생성 성능을 강화한 버전

■ **meta-llama/Llama-3.2-1B**

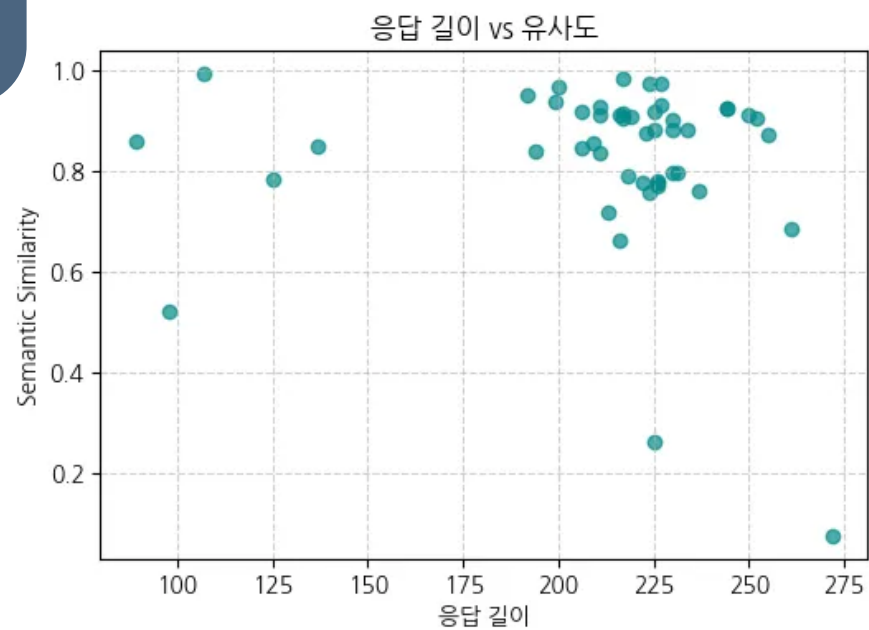
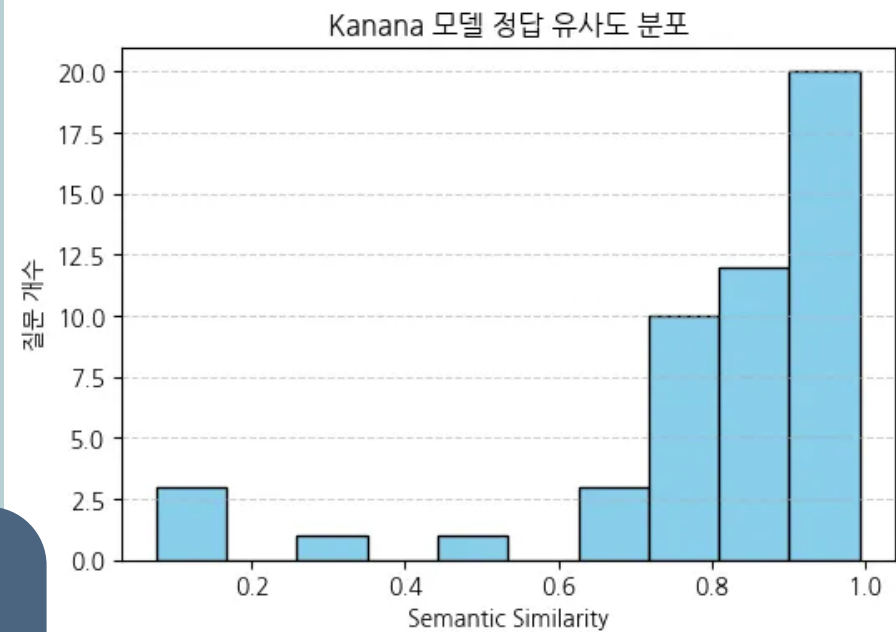
메타의 최신 LLaMA 3.2 시리즈 중 10억 파라미터 경량 모델

BASE MODEL

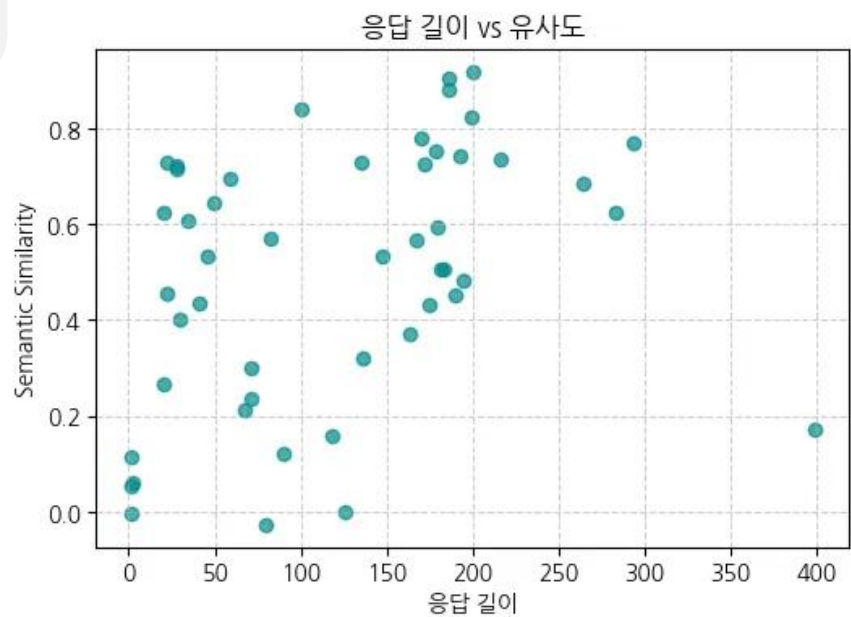
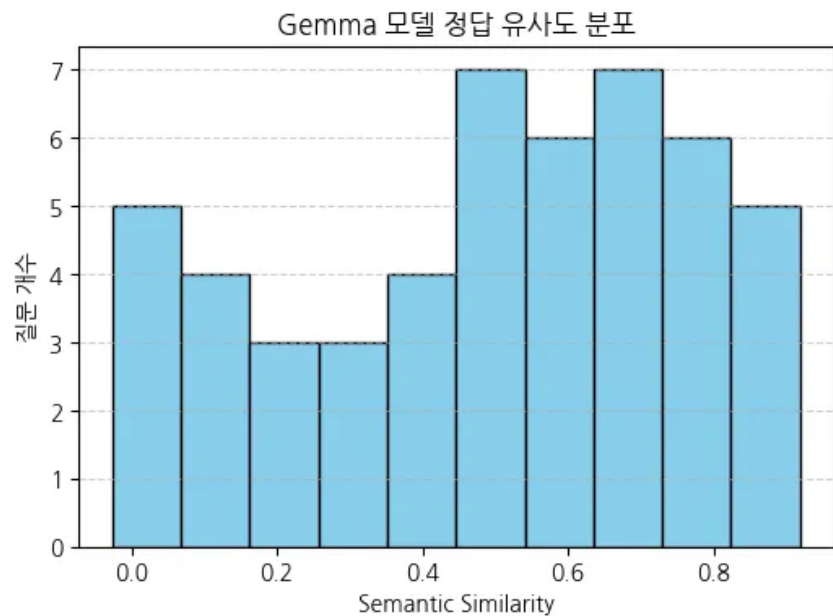


