

2021 국립국어원 인공지능 언어 능력 평가 대회

- K N U A I -

Ju-Seong Do

Kyungpook National University

07, November. 2021.



▶▶ INDEX

▶▶ 팀 원

▶▶ 모델 개발 과정

▶▶ 모델 구동 방법

▶▶ 팀 원

경 북 대 학 교 K N U A I	대 학 원	학 부
인 공 지 능	<div> <div>도주성</div> <div>김준우</div> <div>윤혜경</div> <div>김동현</div> <div>윤은지</div> </div>	<div> <div>이주형</div> <div>윤민후</div> </div>
국 어 국 문 학	<div> <div>안진산</div> <div>고예린</div> </div>	<div> <div>성민규</div> </div>

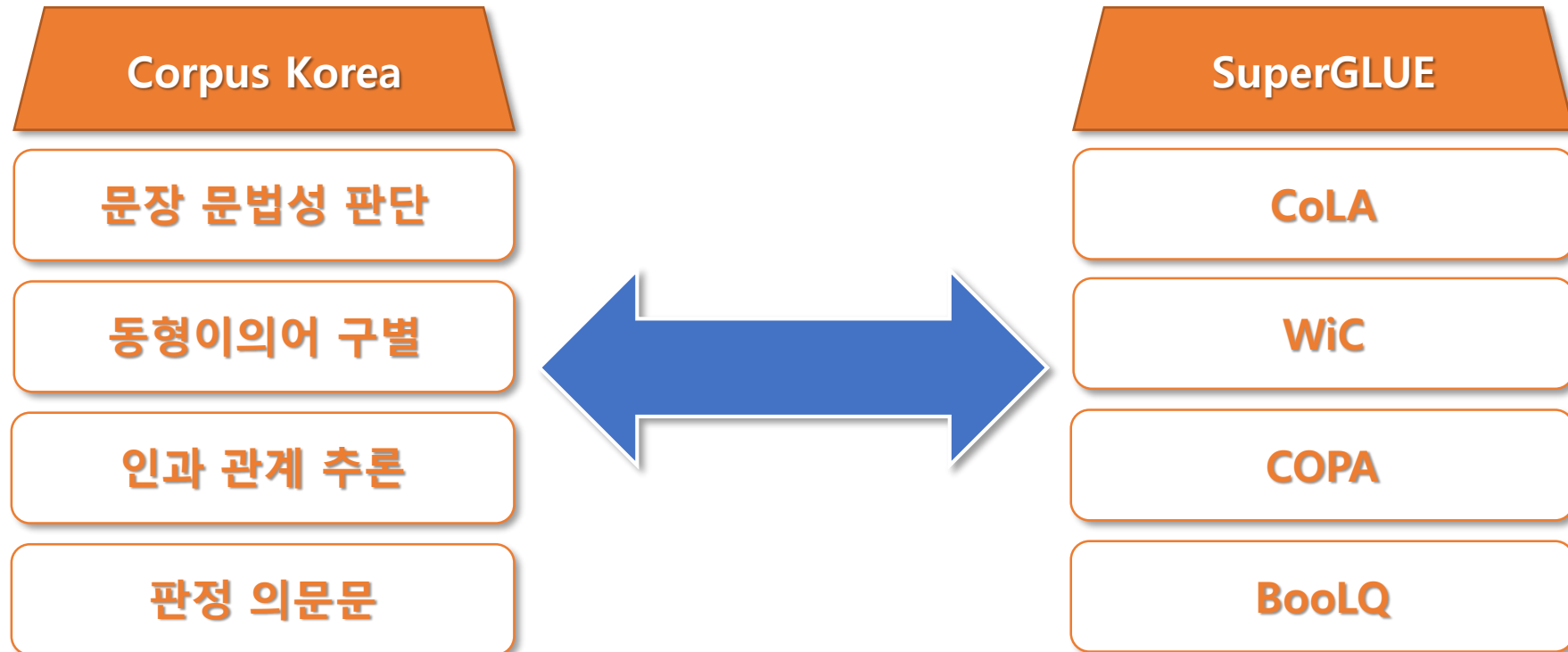
▶▶ 모델 개발 과정

1. 한국어에 최적화된 언어 모델 연구



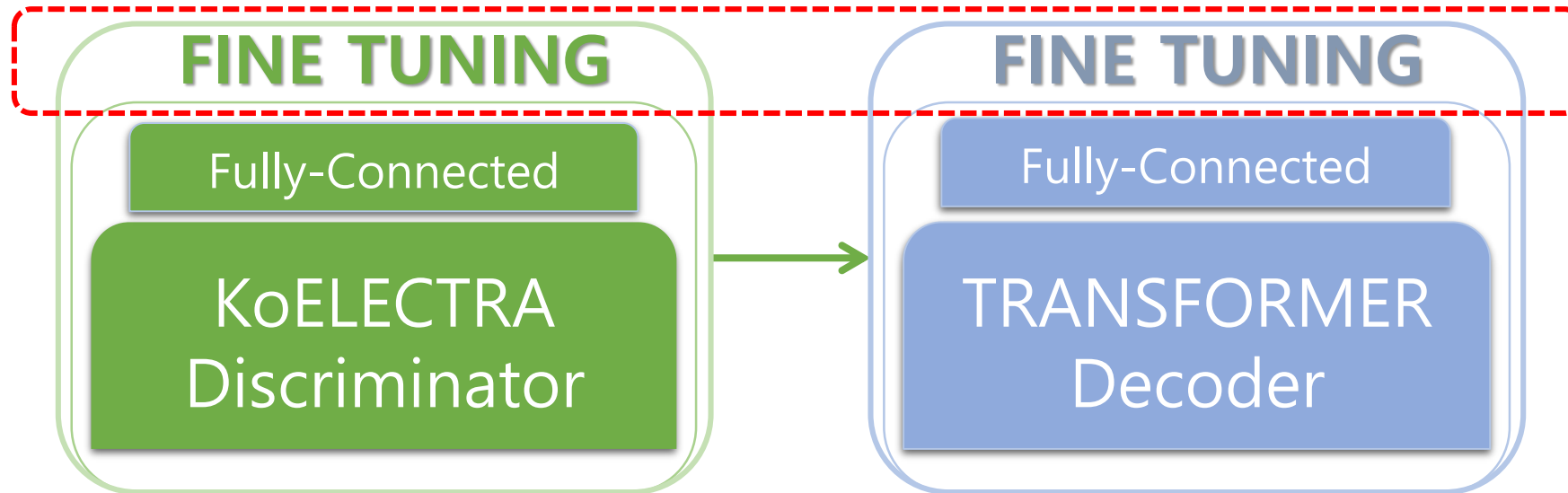
▶▶ 모델 개발 과정

2. 주어진 과제의 Dataset 분석



▶▶ 모델 개발 과정

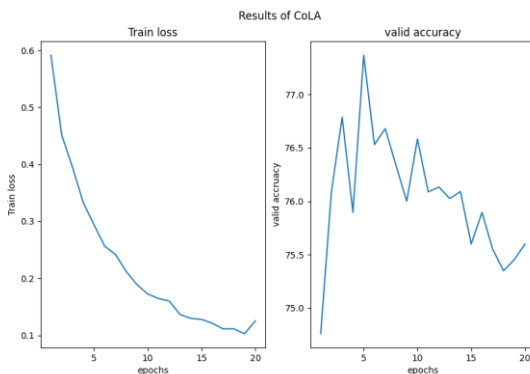
3. **KNUIAI_final** Based on KoELECTRA



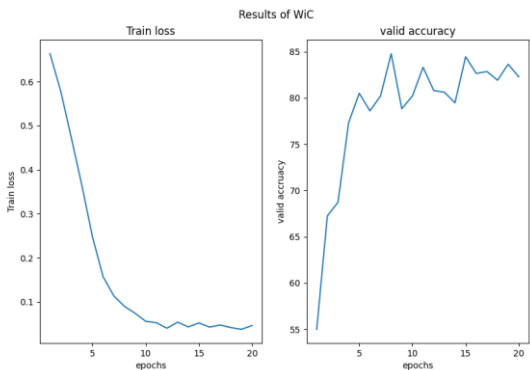
