
KRX LLM COMPETITION

MODEL CARD

Contents

1	Dataset	2
1.1	구성 및 라이선스	2
1.2	데이터 가공	2
1.3	Validation Set	3
2	MCQA	3
3	QA	4
4	Training	5
4.1	Curriculum SFT	5
4.2	DPO	6
5	Conclusion	7
5.1	Strength	7
5.2	Limitation	7
5.3	Future improvements	7
	References	8
A	Dataset Card	10
A.1	References	10
A.2	MCQA	11
A.3	QA	11
B	ModelCard	12

Abstract

shibainu24 모델은 Qwen-2.5-7B-Instruct¹(Yang et al., 2024)를 기반으로 GPT 모델로부터 학습 데이터를 생성해 SFT, DPO(Rafailov et al., 2024) 학습된 모델입니다. 모델은 Orca(Mukherjee et al., 2023) 학습 방식을 따라 Reasoning을 중점적으로 학습되었으며, MCQA와 QA 각각의 방식에 따라 데이터셋이 구축 및 학습이 진행되었습니다.

MCQA의 경우 WizardLM(Xu et al., 2023)의 방식을 기반으로 조금 더 높은 난이도의 학습 데이터셋을 단계적으로 생성합니다. 원천 데이터가 질문과 답변으로 구성된 경우, RFT(Yuan et al., 2023)를 통해 데이터셋을 필터링하여 데이터의 질을 높입니다. MCQA의 경우 선택지를 고르는 것만으로 모델이 지식을 배우는 것이 한계가 있기 때문에 별도의 Reasoning Process를 생성하여 학습에 추가합니다.

다음으로 QA의 경우 용어 사전 등 참조 지식을 기반으로 질문과 첫번째 답변을 생성합니다. 그리고 생성한 질문에 대해 참조 지식 없이 두번째 답변을 생성하며, GPT 모델에게 첫번째와 두번째 답변에 대한 Preference를 평가받습니다.

MCQA와 QA 모두 생성한 데이터들은 fineweb-edu(Penedo et al., 2024a)를 기반으로 교육적 가치를 평가하며 적절한 기준값으로 필터링하게 됩니다. 그리고 마지막으로, 생성된 질문과 답변을 GPT 모델로부터 hallucination 체크를 받아 hallucination 데이터를 제거합니다. SFT 학습이 완료된 후 QA의 2가지 답변과 평가받은 Preference를 기반으로 DPO(Rafailov et al., 2024) 학습을 수행합니다.

복잡한 데이터 생성 과정과 재현성을 해결하기 위해, 데이터 생성 파이프라인은 LangGraph²를 기반으로 자동화되어 있습니다. 이를 통해 GPT 모델의 오류를 해결될 때까지 반복적으로 수행하며, 복잡한 생성 과정과 데이터 후처리 과정을 자동으로 수행하기 때문에 데이터 생성 비용을 줄일 수 있습니다.

1 Dataset

1.1 구성 및 라이선스

학습에 사용된 데이터셋, 구성, 출처, 라이선스, 토큰 수 등의 세부적인 데이터셋의 정보들은 Appendix A Dataset Card에 기재되어 있습니다.

데이터셋의 라이선스 정책을 지키기 위해 다음 사항에 부합하는 데이터셋을 사용합니다.

- Apache License 2.0, MIT License 데이터
- 공공데이터법 기준에 부합하는 허가된 데이터
- 저작권 정책에 따라 비상업적 이용이 자유로운 데이터
- 개별적으로 담당자에게 연락 후 사용이 허가된 데이터

1.2 데이터 가공

학습용 데이터를 생성하기 위해 데이터 원천에 따라 다음 전처리 및 후처리 과정을 거칩니다. 데이터 가공 소스 코드들은 github³에 작성되어 있으며, Huggingface의 Datatrove(Penedo et al., 2024b)를 모티브로 작성되었습니다.

