

Llama 2 Fine-Tuning을 위한 KorQuad 데이터셋 소개

에이아이스쿨(AISchool) 대표
양진호 (솔라리스)

<http://aischool.ai>

<http://solarisailab.com>

SQuAD v1.1 Dataset

- The **S**tanford **Q**uestion **A**nswering **D**ataset (SQuAD v1.1) 는 100K(10만개)의 Question & Answer 쌍으로 구성된 데이터셋입니다.
- Question과 Answer를 위한 문맥 Paragraph가 주어지고 해당 Paragraph에 있는 정보를 통대로 Answer를 맞추는 형태의 데이터셋입니다.
- <https://rajpurkar.github.io/SQuAD-explorer/>

SQuAD v1.1 Dataset

SQuAD

[Home](#)

[Explore 2.0](#)

[Explore 1.1](#)

Super_Bowl_50

The Stanford Question Answering Dataset

Super Bowl 50 was an American football game to determine the champion of the National Football League (NFL) for the 2015 season. The American Football Conference (AFC) champion Denver Broncos defeated the National Football Conference (NFC) champion Carolina Panthers 24–10 to earn their third Super Bowl title. The game was played on February 7, 2016, at Levi's Stadium in the San Francisco Bay Area at Santa Clara, California. As this was the 50th Super Bowl, the league emphasized the "golden anniversary" with various gold-themed initiatives, as well as temporarily suspending the tradition of naming each Super Bowl game with Roman numerals (under which the game would have been known as "Super Bowl L"), so that the logo could prominently feature the Arabic numerals 50.

Which NFL team represented the AFC at Super Bowl 50?

Ground Truth Answers: Denver Broncos Denver Broncos Denver Broncos

Prediction: Denver Broncos

Which NFL team represented the NFC at Super Bowl 50?

Ground Truth Answers: Carolina Panthers Carolina Panthers Carolina Panthers

Prediction: Carolina Panthers

Where did Super Bowl 50 take place?

Ground Truth Answers: Santa Clara, California Levi's Stadium Levi's Stadium in the San Francisco Bay Area at Santa Clara, California.

Prediction: Levi's Stadium in the San Francisco Bay Area at Santa Clara, California

KorQuad v1.0 Dataset

- Korean Question Answering Dataset (KorQuad v1.0) 는 SQuAD 데이터셋의 한국어 버전입니다.
- KorQuAD 데이터셋은 LG CNS에서 구축하였습니다.
- KorQuAD 1.0의 전체 데이터는 1,560 개의 Wikipedia article에 대해 10,645 건의 문단과 66,181 개의 질의응답 쌍으로, Training set 60,407 개, Dev set 5,774 개의 질의응답쌍으로 구성되어 있습니다.
- Question과 Answer를 위한 문맥 Paragraph가 주어지고 해당 Paragraph에 있는 정보를 통대로 Answer를 맞추는 형태의 데이터셋입니다.
- <https://korquad.github.io/KorQuad%201.0/>

KorQuad v1.0 Dataset Example 1

- **Context** : 한편 1840년부터 바그너와 알고 지내던 리스트가 잊혀져 있던 1악장을 부활시켜 1852년에 바이마르에서 연주했다. 이것을 계기로 바그너도 이 작품에 다시 관심을 갖게 되었고, 그 해 9월에는 총보의 반환을 요구하여 이를 서곡으로 간추린 다음 수정을 했고 브라이트코프흐 & 헤르텔 출판사에서 출판할 개정판도 준비했다. 1853년 5월에는 리스트가 이 작품이 수정되었다는 것을 인정했지만, 끝내 바그너의 출판 계획은 무산되고 말았다. 이후 1855년에 리스트가 자신의 작품 파우스트 교향곡을 거의 완성하여 그 사실을 바그너에게 알렸고, 바그너는 다시 개정된 총보를 리스트에게 보내고 브라이트코프흐 & 헤르텔 출판사에는 20루이의 금을 받고 팔았다. 또한 그의 작품을 “하나하나의 음표가 시인의 피로 쓰여졌다”며 극찬했던 한스 폰 뷔로가 그것을 피아노 독주용으로 편곡했는데, 리스트는 그것을 약간 변형되었을 뿐이라고 지적했다. 이 서곡의 총보 첫머리에는 파우스트 1부의 내용 중 한 구절을 인용하고 있다.
- **Question** : 바그너의 작품을 시인의 피로 쓰여졌다고 극찬한 것은 누구인가?
- **Answer** : 한스 폰 뷔로
- **answer_start** : 402

KorQuad v1.0 Dataset Example 2

- **Context** : 한편 1840년부터 바그너와 알고 지내던 리스트가 잊혀져 있던 1악장을 부활시켜 1852년에 바이마르에서 연주했다. 이것을 계기로 바그너도 이 작품에 다시 관심을 갖게 되었고, 그 해 9월에는 총보의 반환을 요구하여 이를 서곡으로 간추린 다음 수정을 했고 브라이트코프흐 & 헤르텔 출판사에서 출판할 개정판도 준비했다. 1853년 5월에는 리스트가 이 작품이 수정되었다는 것을 인정했지만, 끝내 바그너의 출판 계획은 무산되고 말았다. 이후 1855년에 리스트가 자신의 작품 파우스트 교향곡을 거의 완성하여 그 사실을 바그너에게 알렸고, 바그너는 다시 개정된 총보를 리스트에게 보내고 브라이트코프흐 & 헤르텔 출판사에는 20루이의 금을 받고 팔았다. 또한 그의 작품을 “하나하나의 음표가 시인의 피로 쓰여졌다”며 극찬했던 한스 폰 뷔로가 그것을 피아노 독주용으로 편곡했는데, 리스트는 그것을 약간 변형되었을 뿐이라고 지적했다. 이 서곡의 총보 첫머리에는 파우스트 1부의 내용 중 한 구절을 인용하고 있다.
- **Question** : 잊혀져 있는 파우스트 서곡 1악장을 부활시킨 것은 누구인가?
- **Answer** : 리스트
- **answer_start** : 23

