2025-10-02
인공지능(AI)말평

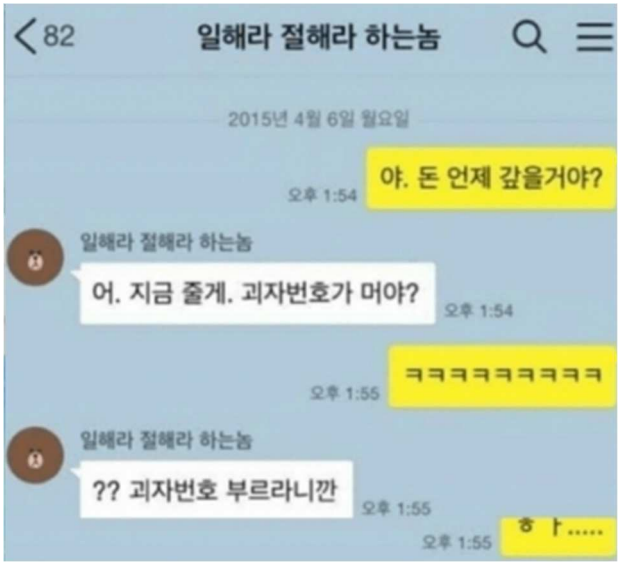
[2025]한국어 어문 규범 기반 생성(RAG)(가 유형)

<overfit-brothers Pro Max>
임형준, 유용상, 이기훈

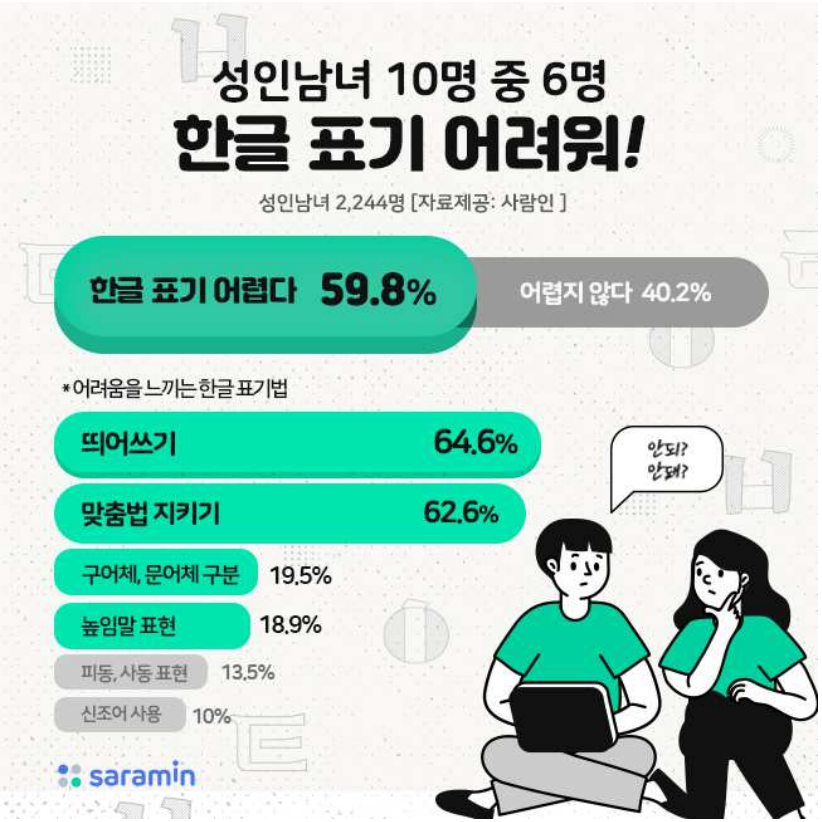
Contents

0 맞춤법은 어려워


27
소개팅남
2024년 12월 20일 금요일 >
오빠 감기 괜찮아?
오전 11:42
애교뿜뿜 어피치
어 많이 낳았어 걱정해져서 고마워
오전 11:43


82
일해라 절해라 하는놈
2015년 4월 6일 월요일
야. 돈 언제 갚을거야?
오후 1:54
일해라 절해라 하는놈
어. 지금 줄게. 괴자번호가 머야?
오후 1:54
ㅋㅋㅋㅋㅋㅋㅋㅋ
오후 1:55
일해라 절해라 하는놈
?? 괴자번호 부르라니깐
오후 1:55
ㅎㅏ.....
오후 1:55