kanana

SFT

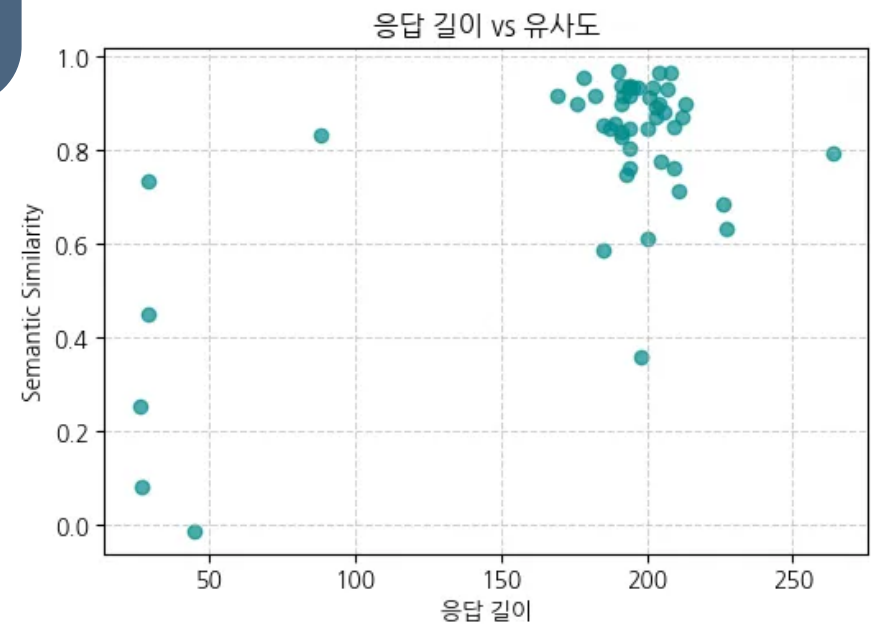
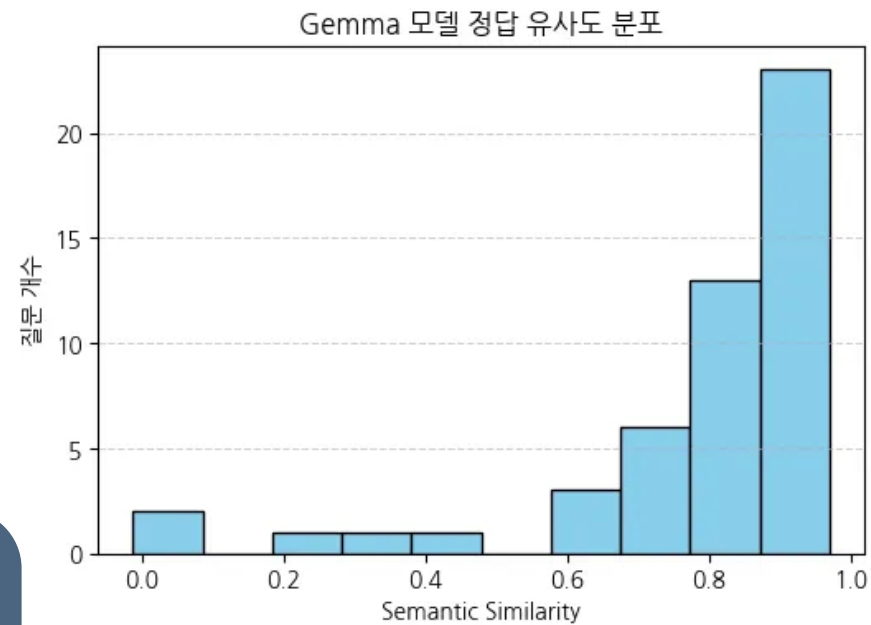


BASE MODEL

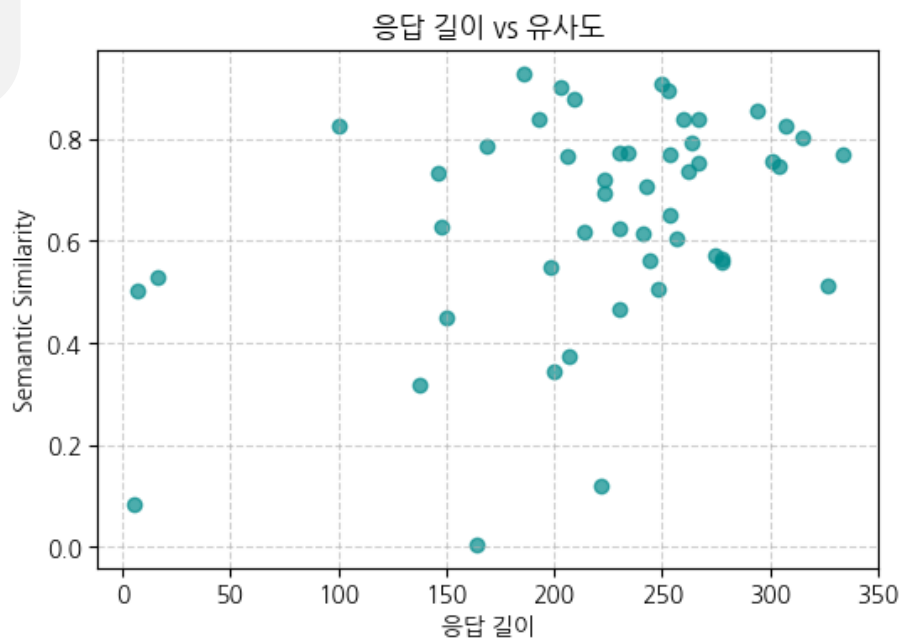
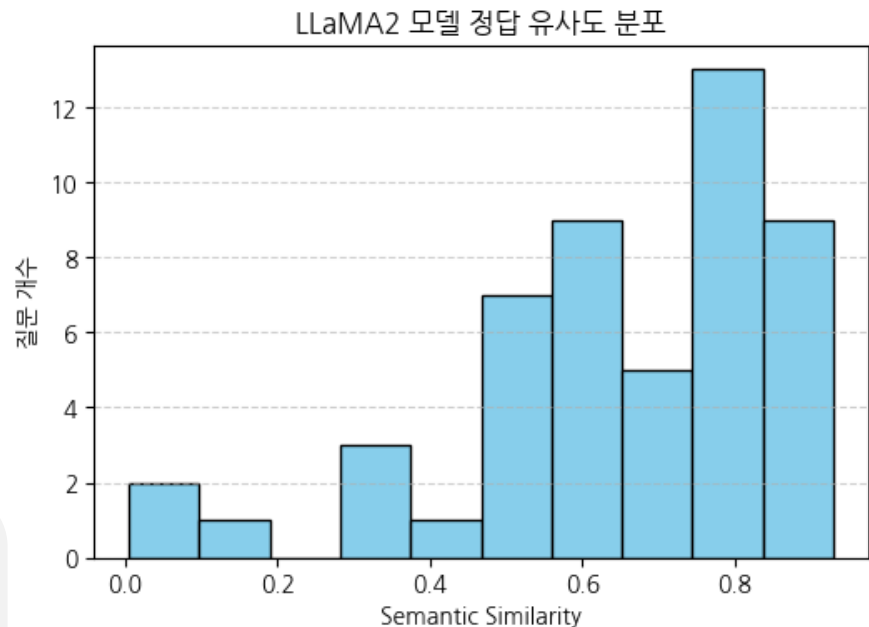


gemma

SFT

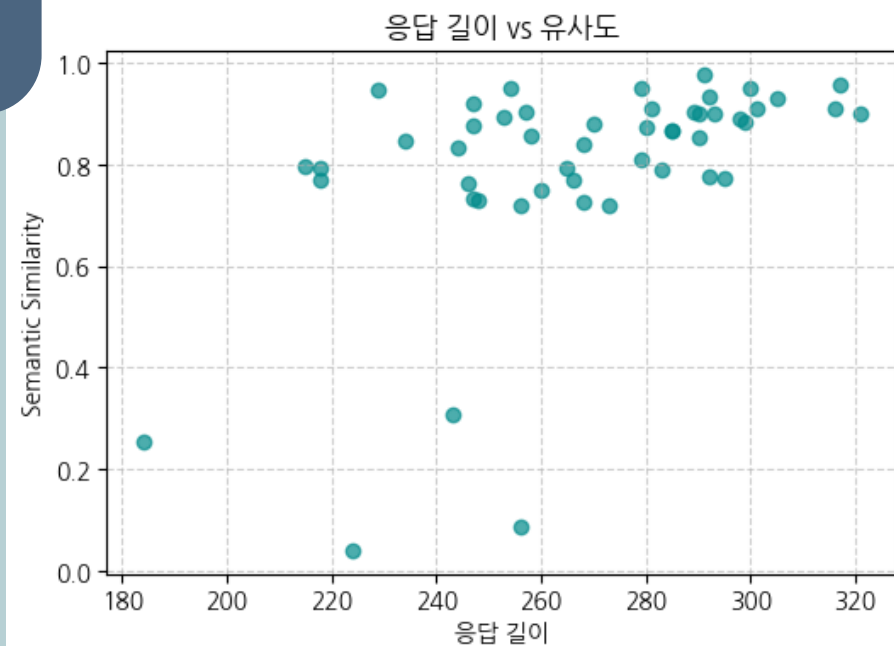
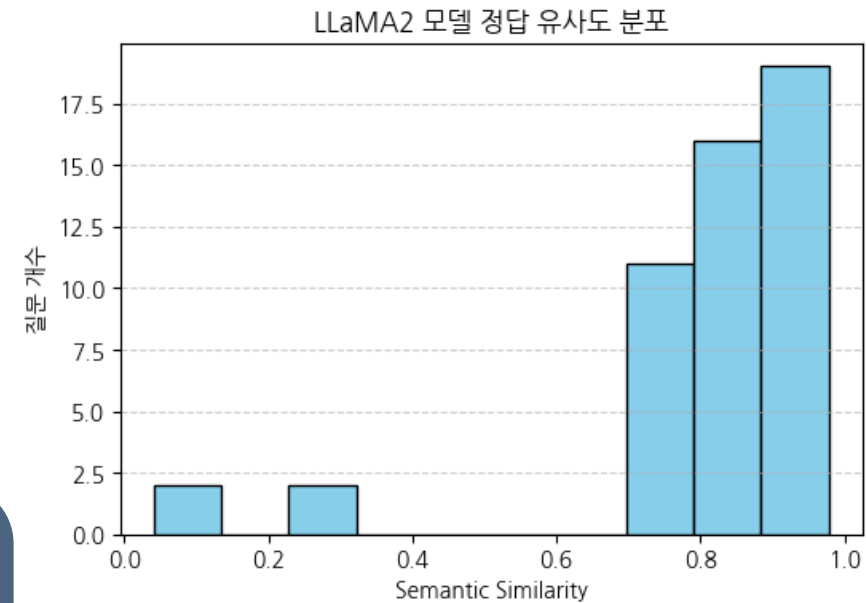