KorQuad v1.0 Dataset Example 3

- **Context** : 일본의 항복으로 태평양 전쟁, 제2차 세계 대전이 모두 끝나게 되었다. 이는 히로시마와 나가사키에 원자폭탄을 투하한 것이 가장 큰 원인이 되었다. 일본의 본토 공습에 수 만명의 민간인이 사망했음에도 불구하고 일본 정부는 항복하기를 거부하며 옥쇄할 각오가 되어 있었다. 일본 정부는 연합군의 본토 침공 작전인 몰락 작전에 철저히 준비해 오키나와와 이오지마에서의 전투처럼 강력히 저항하려고 했다. 일본 전역은 연합군의 침공 이후 일본 본토에 공군과 해군이 연합군을 지원하는 방식으로 진행될 예정이었다. 일본군은 여전히 본토에 200만 명이나 되는 군대를 보유하고 있었으며, 연합군의 침공을 좌절시킬 무기와 가용병력도 풍부했다. 만약 일본군과 연합군 사이에 전투가 벌어진다면 일본 본토 내에서 700만 명의 일본 군인 및 민간인과 50만 명이나 되는 미군 사상자가 발생할 것이라고 전망했다. 히로시마와 나가사키의 원폭 투하는 이러한 일본 본토 침공을 저지함으로써 사상자를 훨씬 줄이는 결과를 낳았다.
- **Question** : 태평양 전쟁과 제2차 세계 대전이 끝난 주요 원인이 된 도시는?
- **Answer** : 히로시마와 나가사키
- **answer_start** : 43

ChatGPT에게 KorQuad 데이터셋 질문 물어보기

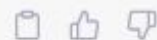
✧ GPT-4

VO

바그너의 작품을 시인의 피로 쓰여졌다고 극찬한 것은 누구인가?



독일의 철학자 프리드리히 니체(Friedrich Nietzsche)는 초기에 리히아르트 바그너(Richard Wagner)와 그의 작품을 매우 높이 평가했습니다. 니체는 바그너의 음악이 독일 문화와 예술에 미칠 영향에 대해 높은 기대를 품었고, 그는 바그너와 개인적으로도 친분을 쌓았습니다. 당시에 니체는 바그너의 작품을 "시인의 피로 쓰여졌다"고 극찬하기도 했습니다.



그러나 이들의 관계는 후에 악화되었고, 니체는 바그너와 그의 음악에 대한 자신의 견해를 변경했습니다. 니체는 바그너의 민족주의와 반유대주의, 그리고 그의 예술에 대한 순수한 예술성의 부재를 비판하기 시작했습니다. 이러한 변화는 니체의 저서 "니체 대 바그너"("Nietzsche contra Wagner")에서도 잘 나타나 있습니다.

니체와 바그너의 이러한 복잡한 관계는 19세기 뒷부분의 독일 문화와 철학, 예술에 깊은 영향을 끼쳤으며, 여러 분야에서도 널리 연구되고 논의되고 있습니다.

ChatGPT에게 KorQuad 데이터셋 질문 물어보기

✦ GPT-4

VO

잊혀져 있는 파우스트 서곡 1악장을 부활시킨 것은 누구인가?



정확한 정보가 없을 수 있지만, 페르디난트 라이스(Ferdinand Ries)가 이 작품을 알리는 데 중요한 역할을 했다고 알려져 있습니다. 페르디난트 라이스는 루트비히 판 베토벤의 제자이자 친구로, 그는 베토벤의 작품들을 널리 알리는 데 크게 기여했습니다. 또한, 그는 여러 가지 작품들을 부활시키거나 공연을 주최하는 등의 활동을 통해 음악계에 큰 영향을 미쳤습니다.



베토벤의 "파우스트 서곡" (WoO 59) 1악장은 완성되지 않은 작품 중 하나로, 오랫동안 많은 주목을 받지 못했습니다. 라이스의 노력 덕분에 이 작품이 다시 주목받게 되었을 가능성이 있습니다.

그러나 위 정보는 변경될 수 있으며, 다른 연구나 정보가 이를 뒷받침하거나 반박할 수 있습니다. 이러한 음악학적 문제는 지속적으로 연구되고 있으므로, 가장 최신의 연구 결과를 참고하는 것이 좋습니다.

ChatGPT에게 KorQuad 데이터셋 질문 물어보기

✦ GPT-4

VO

태평양 전쟁과 제2차 세계 대전이 끝난 주요 원인이 된 도시는?



제2차 세계 대전과 그 일부로 취급되는 태평양 전쟁이 끝난 주요 원인이 된 도시는 일본의 히로시마와 나가사키입니다. 1945년 8월 6일, 미국은 히로시마에 원자폭탄을 투하하였고, 그로 인해 대량의 사상자가 발생하였습니다. 이후 8월 9일에는 나가사키에도 원자폭탄이 투하되었습니다. 이 두 개의 원자폭탄 공격은 일본이 항복하는 데 큰 영향을 미쳤으며, 그 결과 8월 15일에 일본은 무조건 항복을 선언하였습니다. 이로써 제2차 세계 대전과 태평양 전쟁은 종식되었습니다.