¹<https://huggingface.co/Qwen/Qwen2.5-7B-Instruct>

²<https://www.langchain.com/langgraph>

³<https://github.com/aiqwe/krx-llm-competition>

- Deduplication: 중복 제거
- Quality Filtering
 - 텍스트 내 일정 비율 이상의 특정 영어 포함시 필터링(한글 텍스트 사용시)
 - 텍스트 내 일정 비율 이상의 Special Character(Punctuation) 포함시 필터링
 - 텍스트 내 일정 비율 이상의 숫자 포함시 필터링
 - 텍스트 공백 제거
 - 한글, 영어, 숫자, 일부 Punctuation을 제외한 유니코드 제거
 - 특정 토큰 수 미만시 필터링

토큰 수 계산은 tiktoken⁴의 o200k_base를 기반으로 계산합니다.

1.3 Validation Set

학습된 모델의 Checkpoint 중 가장 우수한 Checkpoint를 찾기 위해 Validation Set을 작성합니다. MCQA 평가 사항인 재무제표 분석, 회계 원칙 적용, 재무 비율 계산, 금융 개념, 시장 동향, 투자 전략 항목들의 참조 정보들을 균일한 비율로 검색합니다. 그리고 금융, 회계 지식을 활용하여 참조 정보들의 정확성, 난이도를 평가합니다. 여기서 좋은 평가를 받은 참조 정보들을 활용하여 o1-preview를 사용하고 Validation Set 초안을 작성합니다. o1-preview로 생성된 데이터들도 다시한번 데이터의 질을 평가하고 보완한 뒤, Validation set으로 완성합니다. 이 때 정답의 분포와 평가 항목의 분포를 균일하게 작성하여, 모델들이 특정 평가항목에 편향되는 평가를 받지 않도록 합니다. 완성된 Validation set은 huggingface⁵ 또는 github를 통해 제공합니다.

2 MCQA

데이터셋은 Graph로 구성된 데이터 파이프라인을 따라 자동으로 생성합니다. 사용자는 참조 지식을 그래프에 제공하고 사전에 정의된 단계에 따라 데이터를 생성합니다.

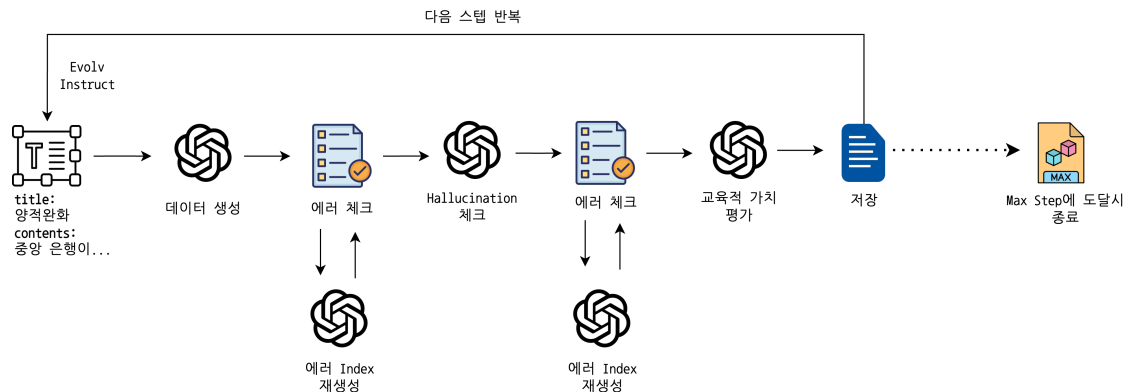


Figure 1: MCQA Graph Pipeline

⁴<https://github.com/openai/tiktoken>

⁵<https://huggingface.co/datasets/aigwe/krx-llm-competition>

MCQA 데이터셋은 아래 단계로 노드와 엣지를 따라 생성됩니다.