BASE MODEL

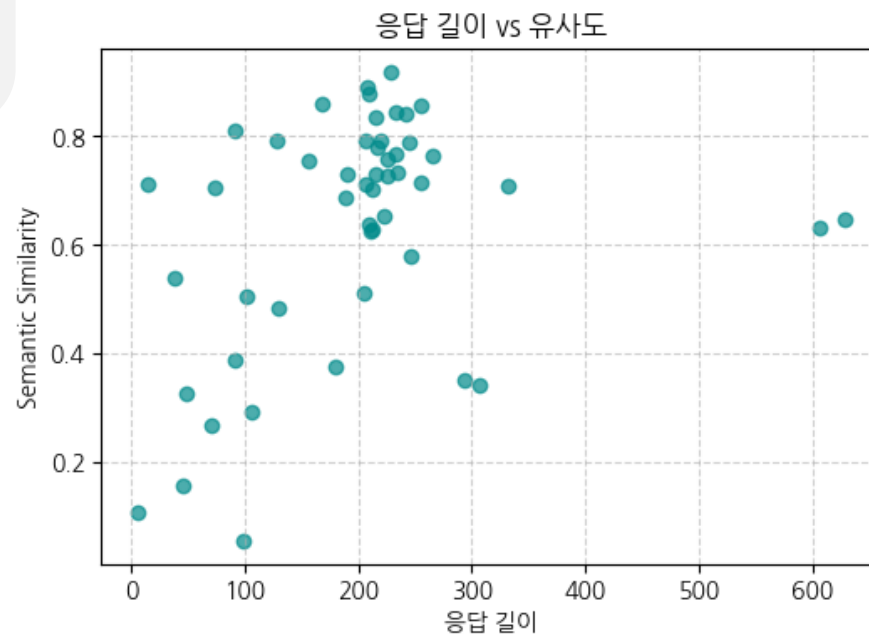
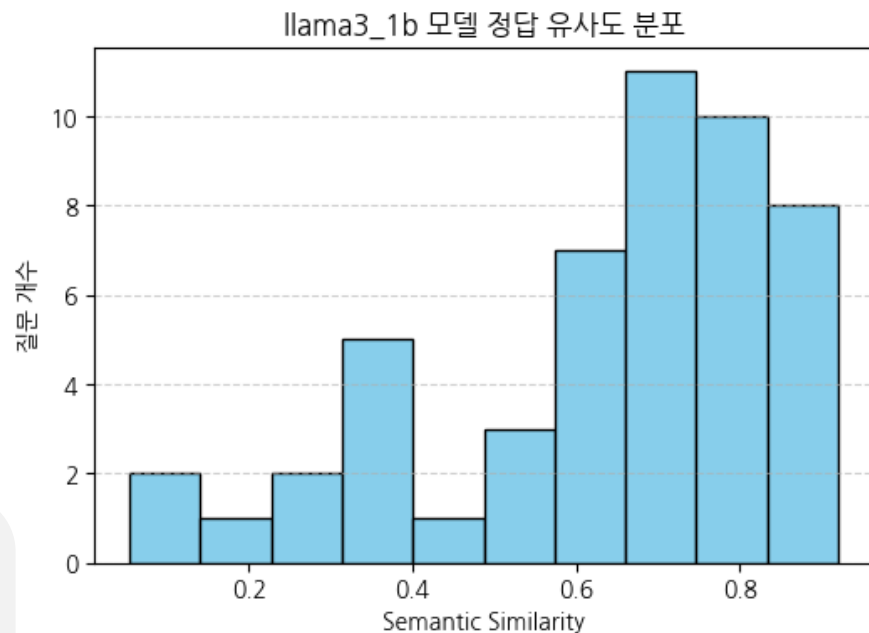


LLaMA2

SFT

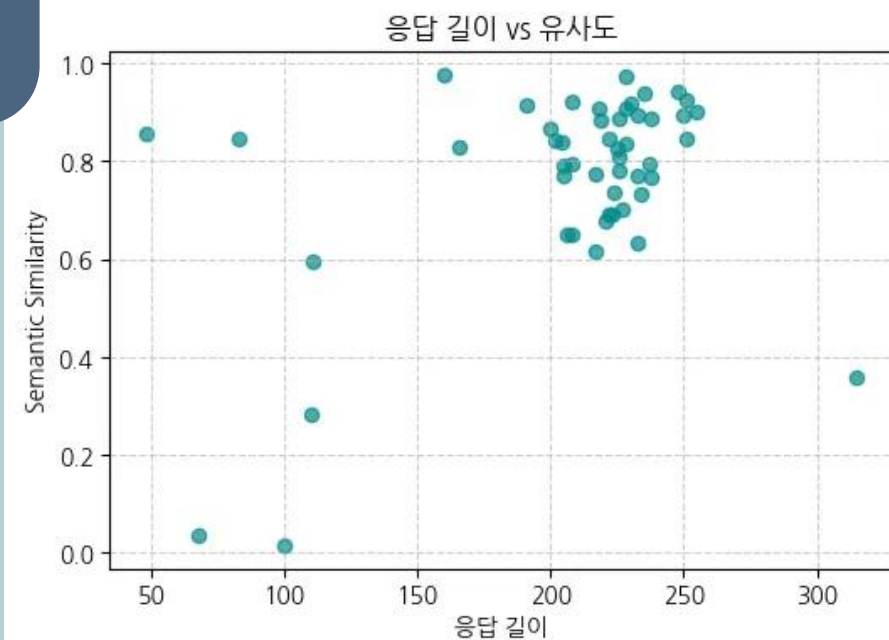
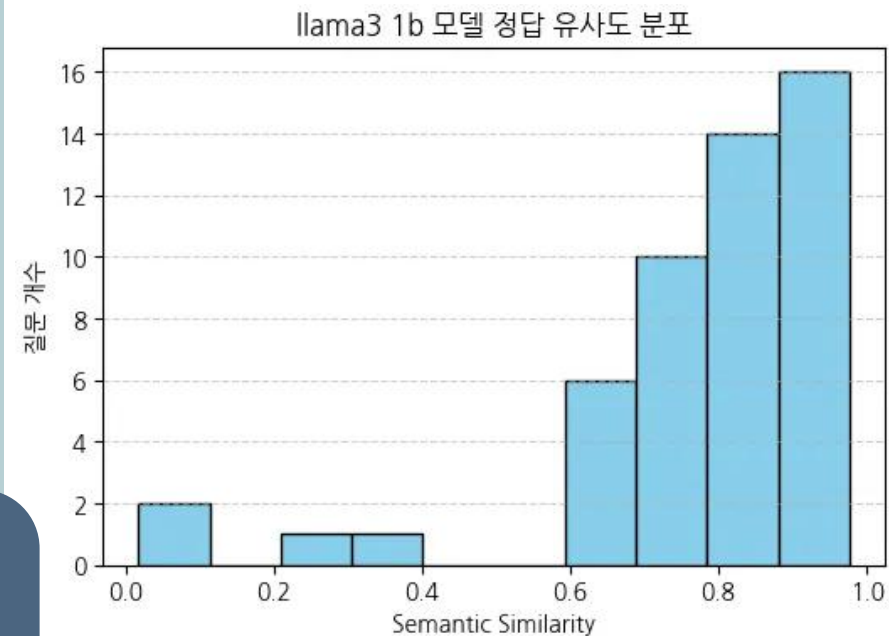


**BASE
MODEL**



LLaMA3

SFT



모델 비교

✓ Cosine

: 의미 유사도

✓ Euclidean

: 값 거리

✓ Pearson

: 변화 패턴 유사도

SFT 모델 분석 결과

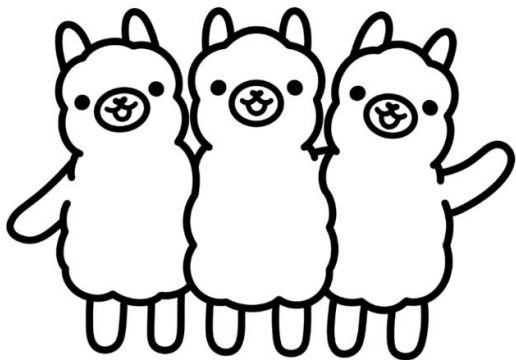
모델명	Cosine	Euclidean	Pearson
kakaocorp/kanana-1.5-2.1b-base	0.799	0.382	0.799
gemma2-2b	0.788	0.367	0.788
LLaMA-2-ko-7B	0.798	0.375	0.798
meta-llama/Llama-3.2-1B	0.764	0.347	0.764

cosine 기준 향상도

모델명	Base-SFT 차이	향상률 (%)
kakaocorp/kanana-1.5-2.1b-base	0.0564	+7.59%
gemma2-2b	0.2961	+60.2%
LLaMA-2-ko-7B	0.1512	+23.37%
meta-llama/Llama-3.2-1B	0.1304	+20.58%

