Lei Li

✉ nlp.lilei@gmail.com · % Homepage · ☎ (+86) 18810138182 · Google Scholar

🎓 Education

The University of Hong Kong, SAR, Hong Kong   2023.07 – Now
PhD candidate in Computational Linguistics
Advisor: Prof. Qi Liu & Lingpeng Kong

Peking University, Beijing, China   2020.09 – 2023.06
Master student in Computational Linguistics
Advisor: Prof. Xu Sun, GPA: 3.6/4.0, **National Scholarship Winner**

Xidian University, Shaanxi, China   2016.09 – 2020.06
B.Sc. in Software Engineering
GPA: 94.2 (Ranking 1 / 448), English: CET-6 638, **National Scholarship Winner**

👥 Experience

Shanghai Ai Lab Shanghai, China   2022.07 – 2023.06
Research Intern   Mentor: Dr. Jingjing Xu & Prof. Lingpeng Kong

Research Topic: Investigating the learned knowledge of large-scale pre-trained cross-modal models such as CLIP. Building a benchmark for comprehensive evaluation, devising novel probing method for CLIP and conducting comparative analysis with text-only models. Corresponding paper is submitted to ICLR 2023.

Toutiao Search, ByteDance Beijing, China   2021.12 – 2022.06
Search Algorithm Intern   Mentor: Dr. Jiacheng Zhang & Chao Qiao

Achievements: (1) Fine-tuning pre-trained language models for optimizing the Top-1 search result for better display effect, including highlighting and structuring the content regarding the query. Our model has been deployed with +25 GSB score. (2) Contributing to inner model training framework about dynamic distillation objective proposed in our EMNLP paper, which has been proved effective in our task.

Wechat AI, Tencent Beijing, China   2020.04 – 2021.11
NLP Research Intern   Mentor: Prof. Peng Li & Prof. Yankai Lin

Research Topic: (1) Analysing the limitation of dynamic early exiting framework, and proposing a cascading framework to speed-up the inference according to instance complexity. (2) Investigating a dynamic knowledge distillation regarding distillation data selection, teacher model adoption and distillation objective adjustment, towards better training efficiency and distillation performance. Publish 2 EMNLP papers.

📕 Selected Publications

M^3IT: A Large-Scale Dataset towards Multi-Modal Multilingual Instruction Tuning
- **Lei Li**, Yuwei Yin, Shicheng Li, Liang Chen, Peiyi Wang, Shuhuai Ren, Mukai Li, Yazheng Yang, Jingjing Xu, Xu Sun, Lingpeng Kong, Qi Liu.
- preprint

Can Language Models Understand Physical Concepts?
- **Lei Li**, Jingjing Xu, Qingxiu Dong, Ce Zheng, Qi Liu, Lingpeng Kong, Xu Sun
- Pre-print

A Survey for In-context Learning
- Qingxiu Dong, **Lei Li**, Damai Dai, Ce Zheng, Zhiyong Wu, Baobao Chang, Xu Sun, Jingjing Xu, Zhifang Sui
- Pre-print

Dynamic Knowledge Distillation for Pre-trained Language Models

- **Lei Li**, Yankai Lin, Shuhuai Ren, Peng Li, Jie Zhou, Xu Sun.
- EMNLP 2021 (Oral)

Distributional Correlation–Aware Knowledge Distillation for Stock Trading Volume Prediction

- **Lei Li**, Zhiyuan Zhang, Ruihan Bao, Keiko Harimoto, Xu Sun.
- ECML-PKDD 2022 (Oral)

CascadeBERT: Accelerating Inference of Pre-trained Language Models via Calibrated Complete Models Cascade

- **Lei Li**, Yankai Lin, Deli Chen, Shuhuai Ren, Peng Li, Jie Zhou, Xu Sun.
- Findings of EMNLP 2021

From Mimicking to Integrating: Knowledge Integration for Pre-trained Language Models

- **Lei Li**, Yankai Lin, Xuancheng Ren, Guangxiang Zhao, Peng Li, Jie Zhou, Xu Sun.
- Findings of EMNLP 2022.

Enhancing Topic-to-Essay Generation with External Commonsense Knowledge

- Pengcheng Yang*, **Lei Li*** (Equal Contribution), Fuli Luo, Tianyu Liu, Xu Sun.
- ACL 2019

Alleviating the Knowledge-Language Inconsistency: A Study for Deep Commonsense Knowledge

- Yi Zhang*, **Lei Li*** (Equal Contribution), Yunfang Wu, Qi Su, Xu Sun.
- IEEE/ACM Transactions on Audio, Speech, and Language Processing (TASLP)

Leveraging Word-Formation Knowledge for Chinese Word Sense Disambiguation

- Hua Zheng*, **Lei Li*** (Equal Contribution), Damai Dai, Deli Chen, Tianyu Liu, Xu Sun, Yang Liu.
- Findings of EMNLP 2021 (Short Paper)

💼 Professional Services

Program Committee (Reviewer):

NeurIPS 2023, ACL 2019-2023, EMNLP 2019-2023, ECML-PKDD 2022

Teaching Assistant:

- Introduction to Natural Language Processing
- Computational Linguistics

⌭ Projects

Multi-Modal Instruction Tuning Dataset (35k downloads)

Multi-modal multilingual instruction tuning dataset with 400 manually written instructions covering 40 tasks.

Hip-pop GAN (100+ stars)

Chinese lyrics generation with modified SeqGAN for rhymed hip-pop lyrics.

Awesome Efficient PLMs (70+ stars)

Paper list for techniques to improve the training and inference efficiency of pre-trained language models.

❈ Honors and Awards

National Scholarship, Ministry of Education	2022.10
Merit Student, Peking University	2022.10
National Scholarship, Ministry of Education	2019.11
Best Demo Award, DeeCamp 2018, Peking University & Sinovation Ventures	2018.08