**Code of Manuscirpt**: Empowering College Students to Select Ideal Advisors: A Keyword-Based Recommendation Model

**1. Running Environment and Configuration**

Memory: 90 GB

GPU: NVIDIA vGPU-32GB

Python==3.10.8

torch==2.6.0

transformers==4.51.2

scipy

jsonlines

typing

tqdm

pandas

**2.Files**

(1) LLM\_BERT\_SIMCES\_AVRD.py: the main results of our model.

(2) baselines.py: the results of baseline algorithms(TF-IDF, LSA, LSA+TF-IDF, Word2Vec).

(3) bert\_albation.py: the ablation experiment results(table 4).

(4) LLM\_Qwen\_mlp\_recom\_loadh5.py: the results of the LLM Qwen.

(5) LLM\_Deepseek\_no\_SIMCES.py: the results of LLM Deepseek without SIMCES.

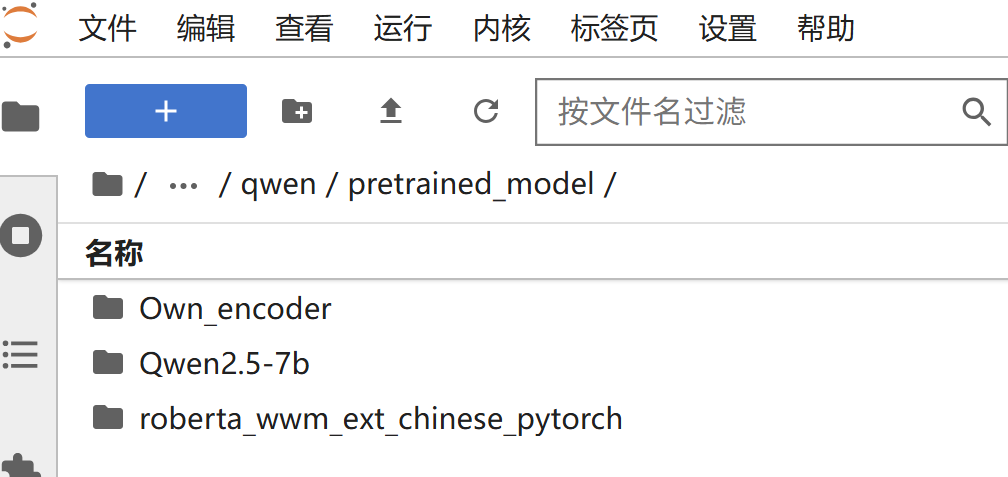
(6) LLM\_Qwen\_no\_SIMCES.py: the results of LLM Qwen without SIMCES.

(7) LLM\_Qwen\_vector\_data.py: To improve computing efficiency, we transform students’ keywords and teachers’ texts to vectors using LLM Qwen.

(8) LLM\_train\_Qwen\_mlp.py: train a Qwen+SimCES model.

(9) LLM\_inference\_Teacher.py: We generated teacher profiles (with a length of less than 250 words for each teacher) based on teaching records by calling the DeepSeek interface to conduct a second round questionnaire adjustments.

Due to the large memory consumption of LLMs, we haven't uploaded them. Please download and place them in the corresponding folder, as shown in the following figure.



The ‘3Questionnaire\_rec\_new.xlsx’ in the folder ‘Data’ is the adjusted results after a second round questionnaire adjustments, and ‘the 0Questionnaire\_word\_segmentation.xlsx’, ‘1Questionnaire\_word\_segmentation.xlsx’ and ‘2Questionnaire\_word\_segmentation.xlsx’ are the first round results.