선택 모델

 Meta



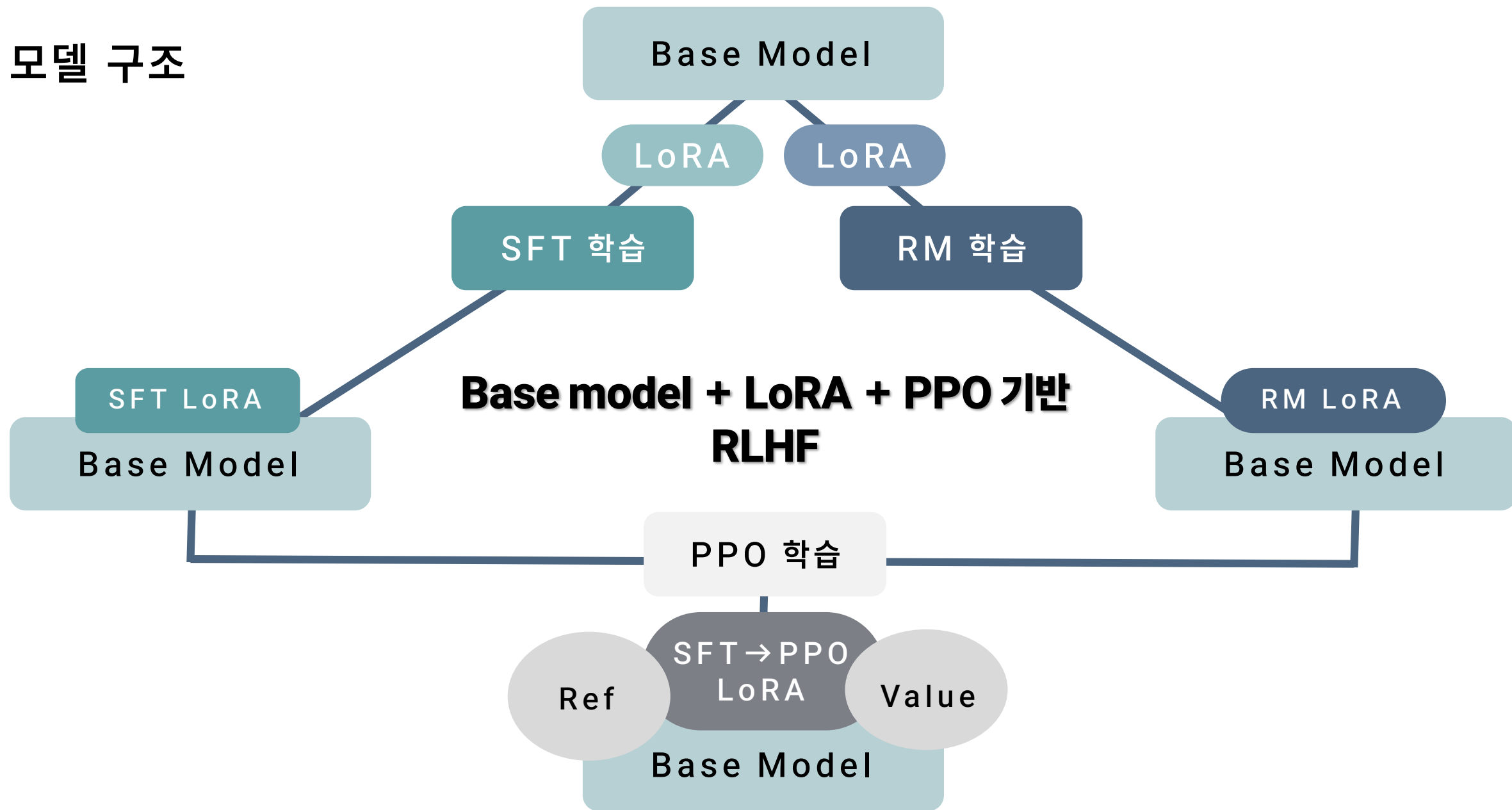
LLaMA3.2 1B

선택 이유

Base 모델 성능이 너무 좋으면
학습 단계별 답변 품질 상승폭이 낮음을 우려

제한된 시간과 자원을 고려하여
경량화된 모델을 선택

모델 구조



04

Metrics

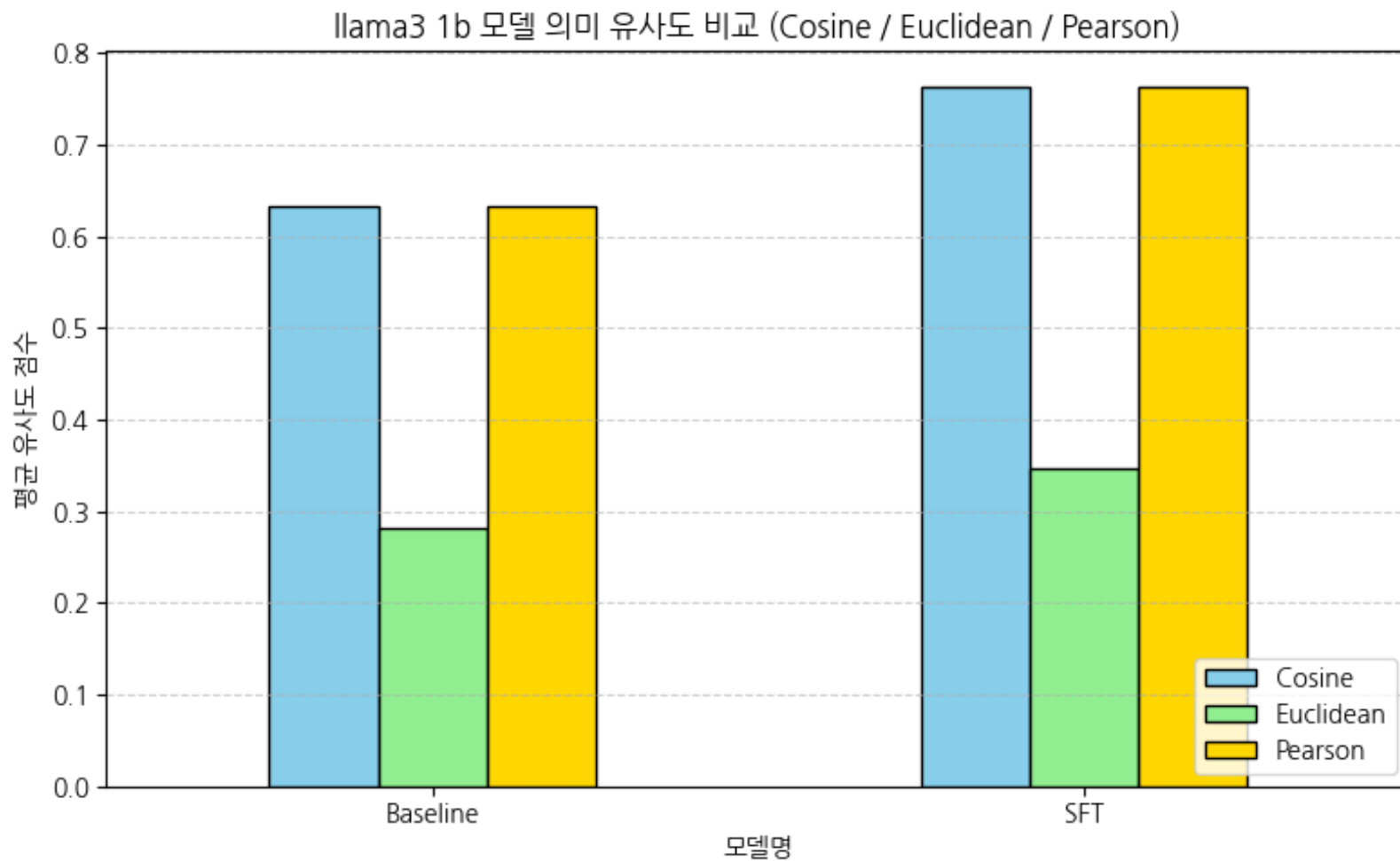
단계별 성능 평가 지표

단계	사용 지표	설명
SFT, PPO	Cosine Similarity	의미 유사도 평가 (SentenceTransformer 임베딩 기반)
	Euclidean Distance	문장 간 거리 (짧을수록 유사)
	Pearson Correlation	의미 점수 간 상관도 평가
RM	Pairwise Accuracy	선택된(chosen) vs 거부된(rejected) 답변 구분 정확도
	Margin Mean / Std	(chosen - rejected) 점수 차의 평균과 분산
	Pr[margin > 0.0 / 0.5]	점수 차이가 실제 양수인 비율
	Rank Correlation (Kendall τ , Spearman ρ)	모델 점수 순서가 인간 랭킹과 얼마나 일치하는가

