Tianjian Li

Updated October 28, 2024

Email: tli104@jhu.edu **Website**: tianjianl.github.io

Education

Johns Hopkins University — Baltimore, MD
PhD in Computer Science — 2024 – Present
Advisor: Daniel Khashabi

Johns Hopkins University — Baltimore, MD
MSE in Computer Science — 2022 – 2024
Advisors: Kenton Murray, Daniel Khashabi, Philipp Koehn

New York University — New York, NY
BA in Computer Science and Mathematics — 2017 – 2021

Publications

***Upsample or Upweight?* Balanced Training on Heavily Imbalanced Datasets**
Tianjian Li, Haoran Xu, Weiting Tan, Kenton Murray, Daniel Khashabi
Under Review. Link

***Verifiable by Design:* Aligning Language Models to Quote from Pre-Training Data**
Jingyu Zhang, Marc Marone, **Tianjian Li**, Benjamin Van Durme, Daniel Khashabi
Under Review. Link

Error Norm Truncation: Robust Training in the Presence of Data Noise for Text Generation Models
Tianjian Li, Haoran Xu, Philipp Koehn, Daniel Khashabi, Kenton Murray
International Conference on Learning Representations (ICLR), 2024. Spotlight (Top 5%). Link

Why Does Zero-Shot Cross-Lingual Generation Fail? An Explanation and a Solution
Tianjian Li, Kenton Murray
Association of Computational Linguistics (ACL) - Findings, 2023. Link

Research experience

Johns Hopkins University — 2022 – 2024
Center for Language and Speech Processing (CLSP)
Research Assistant
Advisors: Kenton Murray, Daniel Khashabi, Philipp Koehn
- Data re-weighting for heavily imbalanced datasets. (Under Review)
- Locating errors in training data. (ICLR' 24)

Tsinghua University & Zhipu.AI — Spring 2022
Research Intern
Advisor: Jie Tang
- Data curation and pre-training of large multilingual language models.
- Multilingual Language Model evaluation.

Industry experience

Baidu Inc. Baidu Maps — Beijing, China
Machine Learning Engineer (Intern) — Fall 2021
- Optimization of estimated trip time with graph neural networks.

Skills

Programming: Python, C/C++, Java
Frameworks: PyTorch, Huggingface (Accelerate), Fairseq, DeepSpeed, Jax
Tools: Docker, git, Hadoop streaming, Spark, Vim, LaTeX

Service

Reviewer: ACL (2023, 2024), EMNLP (2023, 2024), EACL (2024), COLM (2024), ICLR (2025)