1. 데이터셋 생성: 사용자가 제공한 참조 지식을 기반으로 질문과 답변 데이터셋 생성
2. 에러 체크: 생성된 데이터에서 선택지, 답변 등의 포맷에 오류가 있는지 확인
3. 에러 index 재생성: 잘못된 데이터만 분류하여 재생성
4. Hallucination 체크: 사용자의 참조 지식을 제외한 상태로, 생성한 질문만을 대상으로 답변을 위한 정보가 부족한지 확인하며, 또한 답변이 hallucination인지 확인
5. 교육적 가치 평가: 생성된 데이터가 fineweb-edu(Penedo et al., 2024a)의 방식을 차용하여 교육적 가치를 평가합니다.
6. 다음 스텝 반복: WizardLM(Xu et al., 2023)을 참조하여 여러 단계로 다양한 데이터셋을 생성하도록 합니다. 이전 단계에서 생성한 데이터셋을 다시 프롬프트로 입력하여 좀더 높은 난이도의 데이터를 생성하도록 GPT 모델에게 요구합니다. 이후 단계는 이전 Step과 동일한 절차를 반복합니다.
7. Max Step 확인: 사용자가 지정한 Step에 도달하면 다음 단계인 hallucination 체크로 진행합니다.
8. 저장: 최종 데이터셋을 jsonline 형태로 저장

사용자는 참조 지식을 `title`과 `contents`로 작성하여 그래프에 입력 데이터로 제공합니다.

Orca Prompt and Reasoning MCQA 데이터셋은 선택형 문제이기 때문에 모델이 데이터로부터 지식을 제한적으로 학습할 가능성이 높습니다. 따라서 Orca(Mukherjee et al., 2023) 프롬프트 방식에 따라 Reasoning Process를 생성하여 이를 모델이 추가적으로 학습합니다. 반복 실험을 통해 Orca 프롬프트 중 필요한 것만 선택, 요약, 변형하여 데이터셋을 생성하는데 사용합니다.

RFT Filtering 참조할 데이터셋이 지식 형태가 아닌 질문과 답변 형태라면, RFT(Yuan et al., 2023) 방식으로 데이터를 필터링하여 사용합니다. 참조할 데이터의 질문을 GPT 모델에게 전달하고 실제 정답과 GPT의 답변이 일치하는 데이터만 필터링한 뒤, 가공이 필요하다면 그래프에 투입하여 데이터셋을 생성합니다.

3 QA

QA의 경우, 여러 Step에서 질문과 답변을 생성하는 MCQA와 달리 1개의 질문과 2개의 답변을 생성합니다. 그리고 Preference를 평가하여 DPO(Rafailov et al., 2024) 학습에 사용합니다.

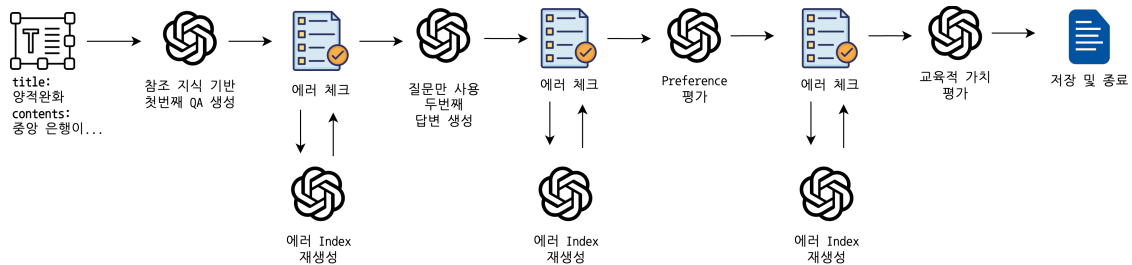


Figure 2: MCQA Graph Pipeline

1. 첫번째 QA 생성: 사용자가 제공한 참조 지식을 기반으로 첫번째 질문과 답변 데이터셋 생성
2. 에러 체크: 생성된 데이터에서 선택지, 답변 등의 포맷에 오류가 있는지 확인
3. 에러 index 재생성: 잘못된 데이터만 분류하여 재생성
4. 두번째 답변 생성: 첫번째 QA 생성에서 생성한 질문만을 사용하여, GPT의 지식과 Reasoning을 최대한 발휘하도록 하는 두번째 답변을 생성합니다.
5. Preference 평가: 첫번째 답변과 두번째 답변에 대해 GPT 모델이 Preference를 평가
6. Educational Value Check: 생성된 데이터가 fineweb-edu(Penedo et al., 2024a)의 방식을 차용하여 교육적 가치를 평가합니다.
7. 저장: 최종 데이터셋을 jsonline 형태로 저장

QA도 MCQA와 동일하게 사용자는 참조 지식을 `title`과 `contents`로 작성하여 그래프에 입력 데이터로 제공합니다.