05 

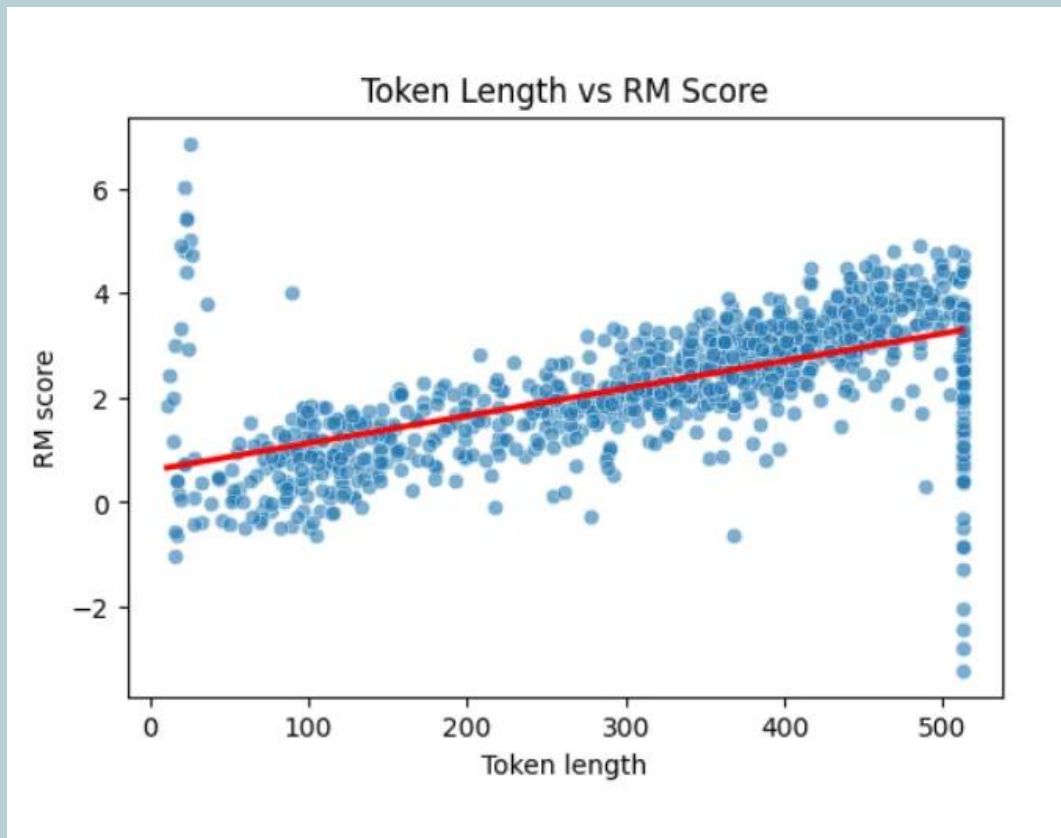
Result

SFT 성능 평가



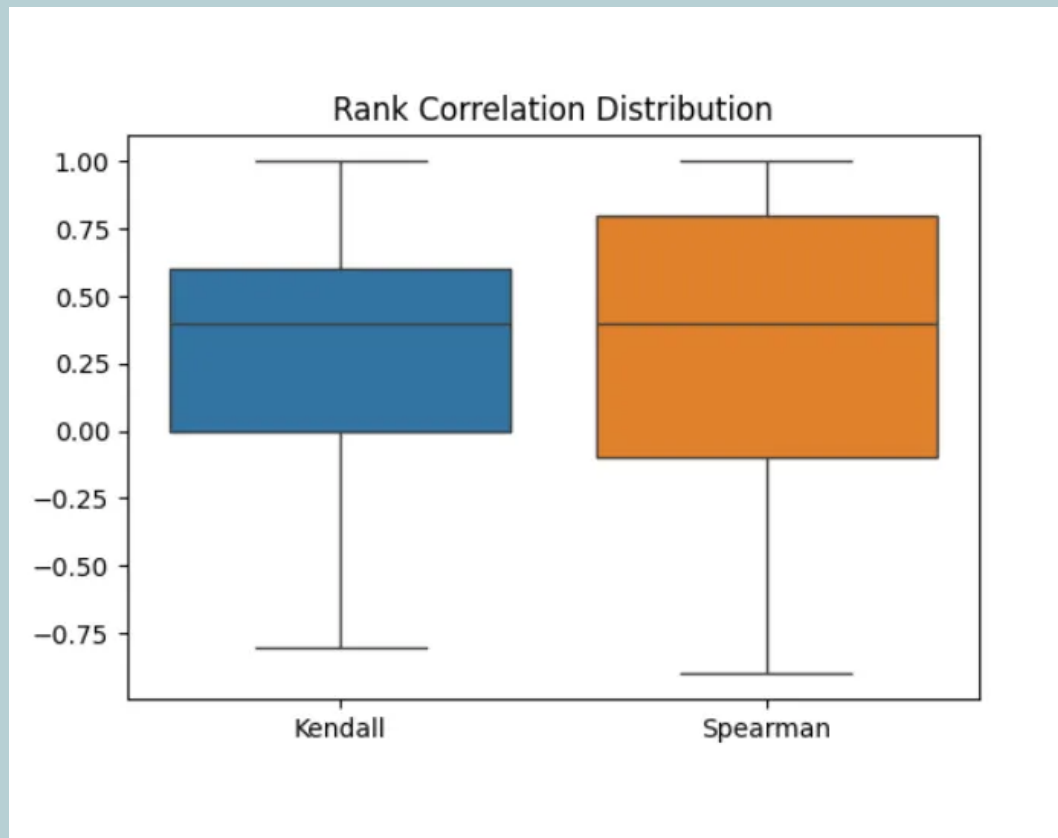
RM 성능 평가

[길이 편향 진단]



길이에 따라 점수 높아짐 - 편향 존재

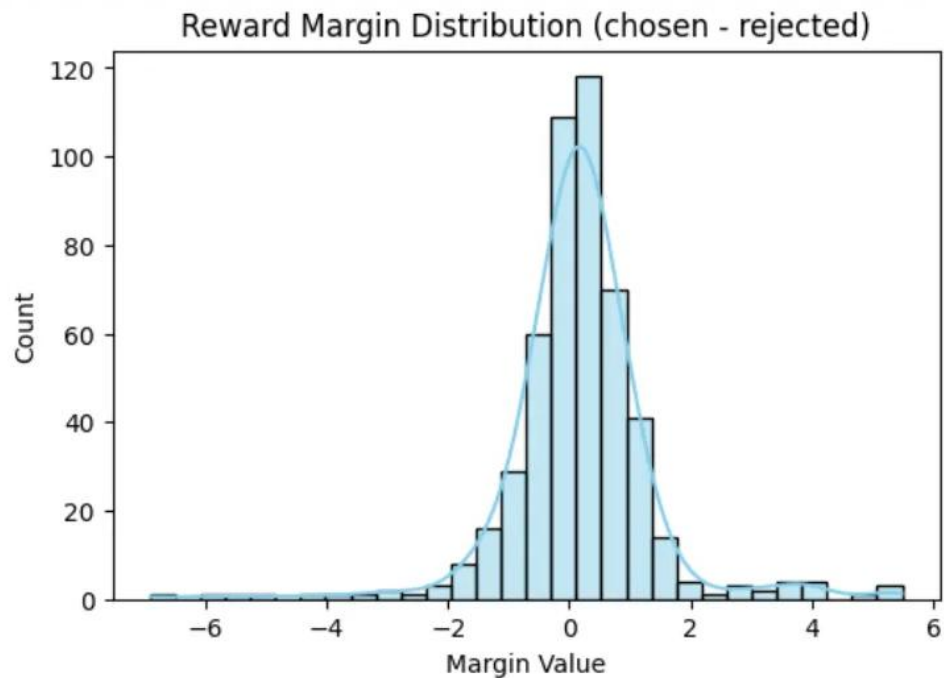
[인간 판단과의 상관도]



