





YUXUAN LU

✉ lu.yuxuan@northeastern.edu ·  yuxuan-lu ·  https://yuxuan.lu ·   

EDUCATION

Northeastern University, MA, USA Sep. 2023 – Present


Ph.D. in Computer Science. Advised by Prof. Dakuo Wang

Beijing University of Technology, Beijing, China Sep. 2019 – Jun. 2023

Graduated **with honor**. Bachelor of Engineering in Computer Science and Technology

HONORS AND AWARDS

 3 × *Bronze Medal*, ICPC Asia Regional Contest 2019, 2020, 2021

 *Global Rank 42 (Top 2%) and 85 (Top 3.5%)*, IEEEExtreme programming competition 2020, 2021

PUBLICATIONS

- Yuxuan Lu, Bingsheng Yao, Shao Zhang, Yun Wang, Peng Zhang, Tun Lu, Toby Jia-Jun Li, Dakuo Wang, *Human Still Wins over LLM: An Empirical Study of Active Learning on Domain-Specific Annotation Tasks* (arxiv preprint, 2311.09825)
- Bingsheng Yao, Guiming Chen, Ruishi Zou, Yuxuan Lu, Jiachen Li, Shao Zhang, Sijia Liu, James Hendler, Dakuo Wang, *More Samples or More Prompt Inputs? Exploring Effective In-Context Sampling for LLM Few-Shot Prompt Engineering* (arxiv preprint, 2311.09782)
- Jiaju Chen, Yuxuan Lu, Shao Zhang, Bingsheng Yao, Yuanzhe Dong, Ying Xu, Yunyao Li, Qianwen Wang, Dakuo Wang, Yuling Sun, *FairytalesCQA: Integrating a Commonsense Knowledge Graph into Children's Storybook Narratives* (arxiv preprint, 2311.09756)
- Bingsheng Yao, Ishan Jindal, Lucian Popa, Yannis Katsis, Sayan Ghosh, Lihong He, Yuxuan Lu, Shashank Srivastava, James Hendler, Dakuo Wang, *Beyond Labels: Empowering Human Annotators with Natural Language Explanations through a Novel Active-Learning Architecture* (EMNLP Findings 2023, arxiv: 2305.12710)
- Yuxuan Lu, Jingya Yan, Zhixuan Qi, Zhongzheng Ge, and Yongping Du, *Contextual embedding and model weighting by fusing domain knowledge on Biomedical Question Answering (ACM-BCB 2022, with oral presentation)*, arxiv: 2206.12866)

RESEARCH EXPERIENCE

Microsoft Research Asia & LinkedIn Beijing, China Jul. 2022 – May. 2023

Machine Learning Researcher Manager: Guangming Lu (Manager) / Lun Du (Sr. Researcher)

Participated in a joint program between Microsoft Research Asia and LinkedIn, to discover AI applications driven by LinkedIn's large-scale high-quality production data.

- Training framework for PyTorch models in LinkedIn's internal Kubernetes cluster
 - Contributed several patches on HDFS support for upstream packages (fsspec, pytorch_lightning)
 - Write training framework for scaleable parallel training in Kubernetes cluster
- Heterogeneous Knowledge-based Person-Job Fit
 - Conducted Person-Job Fit research using heterogeneous GNN pre-training

- Participated in method designing, collecting and preprocessing TBs of data with spark and running experiments; our paper is **accepted to WSDM 2024**

Tsinghua NLP Lab (THUNLP)

Dec. 2021 – Jun. 2022

Intern Research Assistant Supervisor: Prof. Zhiyuan Liu, Dr. Huadong Wang

- Big Model for Knowledge Graph (BMKG)
 - Developed a toolkit to help **train large Knowledge Embedding models** on **large KGs** and **run various downstream tasks**
 - Supports **4 levels of parallel** during the training process of **translation-based or context-based** Knowledge Embedding models
 - Designed the framework and wrote code that needed high performance
- Design / Develop / Maintain multiple demos for NLP models
 - Designed and maintained multiple demos for NLP models to show their performance to non-specialists

Beijing University of Technology

Dec. 2020 – Dec. 2021

Research Assistant Supervisor: Prof. Yongping Du

- Biomedical Machine Reading Comprehension
 - Conducted Machine Reading Comprehension research in Biomedical Domain as the project leader, including designing the model, conducting experiments and writing the paper, which was **published in ACM BCB 2022**
 - Designed a **contextual embedding** and **model weighting** strategy to **learn domain knowledge** in Biomedical Question Answering task, which **outformed SOTA models by a large margin**

PROJECT EXPERIENCE

Course Grading and Feedback System based on Fault-Cause analysis Apr. 2020 – Jun. 2022

- Worked as the project leader, who was responsible for including designing the system architecture, code reviewing, and full stack developing
- Designed an **autograding system** that can help daily teaching and **give accurate scores and feedback** based our **automatic fault-cause clustering method**; the system includes 60k+ lines of code, and 72% of them are covered by unit tests
- Achieved the performance that is capable of handling **500+ QPS** while other similar systems can only do 20+
- **Found a bug in the go compiler** (see [golang/go#44614](https://golang.org/go#44614))

Open-source contributions

- LaTeX-Utilities: VSCode extension to help writing LaTeX. **187,000** downloads
- Projects that I've contributed to: **Pytorch Lightning** / **fsspec** / **dashmap** / **GitLab** / ...
- See more at https://yuxuan.lu/open_source

SKILLS

- Programming: Multilingual. Fluent in C++, Rust, Python, Go, JavaScript, etc.
- Deep Learning Framework: PyTorch, Hands-on experience in large-scale parallel training
- Data processing: Apache Spark, Pandas
- Languages: English - Fluent (TOEFL 108), Mandarin - Native speaker