1 과제 개요 및 목표



해결하고자 하는 문제

- 한국어 어문 규범 관련 질문에 대한 적절한 답변

만들고자 하는 것

- 국립국어원 한국어 규범 문서를 **근거**로 활용하여 신뢰도 높은 답변을 생성

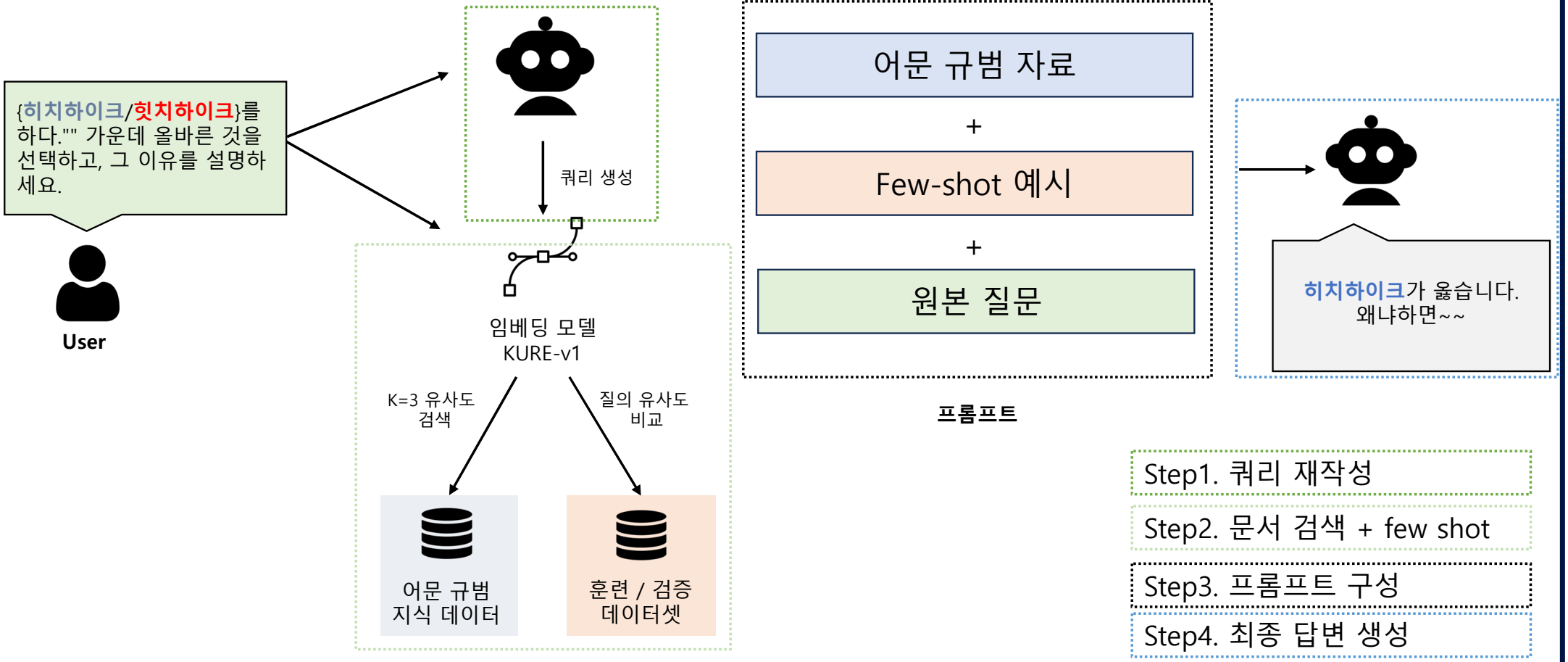
핵심 전략

- **검색과 생성을 결합한 RAG 파이프라인**을 통해 문제 해결

최종 결과

- "한국어 어문 규범 기반 생성"에서 **가장 높은 성능 70.57 달성**

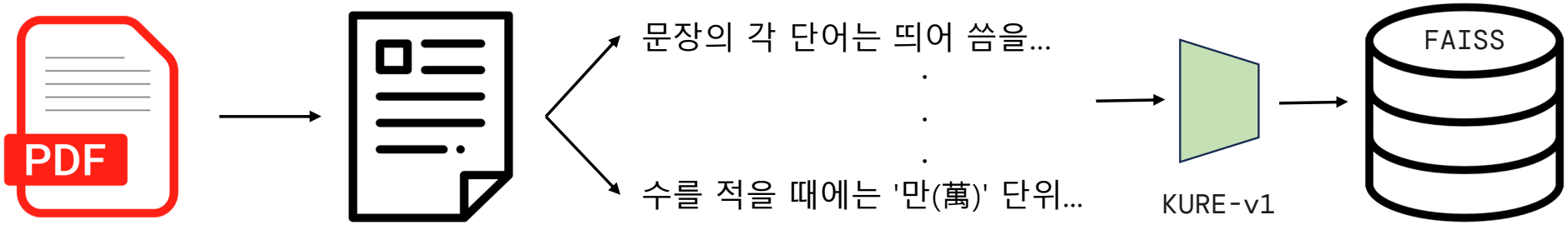
2 전체 시스템 아키텍처


{히치하이크/힛치하이크}를 하다."" 가운데 올바른 것을 선택하고, 그 이유를 설명하세요.
User
쿼리 생성
임베딩 모델
KURE-v1
K=3 유사도 검색
질의 유사도 비교
어문 규범 지식 데이터
훈련 / 검증 데이터셋
어문 규범 자료