Preference Evaluation 사용자가 참조 지식을 제공하여 첫번째 질문과 답변 셋이 생성한 다음, 두번째 답변을 생성할 때는 참조 지식을 제거하고 첫번째 Step에서 생성한 질문만을 제공합니다. GPT 모델이 답변을 생성할 때, 참조 지식이 제공되면 참조 지식을 기반으로 답변을 생성하기 때문에 다양성이 줄어들게 됩니다. 따라서 두번째 답변은 GPT 모델이 학습한 지식을 최대한 사용하여 좀더 다양한 답변을 생성할 수 있도록 참조 지식을 제거합니다.

그리고 첫번째 답변과 두번째 답변을 Judging LLM-as-a-Judge(Zheng et al., 2023)방식으로 Preference를 평가합니다. Preference로 선택된 답변을 바탕으로 SFT학습을 수행한 뒤, 두가지 답변을 바탕으로 DPO 학습을 수행합니다.

4 Training

trl⁶을 사용하여 SFT와 DPO학습을 수행합니다. 자세한 학습에 대한 정보는 ModelCardB과 github⁷에 작성되어 있습니다.

SFT 단계에서는 MCQA와 QA(chosen)를 random shuffle 하여 학습합니다. 8e-6의 cosine learning rate decay를 적용하였고, residual dropout 0.2를 적용하였습니다. 총 16 epochs을 학습하였고 validation set을 통해 best checkpoint를 선정하였습니다. SFT의 하이퍼파라미터는 Instruct-GPT(Ouyang et al., 2022)에서 사용한 값들을 그대로 사용하였습니다.

4.1 Curriculum SFT

SFT 학습은 단순히 총 2 step으로 나누어 진행하였습니다.

1. (step 1) vanilla model(qwen2.5-7b-instruct)위에 난이도가 가장 낮은 데이터(easy set)를 SFT로 학습합니다.

⁶<https://github.com/huggingface/trl>

⁷<https://github.com/aiqwe/krx-llm-competition>

2. (step 2) step 1으로 생성된 SFT checkpoint위에 step1에서 easy set과 그 보다 더 어려운 hard set을 합하여 SFT로 학습합니다.

학습 데이터셋은 WizardLM(Xu et al., 2023)기반의 evolve-instruct 방식으로 제작 했기에 문제의 난이도를 나눌수 있었고, 이를 토대로 결과적으로 효과적인 curriculum learning을 수행할 수 있었습니다.

4.2 DPO

Longform QA의 성능을 높이기위해 Curriculum SFT로 도출된 checkpoint위에 DPO를 수행하였습니다. QA 데이터 제작시 사용된 두개의 응답중 LLM-as-a-Judge를 통해 판별된 chosen, rejected response를 pair 데이터로 사용하였습니다. 반복적인 하이퍼파라미터 서치를 통해 DPO β 는 0.05로 설정하였습니다.

