

# Tianjian Li

Updated July 1, 2024

Email: tli104@jhu.edu Website: [tianjianl.github.io](https://tianjianl.github.io)

## Education

### Johns Hopkins University

PhD in Computer Science  
Advisor: Daniel Khashabi

Baltimore, MD

2024 – Present

### Johns Hopkins University

MSE in Computer Science  
Advisors: Kenton Murray, Daniel Khashabi, Philipp Koehn

Baltimore, MD

2022 – 2024

### New York University

BA in Computer Science and Mathematics

New York, NY

2017 – 2021

## Publications

### **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*. [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

#### **Center for Language and Speech Processing (CLSP)**

Advisors: Kenton Murray, Daniel Khashabi, Philipp Koehn  
- Efficient online domain weighting for training language models.  
- (ICLR '24) Removing data errors during training for language models.  
- (ACL '23) Improving cross-lingual transfer for text generation.

2022 – 2024

### Tsinghua University

Advisor: Jie Tang  
- Pre-training of large multilingual language models.  
- Cross-lingual summarization with multilingual LMs.

Spring 2022

## Teaching experience

### **Course Assistant, Department of Computer Science**

CSCI 601.471/671 NLP: Self-Supervised Models

Spring 2024

## Industry experience

### **Baidu Inc.** Baidu Maps

Machine Learning Engineer (Intern)  
- Optimization of estimated trip time with Graph Neural Networks.

Beijing, China

Fall 2021

## Skills

**Programming:** Python, C/C++, Java

**Frameworks:** PyTorch, Huggingface (Accelerate), Fairseq, DeepSpeed, Jax

**Tools:** Docker, git, Hadoop streaming, Spark, Vim, LaTeX

**Languages:** Chinese (Native), English, French (Beginner)

## Service

**Reviewer:** ACL (2023, 2024), EMNLP (2023), EACL (2024), COLM (2024)

**Organizer:** MASC-SLL (2024)