+
Few-shot 예시
+
원본 질문
프롬프트
히치하이크가 옳습니다. 왜냐하면~~
Step1. 쿼리 재작성
Step2. 문서 검색 + few shot
Step3. 프롬프트 구성
Step4. 최종 답변 생성

3 데이터 분석 및 전처리 전략

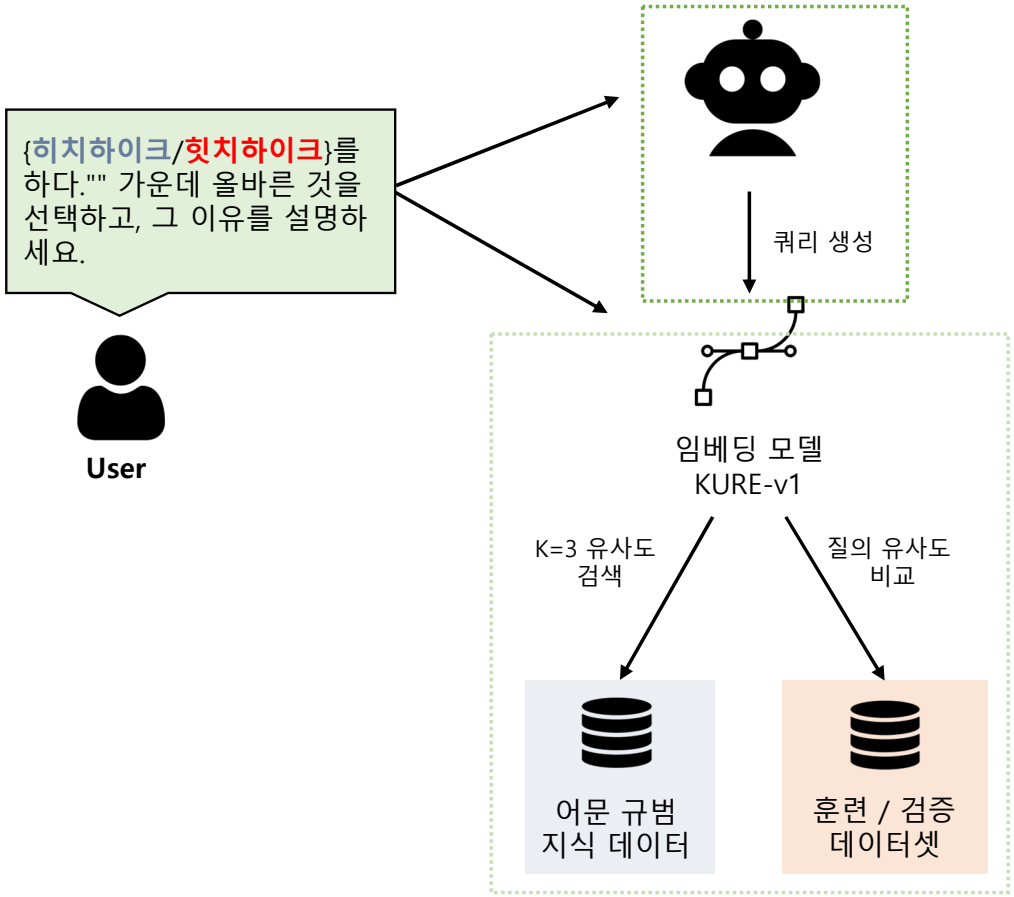
순위	오류 유형	비율	문항 수	세부 내용
1	띄어쓰기·문장 부호	**34%**	212개	조사 및 어미 연결, 복합어 띄어쓰기, 문장 부호 활용
2	외래어·외국어 표기	**29%**	180개	외래어 표기법 준수, 외국 고유명사 표기
3	음운 변동·두음 법칙	14%	87개	ㄴ, ㄹ 두음 법칙, 음성 변화 규칙
4	용언 활용	12%	75개	어간·어미 결합, 불규칙 활용
5	어휘 선택	7%	44개	동음이의어 구별, 의미 차이 인식
6	기타	4%	24개	복합 규칙 적용, 특수 사례