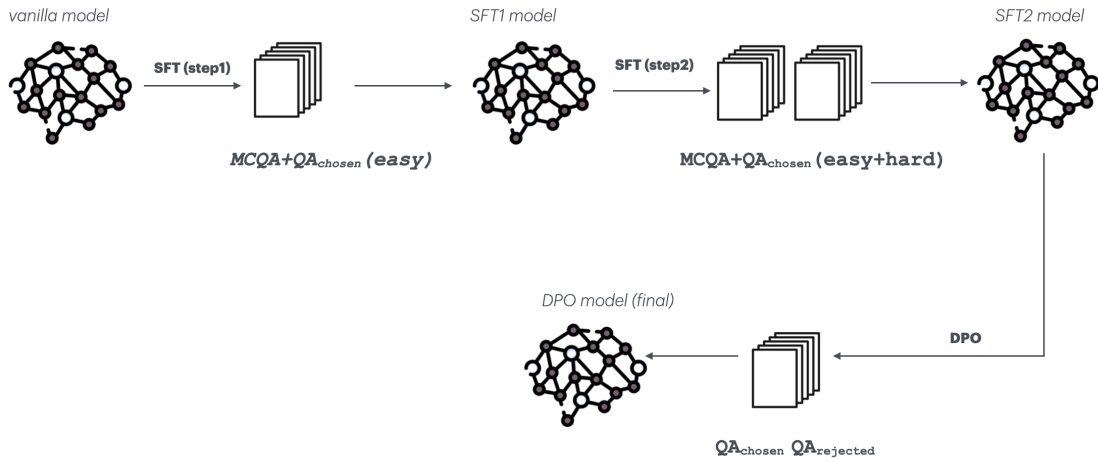


Figure 3: Training Pipeline

5 Conclusion

5.1 Strength

Reasoning 본 모델은 OrcaMukherjee et al. (2023)에 기반하여 Reasoning을 중점적으로 학습하였기 때문에 텍스트 생성과 Reasoning에 강점을 보입니다. 특히 회계 원칙과 금융 개념을 중점적으로 학습하였기 때문에 일반인이 어려워할 수 있는 전문적인 지식에 잘 답변할 수 있습니다.

Educational Value fineweb-edu(Penedo et al., 2024a)를 통해 데이터의 질을 관리하는 방법을 사용했기 때문에, 적은 데이터로도 좋은 평가를 받을 수 있습니다. 11월 26일부터 교육적 가치를 평가하는 방법을 도입하였고 데이터의 규모는 절반 수준으로 떨어뜨렸으나, 오히려 MCQA 평가 점수는 38.5점에서 41점으로 상승하는 좋은 결과를 얻었습니다. 데이터 규모를 줄이면서 한정된 Computational Power로 좀더 많은 실험을 할 수 있었고, 이후의 학습에서 이는 큰 강점이 되었습니다.

5.2 Limitation

전문 지식 학습의 어려움 한국 제도를 따르는 규정이나 금융 지식들은 데이터셋의 양이 적습니다. 또한 전문적인 지식과 정보를 갖춘 데이터들은 보통 PDF로 존재하며 저작권의 보호를 받기 때문에 학습시키기가 어렵습니다. 이러한 점을 극복하기 위해 본 모델은 PDF 데이터 전처리 기능을 강화하고 다양한 공공기관의 간행물을 사용하여 최대한 질 높은 데이터를 학습시켰습니다.

변화하는 데이터의 학습의 어려움 산업 동향과 같은 데이터는 시기에 따라 변동되며, 데이터 산출 기관에 따라 수치가 달라질 수 있습니다. 이에 따라 모델이 잘못된 수치를 학습할 위험이 있습니다. 따라서 본 모델은 시간에 따라 변화하거나 산출 기관에 따라 달라지는 수치가 있는 데이터들은 제외시켰습니다. 이러한 평가에서 모델이 낮은 점수를 기록할 수는 있지만 학습 후 RAG(Lewis et al., 2021)를 통해 실제 서비스에선 좋은 성능을 발휘할 수 있습니다.

주관적인 데이터 학습의 어려움 투자 전략과 같은 데이터는 차트나 패턴에 따라 주관적인 해석이 포함되기에 모델의 성능을 떨어뜨릴 수 있습니다. 이러한 위험을 피하기 위해 여러 단계의 필터링을 거쳤으며, 되도록 객관적인 지식을 포함한 데이터만 학습하였습니다. 모델은 최대한 원리와 이론적인 답변만 제공하고 실제 사용자들이 주관적인 판단을 하도록 해야 할 것으로 생각됩니다.

5.3 Future improvements

RAG 금융 모델의 목적은 사용자에게 전문적인 금융 지식 뿐만 아니라 현재의 투자 전략, 금융 상황, 시황 등의 정보를 제공할 가능성이 높습니다. LLM 모델은 Online으로 학습하는 비용이 크거나 불가능하기 때문에 RAG(Lewis et al., 2021)를 사용하여 현재 데이터를 잘 답변을 할 수 있도록 하면 성능을 높일 수 있습니다.