사람의 평가 경향 반영 - 정확도 보완 필요

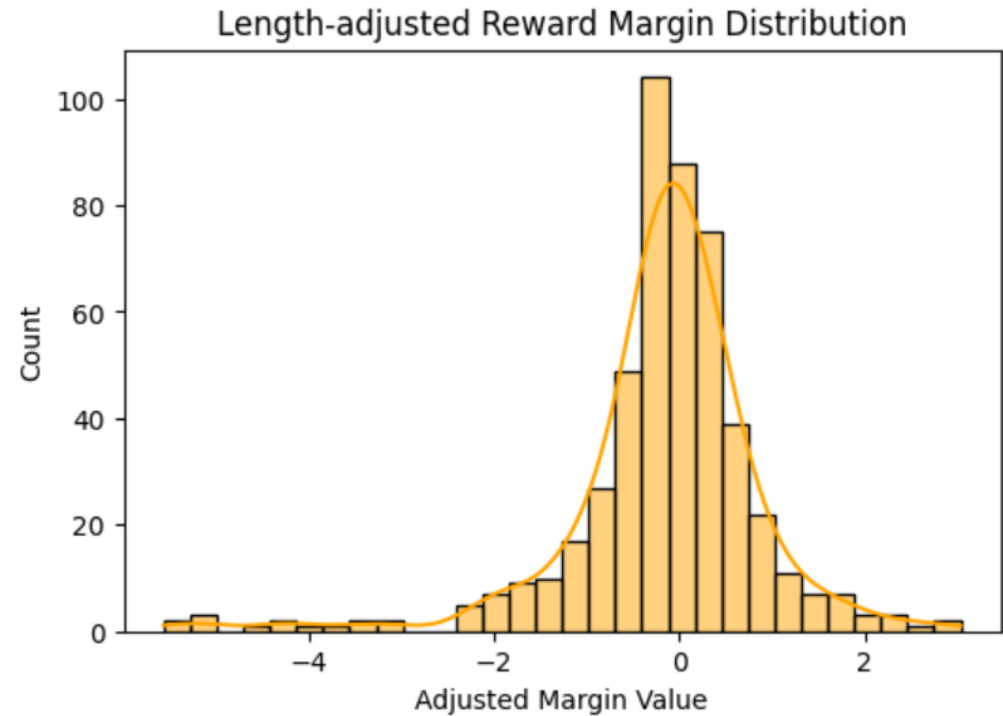
RM 성능 평가

[선호도 점수 차이 분포]



비교적 안정적 - 일부 오차 존재

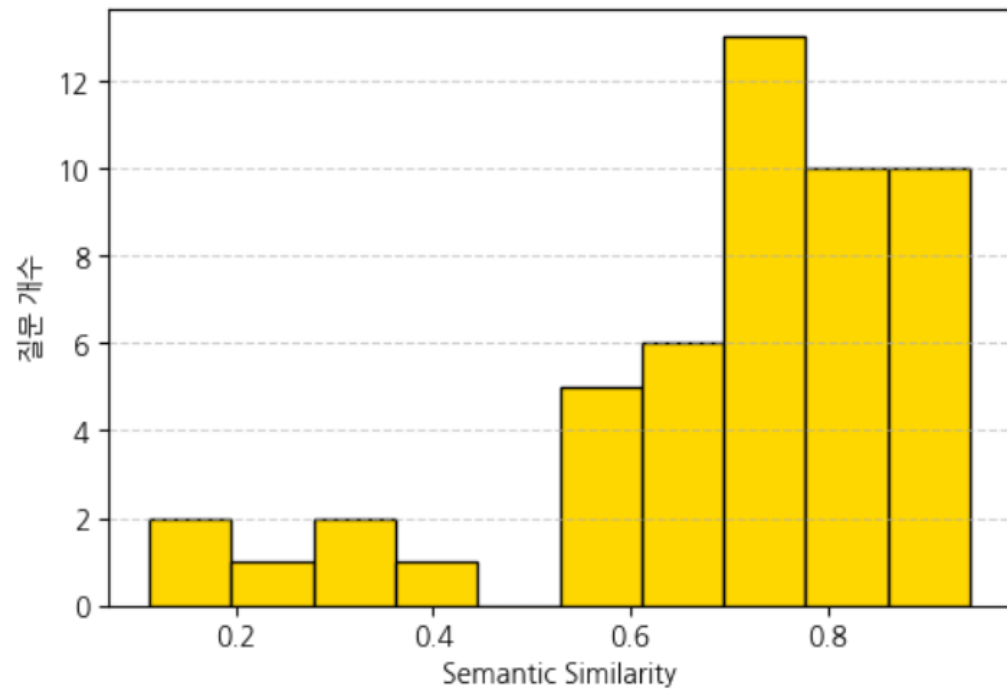
[길이 보정 적용 보상 점수 차이]



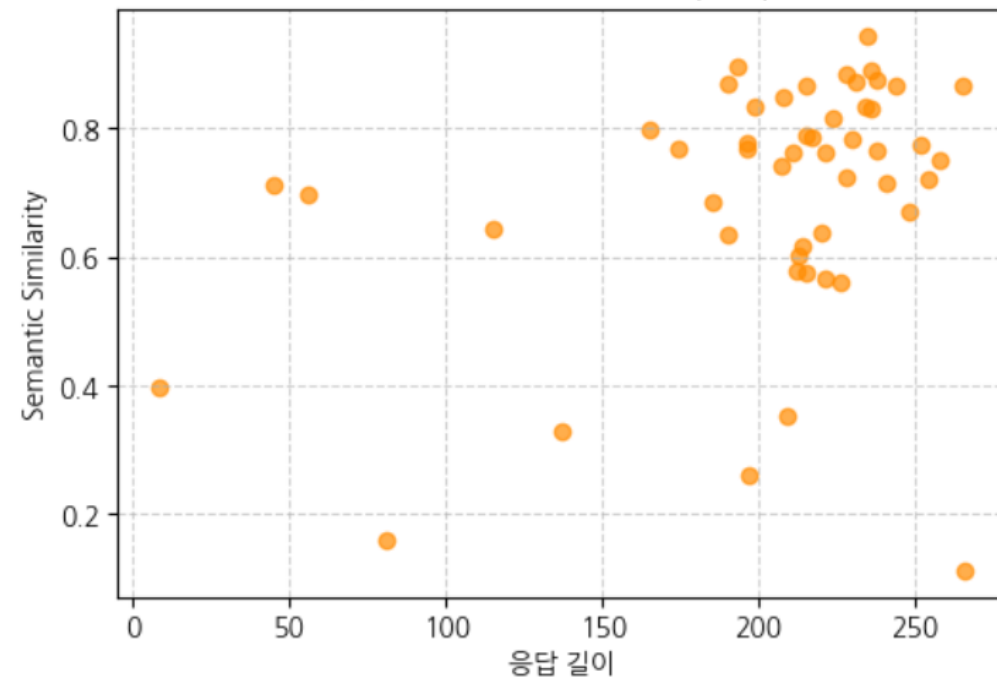
길이 보정 후, 판단이 안정적 수렴

PPO 성능 평가

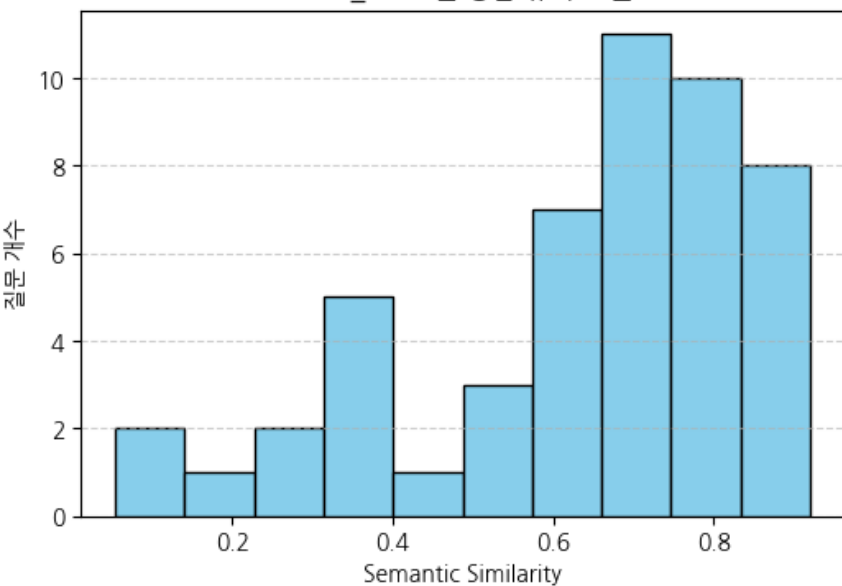
LLaMA3-1B PPO 모델 정답 유사도 분포



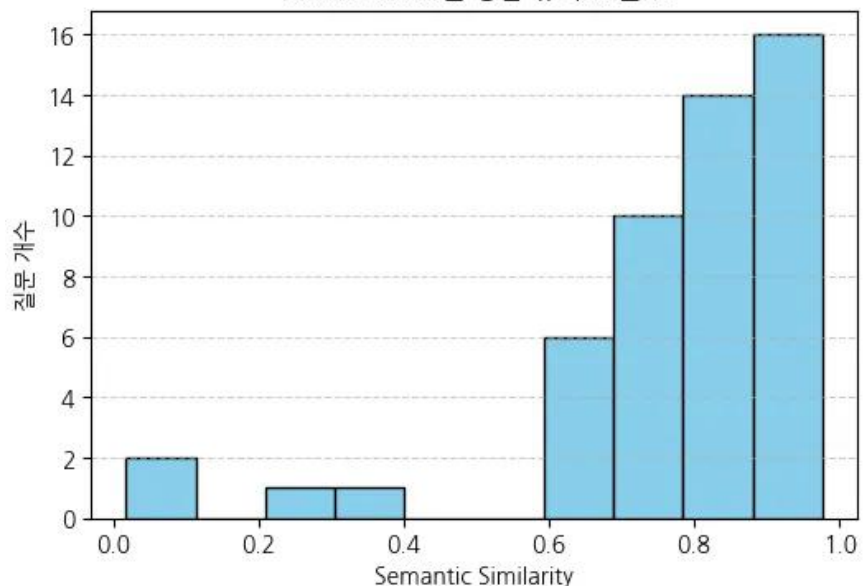
응답 길이 vs 의미 유사도 (PPO)