3 데이터 분석 및 전처리 전략



- **PDF -> TXT**: PyMuPDF로 PDF 문서에서 1,287줄의 텍스트 추출 및 정제
- **의미론적 청킹**: RecursiveCharacterTextSplitter로 250자 단위로 분할, 25자 중첩 (총 11,400개 청크)
- **벡터화 및 인덱싱**: 한국어 특화 모델(KURE-v1)로 768차원 벡터 변환 후 FAISS에 인덱싱

4 핵심 전략 1: Query expansion



문제점: 사용자의 자연어 질문은 검색에 비효율적

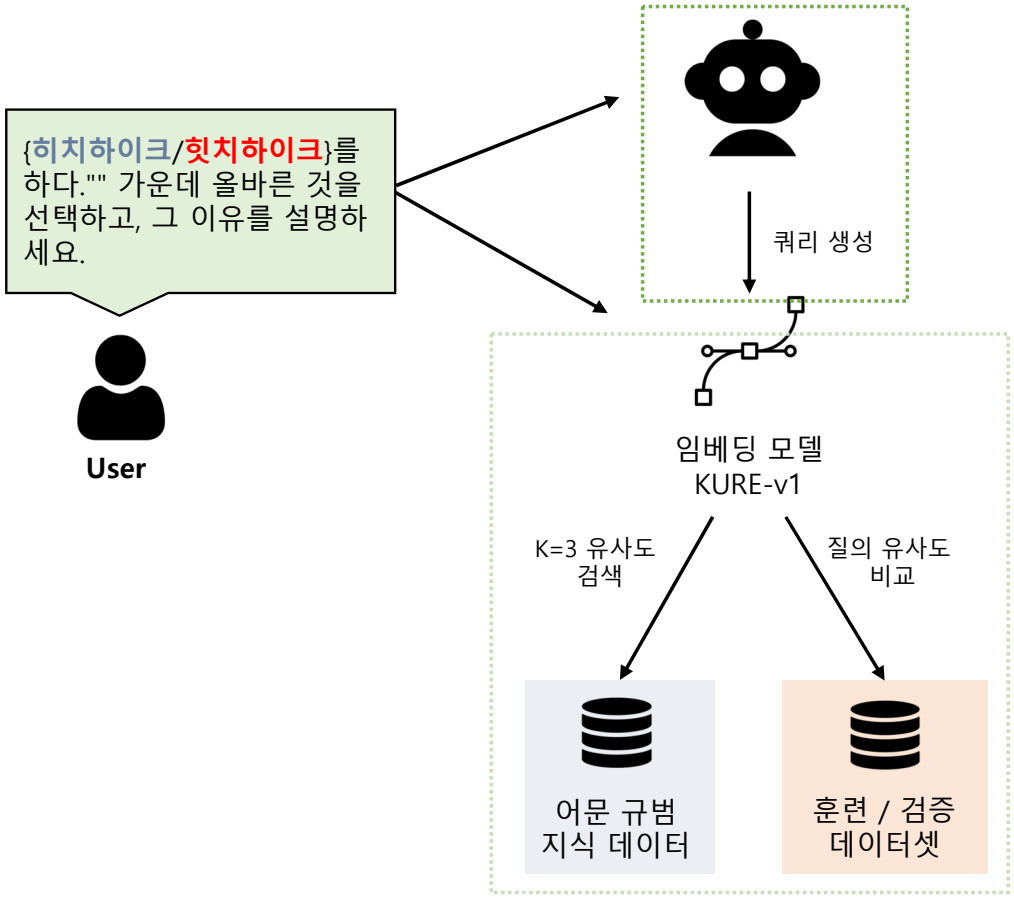
해결책: LLM을 활용한 Query expansion

작동 방식:

- **원본 질문**: "가축을 기를 때에는 {먹이량/먹이양}을 조절해 주어야 한다."
- **지시 사항**: ...하나의 검색용 쿼리들을 생성하는 것입니다...
- **생성된 쿼리**: "의존명사 양 량 두음법칙"

- 생성된 쿼리를 이용한 Retrieval

4 핵심 전략 2: Dynamic Few-shot



문제점: 잘못된 검색과 고정된 Few-shot 예제는 강건함이 부족함

해결책: 질문의 의도를 파악해 최적의 예제를 동적으로 선택

작동 방식:

- **입력 질문**: {히치하이크/힛치하이크}를 하다.
- **생성된 쿼리**: "외래어 표기법 준수"