Accounting Math 회계 기준은 각 국가의 상황에 맞게 변형하여 사용하기 때문에 국내 회계 기준과 맞는 회계 계산 데이터가 필요합니다. 따라서 국내 회계 기준에 맞는 회계 계산 데이터는 확보하기가 매우 어렵습니다. 추후에는 이러한 데이터들을 많이 생산할 수 있는 환경이 조성되어야 할 것으로 생각됩니다.

PDF 공공기관은 대부분 PDF를 사용하여 문서를 배포합니다. 공공기관의 데이터는 질이 굉장히 좋은 편인데, OCR로 가공되지 않으면 학습에 사용하기가 어렵습니다. 따라서 PDF 데이터를 자연어로 처리할 수 있도록 OCR 기반의 PDF를 사용할 수 있는 환경이 조성되는 것이 LLM을 학습하는 데에 도움이 될 것으로 생각됩니다.

References

- Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. Retrieval-augmented generation for knowledge-intensive nlp tasks, 2021. URL <https://arxiv.org/abs/2005.11401>.
- Subhabrata Mukherjee, Arindam Mitra, Ganesh Jawahar, Sahaj Agarwal, Hamid Palangi, and Ahmed Awadallah. Orca: Progressive learning from complex explanation traces of gpt-4, 2023. URL <https://arxiv.org/abs/2306.02707>.
- Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and Ryan Lowe. Training language models to follow instructions with human feedback, 2022. URL <https://arxiv.org/abs/2203.02155>.
- Guilherme Penedo, Hynek Kydlíček, Loubna Ben allal, Anton Lozhkov, Margaret Mitchell, Colin Raffel, Leandro Von Werra, and Thomas Wolf. The fineweb datasets: Decanting the web for the finest text data at scale, 2024a. URL <https://arxiv.org/abs/2406.17557>.
- Guilherme Penedo, Hynek Kydlíček, Alessandro Cappelli, Mario Sasko, and Thomas Wolf. Datatrove: large scale data processing, 2024b. URL <https://github.com/huggingface/datatrove>.
- Rafael Rafailov, Archit Sharma, Eric Mitchell, Stefano Ermon, Christopher D. Manning, and Chelsea Finn. Direct preference optimization: Your language model is secretly a reward model, 2024. URL <https://arxiv.org/abs/2305.18290>.
- Can Xu, Qingfeng Sun, Kai Zheng, Xiubo Geng, Pu Zhao, Jiazhan Feng, Chongyang Tao, and Daxin Jiang. Wizardlm: Empowering large language models to follow complex instructions, 2023. URL <https://arxiv.org/abs/2304.12244>.
- An Yang, Baosong Yang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Zhou, Chengpeng Li, Chengyuan Li, Dayiheng Liu, Fei Huang, Guanting Dong, Haoran Wei, Huan Lin, Jialong Tang, Jialin Wang, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Ma, Jianxin Yang, Jin Xu, Jingren Zhou, Jinze Bai, Jinzheng He, Junyang Lin, Kai Dang, Keming Lu, Keqin Chen, Kexin Yang, Mei Li, Mingfeng Xue, Na Ni, Pei Zhang, Peng Wang, Ru Peng, Rui Men, Ruize Gao, Runji Lin, Shijie Wang, Shuai Bai, Sinan Tan, Tianhang Zhu, Tianhao Li, Tianyu Liu, Wenbin Ge, Xiaodong Deng, Xiaohuan Zhou, Xingzhang Ren, Xinyu Zhang, Xipin Wei, Xuancheng Ren, Xuejing Liu, Yang Fan, Yang Yao, Yichang Zhang, Yu Wan, Yunfei Chu, Yuqiong Liu, Zeyu Cui, Zhenru Zhang, Zhifang Guo, and Zhihao Fan. Qwen2 technical report, 2024. URL <https://arxiv.org/abs/2407.10671>.
- Zheng Yuan, Hongyi Yuan, Chengpeng Li, Guanting Dong, Keming Lu, Chuanqi Tan, Chang Zhou, and Jingren Zhou. Scaling relationship on learning mathematical reasoning with large language models, 2023. URL <https://arxiv.org/abs/2308.01825>.