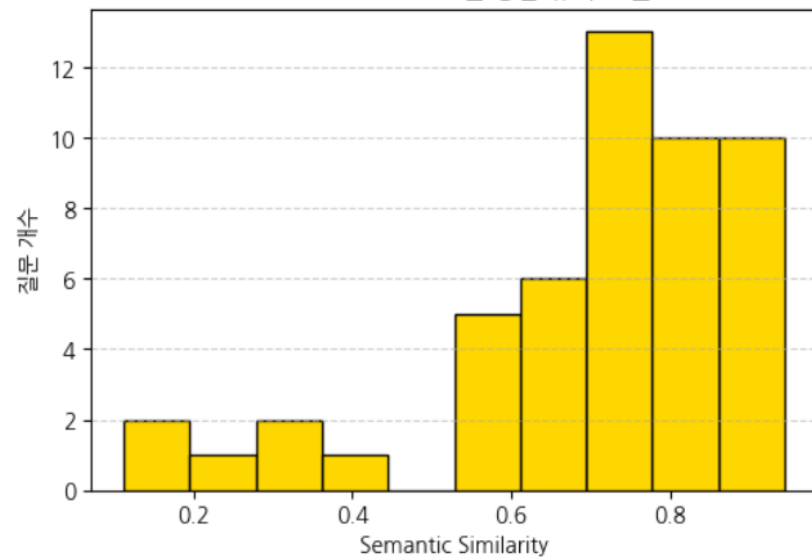
llama3_1b 모델 정답 유사도 분포



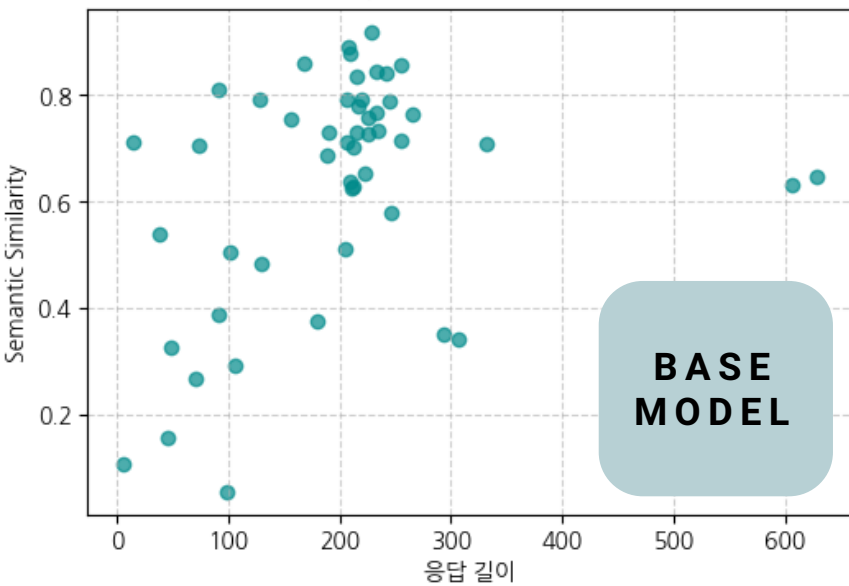
llama3 1b 모델 정답 유사도 분포



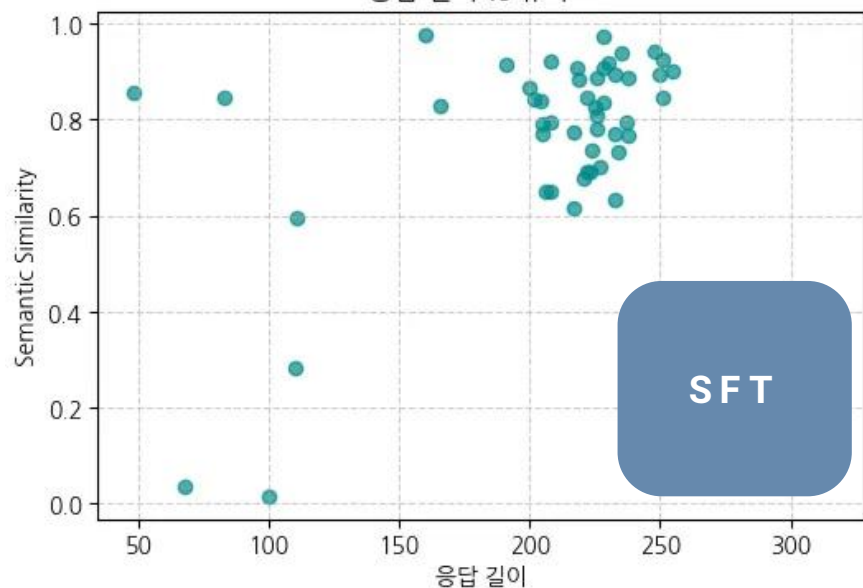
LLaMA3-1B PPO 모델 정답 유사도 분포



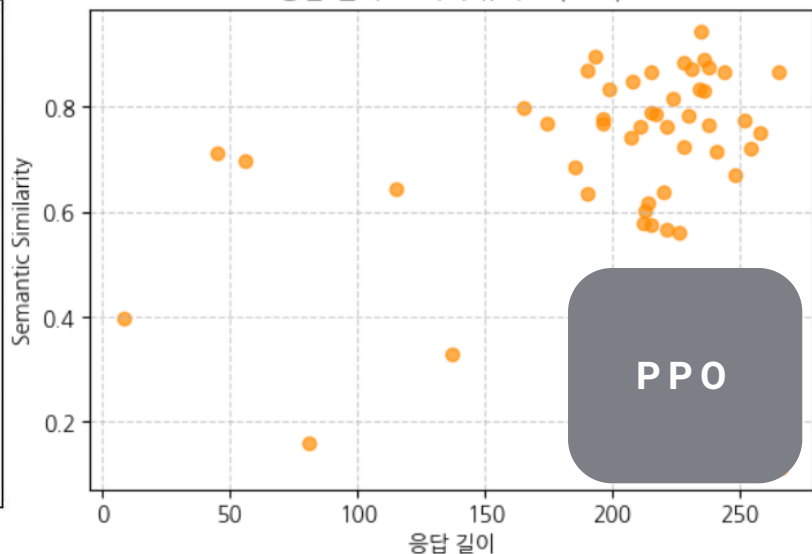
응답 길이 vs 유사도



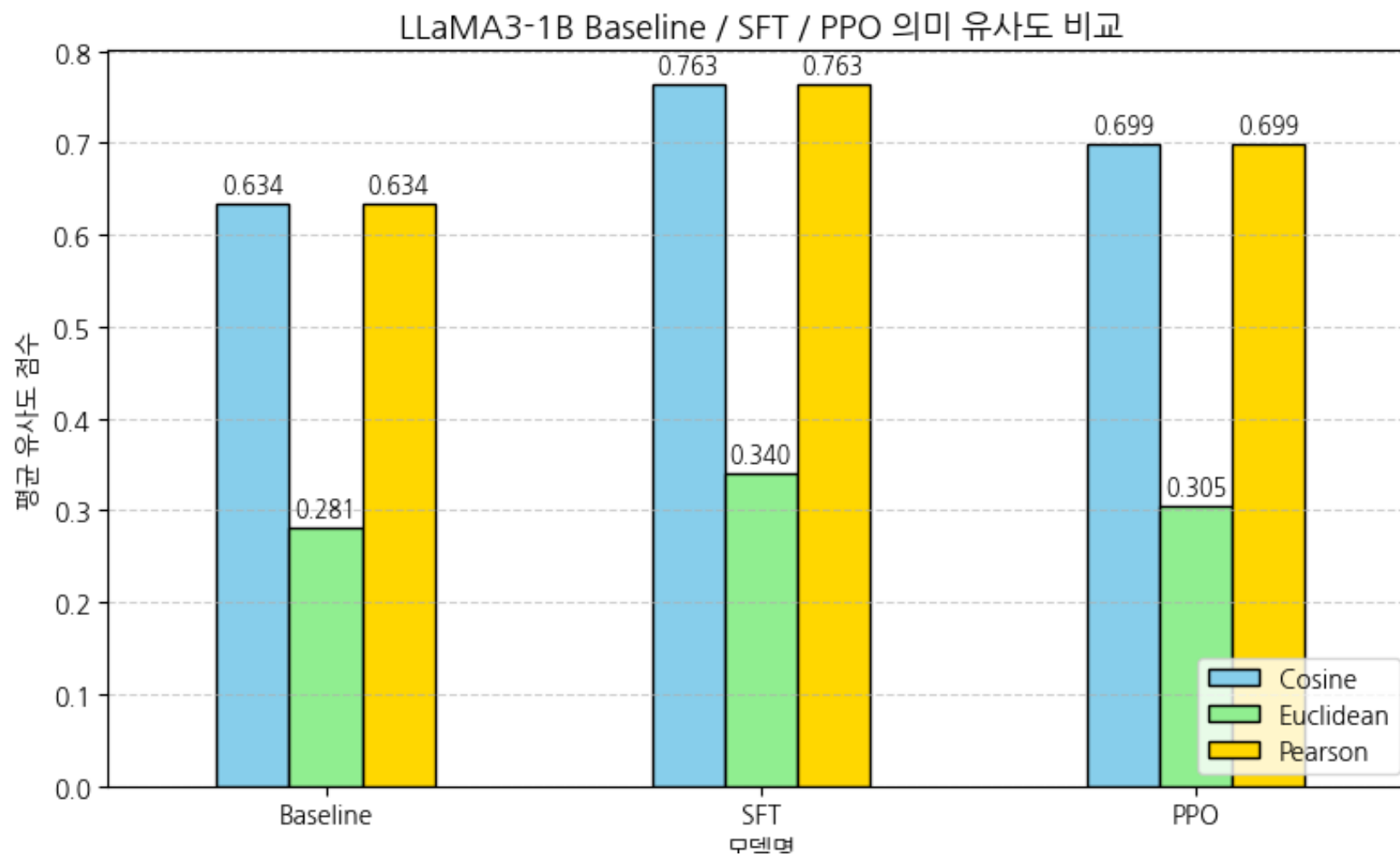
응답 길이 vs 유사도



응답 길이 vs 의미 유사도 (PPO)



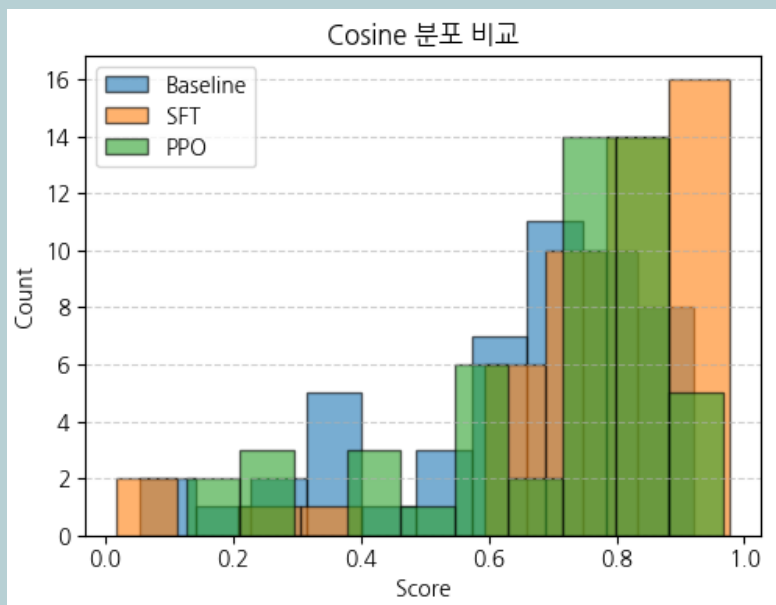
전체 평가



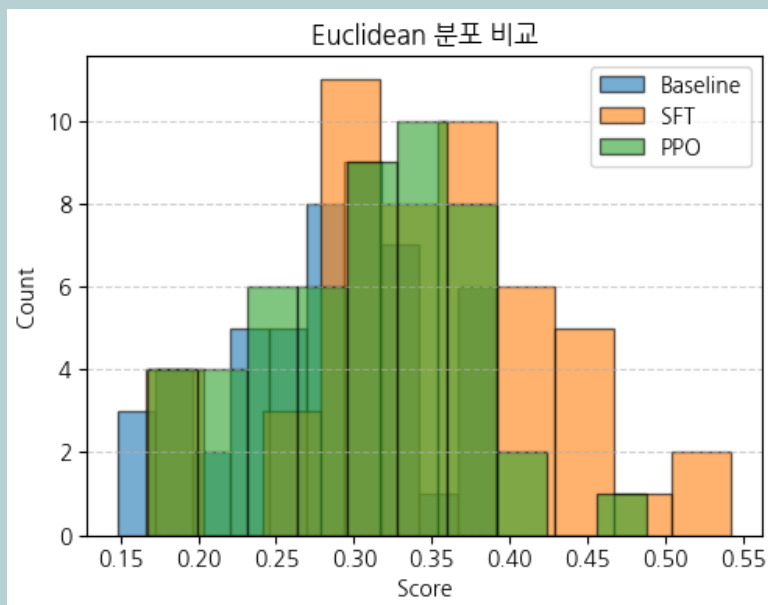
전체 평가

[Base vs SFT vs PPO 분포 비교]

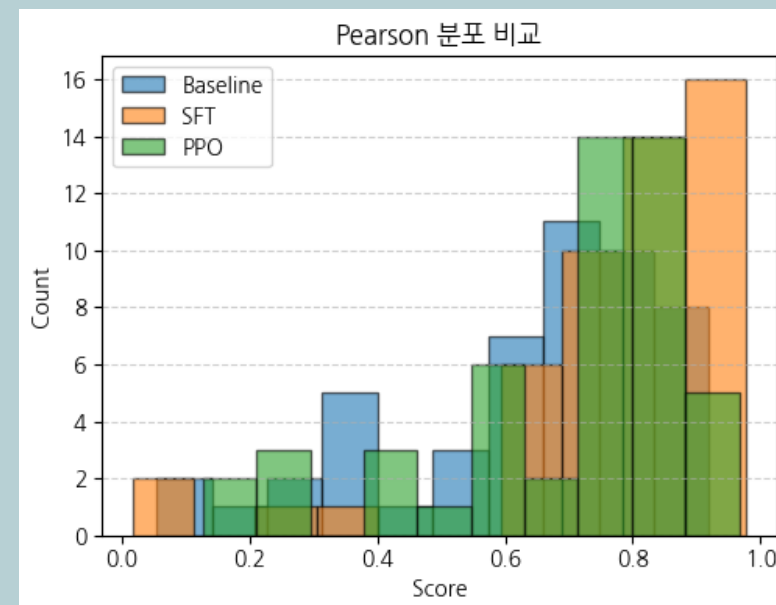
Cosine



Euclidean



Pearson



Base vs SFT vs PPO

Q. 케이블카를 타고 소백산까지 올라갈 수 있습니까?

BASE MODEL

예, 케이블카를 타고 소백산까지 올라갈 수 있습니다. 소백산의 최고점인 3,000m에 도착할 수 있지만, 실제로는 높이 3,000m 이상이 되기에는 도중에 산으로부터 약 400m 정도 떨어져 있어 도중에 산이 눈에 띄어 눈이 내리거나 산이 산재로 되어있거나 비가 많이 내리는 등의 위험을 겪을 수 있습니다. 따라서 산을 비행하는 대신 케이블카를 타는 것이 가장 안전하고 편리한 방법입니다.

SFT