▶▶ 모델 개발 과정

3. KNUAI_final Based on KoELECTRA

CoLA

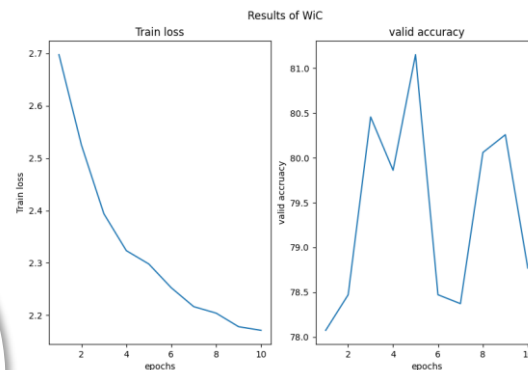


WiC

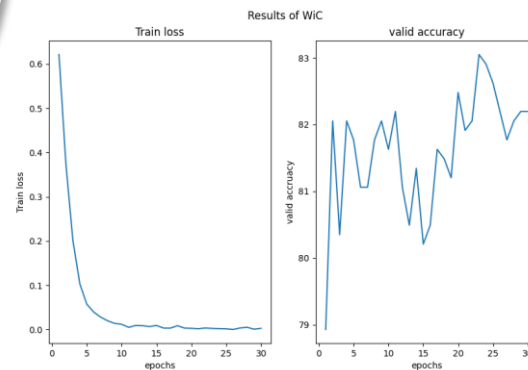


FINE-TUNING

COPA

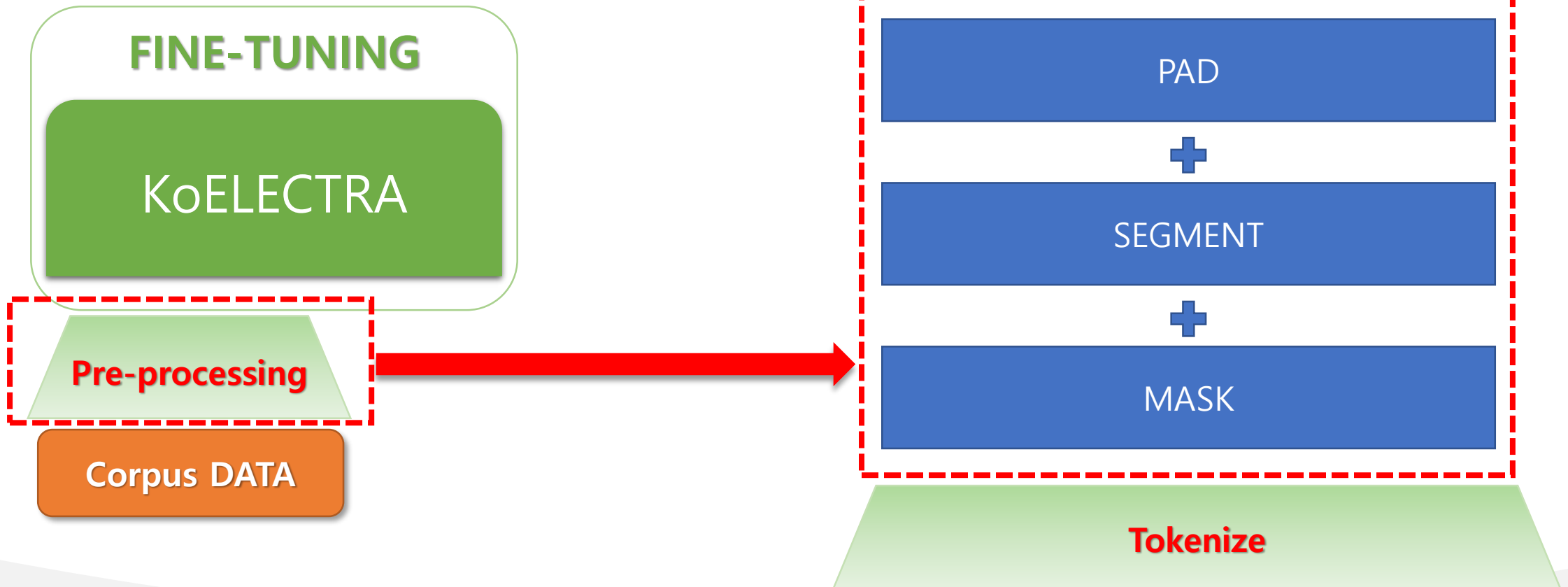


BoolQ



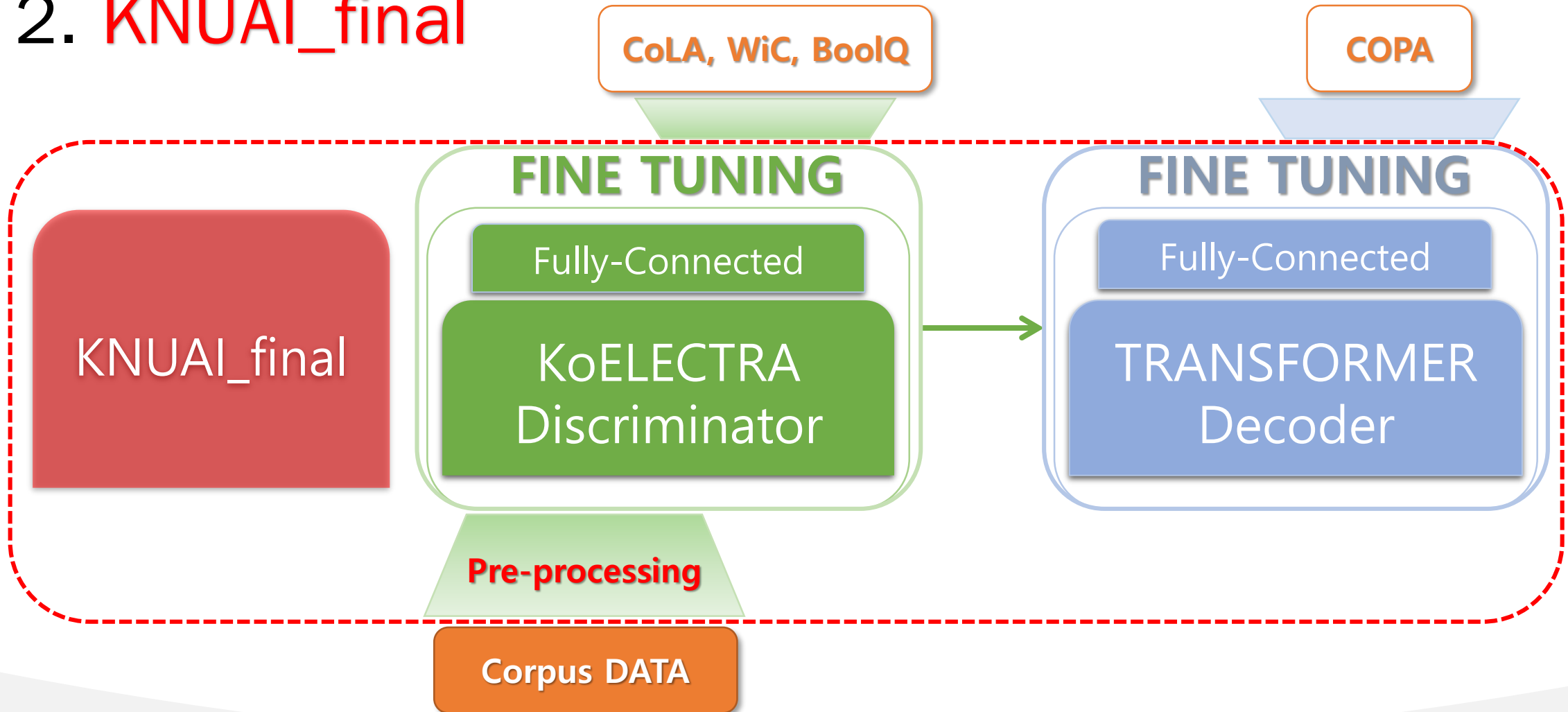
▶▶ 모델 구동 방법

1. Pre-processing



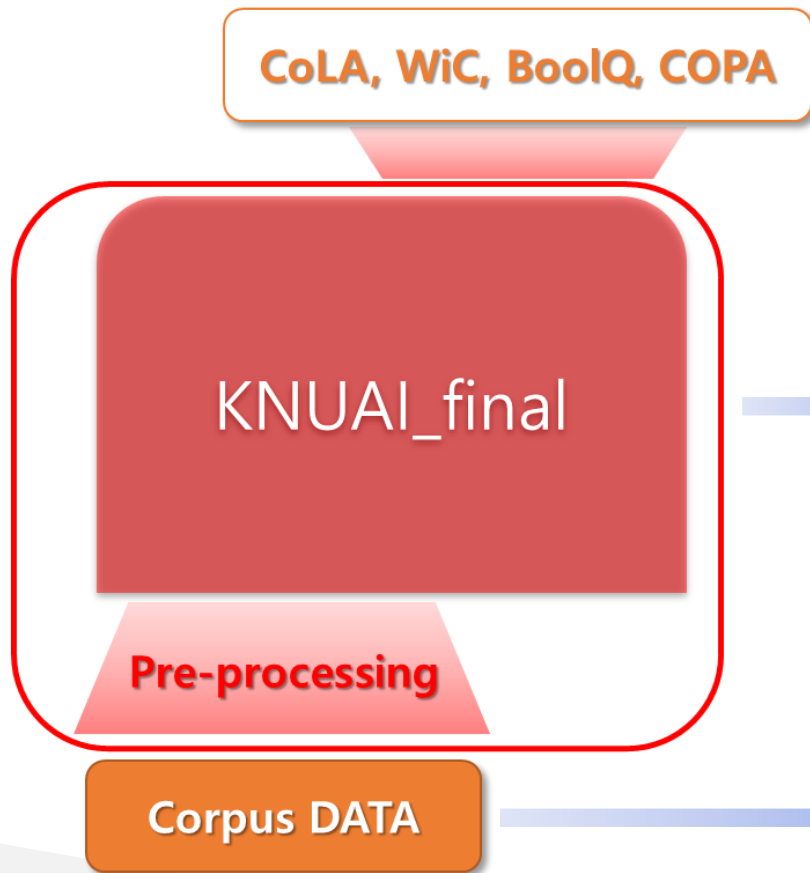
▶▶ 모델 구동 방법

2. KNUAI_final



모델 구동 방법

2. KNUAI_final



https://github.com/KNUAI2021/CorpusKorea_KNUAI/blob/main/evaluation.py

※ cf) Command Line:

```
~$ python evaluation.py --device cuda:0 --test_file_dir ./corpus/ --state_dir ./weights/ --output_dir ./result
```

③

Label Data: `--output_dir "output file path"`

(JSON 파일 저장할 경로)

②

Load Model: `--state_dir "state path"`

(Model Weight 불러올 경로)

①

Input Data: `--device "device"`
`--test_file_dir "test file path"`

※ The name of the test files should include one of (CoLA|cola), (WiC|wic), (COPA|copa), (BoolQ|boolq).

(GPU 설정)
(Dataset 불러올 경로)



▶ ▶ 마무리

- ▶ 한국어 NLP에 관심을 가지는 계기
- ▶ 한국어에 최적화된 언어 모델에 대해서 꾸준한 연구 필요성
- ▶ 문어체에 치중되어 있는 데이터 셋 이외의 다양한 텍스트 장르 데이터 셋 구축될 것이라 기대