- **Few-shot 질문**: {에스커레이터/에스컬레이터} 손잡이를 잡아야 한다.
- **생성된 쿼리**: "외래어 표기법 관련 오류 사항"

5 학습 데이터셋 구성

어문 규범 자료

+

Few-shot 예시

+

원본 질문

다음은 어문 규범에 대한 자료입니다.

...
- flash[flæʃ] 플래시, shrub[ʃrʌb] 슈러브, shark[ʃɑːk] 샤크, shank[ʃæŋk] 섕크,fashion[fæʃən] 패션, sheriff[ʃerif] 셰리프, shopping[ʃɔpiŋ] 쇼핑, shoe[ʃuː] 슈, shim[ʃim]
심
3. 어말 또는 자음 앞의 [ʒ]는 '지'로 적고, 모음 앞의 [ʒ]는 'ㅈ'으로 적는다.
- mirage[mirɑːʒ] 미라지, vision[viʒən] 비전
...

아래는 질문과 답변의 예시입니다. 이 예시의 형식에 맞춰 질문에 답해주세요.

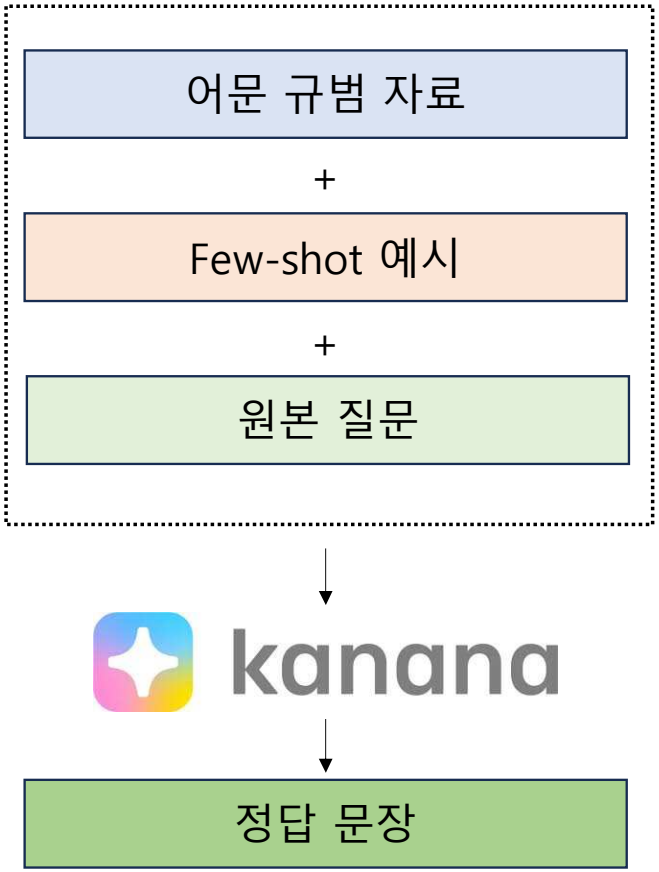
...
question:"{풀루트/플루트} 소리가 좋다." 가운데 올바른 것을 선택하고, 그 이유를 설명하세요.
answer:"플루트 소리가 좋다."가 옳다. 플루트(flute)의 원어 발음은 [flut]로 [f]는 자음 앞에서 '프'로 표기한다. 따라서 '플루트'로 표기하는 것이 옳다.
...