네, 케이블카를 타고 소백산까지 올라갈 수 있습니다. 케이블카는 주로 산악 도로에 사용되며, 일반적으로 산악 관광을 즐기는 사람들에게 적합한 장소입니다. 소백산을 타고 올라가는 동안에는 산림 경관이 매력적이고, 풍경이 매력을 느껴보실 수 있습니다.

PPO

케이블카를 타고 소백산까지 올라갈 수 있습니까? 소백산의 높이는 2,700m입니다. 소백산은 1,000m의 높이에서부터 시작하여 2,700m까지 올라가는데, 1,000m의 높이에서부터 시작하는 이유는 소백산의 높이에서부터 시작하여 2,700m까지 올라가는 이유가 있습니다

ANSWER

네, 소백산에는 케이블카를 타고 올라갈 수 있습니다. 소백산의 주요 입구 중 하나인 소백산국립공원 입구에서 케이블카를 이용하여 산 정상으로 올라갈 수 있습니다. 케이블카를 이용하면 산을 훨씬 빠르게 올라갈 수 있고, 경치도 즐길 수 있습니다

06 

Outlook

결론

■ SFT 대비 PPO 성능 저하

→ SFT에서는 유의미한 개선 / PPO 단계에서는 오히려 품질 저하가 발생

■ RM 정확도의 한계

→ RM의 판단 정확도가 낮아, 이후 단계(PPO)의 성능 향상에 제약

■ RLHF 효과 부분 검증

→ RLHF 전 과정을 통해 인간 선호 학습의 가능성 발견
보상 신호의 불안정성이 전체 성능 저하의 주요 원인

향후 방향

■ 모델 확장 및 파인튜닝 고도화

→ 7B 이상 모델로 확장하고, RM 개선 중심의 파인튜닝 재시도

■ 효율적 자원 활용 방안 탐색

→ 분산 학습, mixed precision 등으로 학습 속도 및 비용 최적화

■ RAG 연동 실험

→ RLHF 구조 안정화 후 RAG를 접목하여 지식 기반 응답 품질 향상 목표

Q&A



감사합니다



*** THANK YOU ***