Lianmin Zheng, Wei-Lin Chiang, Ying Sheng, Siyuan Zhuang, Zhanghao Wu, Yonghao Zhuang, Zi Lin, Zhuohan Li, Dacheng Li, Eric P. Xing, Hao Zhang, Joseph E. Gonzalez, and Ion Stoica. Judging llm-as-a-judge with mt-bench and chatbot arena, 2023. URL <https://arxiv.org/abs/2306.05685>.

A Dataset Card

데이터셋의 토큰 수는 tiktoken의 o200k_base 기준이며, 데이터셋은 Huggingface⁸와 Github⁹로 제공합니다.

A.1 References

데이터명	url	license
한국은행 경제금융 용어 700선	https://www.bok.or.kr/portal/bbs/B0000249/view.do?nttId=235017&menuNo=200765	사용 가능
재무회계 합성 데이터	-	자체 생성
금융감독용어사전	https://terms.naver.com/list.naver?cid=42088&categoryId=42088	사용 가능
web_text_synthetic_dataset_50k	https://huggingface.co/datasets/Cartinoe5930/web_text_synthetic_dataset_50k	사용 가능
지식경제용어사전	https://terms.naver.com/list.naver?cid=43668&categoryId=43668	공공데이터
한국거래소 비정기 간행물	http://open.krx.co.kr/contents/OPN/04/04020000/OPN04020000.jsp#b8943a5f87282cde0d653d1ae73431c9=1	KRX데이터
한국거래소규정	https://law.krx.co.kr/las/TopFrame.jsp&KRX	
초보투자자 증권따라잡기	https://main.krxverse.co.kr/_contents/ACA/02/02010200/file/220104_beginner.pdf	KRX데이터
청소년을위한 증권투자	https://main.krxverse.co.kr/_contents/ACA/02/02010200/file/220104_teen.pdf	KRX데이터
기업사업보고서 공시자료	https://opendart.fss.or.kr/	공공데이터
시사경제용어사전	https://terms.naver.com/list.naver?cid=43665&categoryId=43665	공공데이터

⁸<https://huggingface.co/datasets/aiqwe/krx-llm-competition>

⁹<https://github.com/aiqwe/krx-llm-competition>

A.2 MCQA

학습에 사용된 MCQA 데이터셋의 수는 약 4.6만개이며, 약 2천만개의 토큰으로 구성되어 있습니다.

데이터명	데이터 수	토큰 수
한국은행 경제금융 용어 700선	1,203	277,114
재무회계 목차를 이용한 합성데이터	451	99,770
금융감독용어사전	827	214,297
hf_web_text_synthetic_dataset_50k	25,461	7,563,529
지식경제용어사전	2,314	589,763
한국거래소 비정기 간행물	1,183	230,148
한국거래소규정	3,015	580,556
초보투자자 증권따라잡기	599	116,472
청소년을 위한 증권 투자	408	77,037
기업사업보고서 공시자료	3,574	629,807
시사경제용어사전	7,410	1,545,842
합계	46,445	19,998,931

A.3 QA

QA 데이터셋은 2가지 답변중 SFT 학습에 사용된 답변을 기준으로 산출하였습니다. QA 학습에 사용된 데이터 수는 약 4.8만개이며, 약 2억개의 토큰으로 구성되어 있습니다.

데이터명	데이터 수	토큰 수
한국은행 경제금융 용어 700선	1,023	846,970
금융감독용어사전	4,128	3,181,831
지식경제용어사전	6,526	5,311,890
한국거래소 비정기 간행물	1,510	1,089,342
한국거래소규정	4,858	3,587,059
기업사업보고서 공시자료	3,574	629,807
시사경제용어사전	29,920	5,981,839
합계	47,965	199,998,931

B ModelCard

Contents	Spec
Base model	Qwen2.5-7B-Instruct
dtype	bfloat16
PEFT	LoRA (r=8, alpha=64)
Learning Rate	1e-5 (varies by further training)
LRScheduler	Cosine (warm-up: 0.05%)
Optimizer	AdamW
Distributed / Efficient Tuning	DeepSpeed v3, Flash Attention