



THE UNIVERSITY OF TEXAS AT DALLAS

# Korean Question Generation

# Question Generation

- Task
  - Question Generation
    - Generate a question based on a given piece of text.
    - Requires deep understanding of language and ability to generate relevant questions.

Irrelevant Generation		Relevant Generation	
Text	The cat sat on the mat. The dog barked loudly.	Text	The capital of France is Paris. The population of Paris is over 2 million people.
Output	What is the meaning of life?	Output	What is the capital of France?

# Korean Question Generation

## Models:

- T5-based models
- BART-based models
- GPT-based models
- LLaMA-based models

## Applications

- Japanese
- <https://huggingface.co/sonoisa/t5-base-japanese-question-generation>
- Spanish
- <https://huggingface.co/mrm8488/bert2bert-spanish-question-generation>

# Korean Question Generation

- Dataset
  - KorQuAD
    - The Korean Question Answering Dataset
    - Stats
      - Train Data: 83,486
      - Test Data: 10,165
    - From 47,957 Wiki articles

Type	Text
Label	임종석이 여의도 농민 폭력 시위를 주도한 혐의로 지명수배된 연도는?
Answer	1989년
Prompt	1989년 2월 15일 여의도 농민 폭력 시위를 주도한 혐의(폭력행위등처벌에관한법률위반)으로 지명수배되었다...

# Experiment Analysis

1. Human Evaluation
  - a. Manually compare label to generated text
2. Quantitative evaluation
  - b. BLEU(n-gram based)
3. Try different measurements

Label	임종석이 여의도 농민 폭력 시위를 주도한 혐의로 지명수배된 연도는?
Our proposed model	임종석이 서울지방검찰청 공안부에서 사전구속영장을 발부받은 해는?