위 'answer:' 뒤의 출력 형식을 엄격히 준수하여 답변을 생성하십시오.

이제 다음 질문에 답변하세요.
question:"{프레시/프레쉬}한 음식을 먹고 싶습니다." 가운데 올바른 것을 선택하고, 그 이유를 설명하세요.

Query expansion

6 모델 학습 및 최적화



베이스 모델

- **Kanana-1.5-8B-Instruct**(한국어 어문 및 명령어 수행 능력 최우수)

학습 전략: **LoRA** 기반 파인튜닝

- 전체 파라미터의 0.8%만 학습하여 과적합 방지 및 VRAM 사용량 최소화
- r=32, alpha=64 등 최적의 하이퍼파라미터 조합 사용

학습 효율화: **DataCollatorForCompletionOnlyLM**

- 질문 부분은 제외하고 **답변에만 loss를 계산**하여 모델이 '답변 생성'에만 집중하도록 훈련

7 Ablation study

전체 60팀

순위	팀명	모델명	평가 점수
1	overfit-brothers Pro Max	10	70.5733073
		45	70.5330422
		103	70.1302871
		41	70.0977738
		60	70.0881316
		51	70.0732528
		31	69.9489721
		58	69.9405839
		42	69.8879356

다양한 모델 테스트

- **Kanana-1.5-8B-Instruct**
- Midm-2.0
- Qwen
- Exaone
- A.X…

다양한 학습 기법

- 강화학습
- **Lora**, Q-Lora, Full-tuning
- 모델 병합 (**SLERP**, Linear, TIES…)
- **DataCollatorForCompletionOnlyLM**

학습 데이터 최적화

- 다양한 임베딩 모델(**KURE-v1**, Qwen-embedding, bge-m3-korean…)
- 문장 청크와 중첩 크기(**250, 25**)
- 검색 문장과 Few-shot 개수(**각 3개씩**)

8 최종 결과

전체 60팀

※ 마지막 평가 일시 : 2025년 08월 01일 00시 06분

순위	팀명	모델명	평가 점수	한국어 어문 규범 기반 생성(RAG)(가 유형) (평균)				제출 일시
				Exact Match	BLEURT	BERTScore	ROUGE-1	
1	overfit-brothers Pro Max	10	70.5733073	66.8494258 (bleurt, bertscore, ROUGE-1 의 평균)				2025.07.19. 22:16
				74.2971888	63.7067991	82.8148007	54.0266776	
2	심플리 러블리	천 리 길도 스물여섯 걸음부터	69.6511129	65.2058402 (bleurt, bertscore, ROUGE-1 의 평균)				2025.07.27. 10:04
				74.0963855	61.8564570	81.6914852	52.0695784	
3	Real Awesome Gyubeom	베이스라인찾기22	69.2785804	64.8623816 (bleurt, bertscore, ROUGE-1 의 평균)				2025.07.30. 08:45
				73.6947791	61.4623370	81.5124649	51.6123430	
4	최강인공지능	Model02_18	68.6583414	63.6219038 (bleurt, bertscore, ROUGE-1 의 평균)				2025.07.14. 20:46
				73.6947791	60.9829136	80.9055171	48.9772806	
5	AnIma	f_1	68.5236496	61.5452911 (bleurt, bertscore, ROUGE-1 의 평균)				2025.07.31. 21:40
				75.5020080	59.0623124	80.0183458	45.5552153	

학습 및 추론 코드: https://github.com/overfit-brothers/KRAG
모델 링크: https://huggingface.co/overfit-brothers/KRAG-SOTA
리더보드 결과: https://zrr.kr/Qu3lvv

Q&A