

ZI'AN ZHENG



Addr: West Xi'erqi Road, Haidian District, Beijing, P.R. China

Tel: +86-17610826711/+65-93759156

Email: zheng_zian@u.nus.edu

Website: orion-zheng.github.io

EDUCATION BACKGROUND

School of Information Science & Engineering, Lanzhou University (LZU, Project 985) 09/2018 - 07/2022

- **B. Eng. in Data Science and Big Data Technology**
- **GPA:** 92.8/100, **Ranking:** 1/192
- **Honors:** China National Scholarship (Top 0.1%), 12/2019 & 12/2021
Merit Student in Colleges and Universities in Gansu Province (Top 1%), 05/2021
LZU First-class Scholarship (Top 3%) & LZU Outstanding Model Student Award (Top5%), 12/2020

School of Computing, National University of Singapore (NUS) 08/2022 - 08/2024 (Expected)

- **M. Comp in Artificial Intelligence Specialization (Dissertation Option)**
- **GPA:** 4.42/5, **Supervisor:** Prof. Yang You (Presidential Young Professor)
- **Coursework about Deep Learning and Natural Language Processing:**
CS6207 Advanced NLP (A), CS5242/CS5260 Deep Learning I/II (A-), CS4244 NLP(A-)

RESEARCH INTERESTS

I have a deep interest in the Large Language Model (LLM), especially in the following areas:

- **Data-Centric Approaches:** Focusing on data quality and data strategies (e.g. Data Mixture and Data Curriculum)
- **Efficient LLM Design and Training:** MoE (Mixture-of-Experts) Model, Efficient Context Extrapolation Method.
- **Maximizing Trained LLM Availability/Capability:** Efficient Inference, Prompt Engineering, LLM-based Agent.

RESEARCH EXPERIENCE

Master Student in HPC-AI Lab, National University of Singapore 05/2023 – Now

Dissertation: OpenMoE: An Open-sourced Conditional Computation LLM (**Mentor:** Fuzhao Xue)

Keywords: Data-Centric methods, Mixture-of-Experts Model

- Working on OpenMoE project (second author) with Fuzhao Xue, which is the **first open-source, decoder-only MoE language model**. The OpenMoE-8B with 2B active parameters has been trained on 1.1T tokens. We released the checkpoint and got **~750 stars** on [GitHub](https://github.com) and **~500 likes** on [Twitter](https://twitter.com).
- Investigated publicly available pre-training corpus (English, Chinese, multilingual, code, etc), preprocessing methods and tokenization techniques. Do experiments comparing tokenizers. Prepare the pre-training, SFT and evaluation datasets in TFDS format.
- Worked on the [Pytorch implementation](#) of OpenMoE with ColossalAI team. Now conducting literature review of Mixture-of-Experts models and writing the paper.

WORK EXPERIENCE

Artificial Intelligence Engineer Intern, HPC-AI Tech 07/2023 – 11/2023

Keywords: Data-Centric methods, Long Context LLM, Retrieval Augmented Generation

- Extended the LLaMA's vocabulary for Chinese text and participated in the data cleaning and preparation process in the [Colossal-LLaMA-2](#) project (achieved **186k+ downloads** on Huggingface so far).
- Context length extrapolation: Investigated common context extrapolation techniques (e.g. PI, NTK, LongLLaMA, LongLoRA, etc.), training corpus with long data and long-text evaluation methods. Working on constructing Chinese long text training data and doing multi-GPU training to extrapolate Colossal-LLaMA-2.
- Working on the [ColossalQA](#) project, a retrieval-based QA framework based on Langchain.
- Involved in the writing of the book 'Practical Large AI Models', edited by Professor Yang You.

ZI'AN ZHENG



Addr: West Xi'erqi Road, Haidian District, Beijing, P.R. China

Tel: +86-17610826711/+65-93759156

Email: zheng_zian@u.nus.edu

Website: orion-zheng.github.io

PROJECT EXPERIENCE

Course Project of CS5260 Deep Learning II

02/2023 - 04/2023

Crafting Malicious Soft and Hard Prompts through Gradient Ascent with Auxiliary Loss

Keywords: Prompt-Tuning, Prompt Injection Attack

- Proposed a gradient-based prompt injection attack method with an auxiliary loss, decreasing the classification accuracy of fine-tuned and prompt-tuned GPT-2 models by 9.7% and 3.4%, respectively.
- Implemented the code for the prompt-tuning and prompt injection attack.

Course Project of CS4248 Natural Language Processing

02/2023 – 04/2023

-ELCo: Towards Gender Equality in Emoji Composition (**Keywords:** Contrastive Learning, Text Similarity)

- We developed a framework that translate emoji compositions into English descriptions and mitigated the possible gender bias problem through data augmentation, e.g. 🐻*🐻 → Polar Bear.
- First, we used ChatGPT API to generate possible descriptions candidates for a given emoji sequence, and then used our trained ranking model to select the most plausible description from all candidates. During the training, we mitigated the gender bias problem in the training dataset by generating data for both genders.
- I was responsible for implementing the SimCSE model to encode emoji sequences and their text descriptions into embeddings. I also implemented the code for ranking and evaluation part (spearman correlation).

Course Project of CS6207 Advanced Natural Language Processing

02/2023 – 04/2023

Uncovering and Addressing Biases in Machine Translation: Insights from WMT22 News Commentary Dataset

Keywords: Machine Translation, Bias Evaluation/Mitigation

- We optimized the sentiment bias evaluation method from the ACL2022 paper ‘Measuring and Mitigating Name Biases in Neural Machine Translation’ and evaluated the bias towards politicians in the News Commentary dataset.
- We found that the models trained on news commentary dataset translated positively for western and male politicians more than non-western and female politicians. To mitigate the bias, we simplified the *Switch Entity Method* in the previous work by only swapping the subject words and achieved similarly desirable results.
- I proposed the idea and worked on implementing the bias evaluation pipeline, as well as the code and experiments to fine-tune the mT5-base machine translation model. I also participated in the report writing and presentation.

Lanzhou University Undergraduate Capstone Project

12/2021 - 4/2022

Domain-specific Subject Word Extraction for Weibo Topics (**Keywords:** Text Classification/Clustering)

- Developed a technical solution for extracting topic words from many Weibo topic tags by dividing the overall task into three sub-tasks: text classification, topic text clustering, and keyword extraction.
- Performed multi-label classification on Weibo topic short texts based on RoBERTa, dividing them into 9 different domains, and improved the model performance by introducing label smoothing and adversarial training, with a final model accuracy of 75.3% on the test set.
- Utilized Sentence-BERT to generate semantic representations for short texts for calculating semantic similarity and applied DBSCAN clustering method to obtain text clusters under the same topic. Finally, applied text summarization methods to obtain the key phrase of each topic cluster.

ENGLISH & COMPUTER SKILLS

English: TOEFL:105 | GRE: 322 (Q:168, V:154) + Writing 3

Programming & Software: Python, Java, Linux, C, SQL, Tableau, Echarts

Libraries: PyTorch, Huggingface, Langchain, Pandas, NumPy, Scikit-Learn, PyQt5