

# QAEncoder: Towards Aligned Representation Learning in Question Answering System

---

The project repository consists of the following entries:

- **Demo.py**: This script provides a demonstration of the QAEncoder, corresponding to the QAE\_base method. It serves as a basic introduction and the motivation experiment.
- **EX1**: This directory contains the experimental code required for Experiment 1 (EXPERIMENTS ON CLASSICAL DATASETS), including both dense and sparse methods, along with the necessary datasets: MS-MARCO, SQuAD, and KILT datasets comprising NQ, ELI5, TriviaQA and HotPotQA.
- **EX2**: This directory includes the experimental code required for Experiment 2 (EXPERIMENTS ON LATEST DATASETS), as well as the associated datasets from FIGNEWS and CRUD-RAG, covering multiple languages such as English, Arabic, French, Hindi, Hebrew and Chinese.
- **EX3**: ABLATION STUDIES
  - **EX3(naive)**: This directory holds the extra code relevant to the naive method, utilizing datasets from EX2.
- **EX4**: TRAINING-BASED AND DOCUMENT-CENTRIC METHODS
  - **EX4(instruct)**: This directory holds the extra code for the instruction-tuned method, also leveraging datasets from EX2.
  - **EX4(document)**: This directory holds the extra code for the document-centric method, again using datasets from EX2.
- **Data**: This directory contains some organized experimental tables.

Note: the experiments require the completion of model paths. Search "need complete embed\_model\_path and uncomment the next line" in the code files.