Shiyue Zhang
Ph.D. Student
https://www.cs.unc.edu/shiyue

July 6, 2023
Mobile:+1-(919) 448-5607
Email:shiyue@cs.unc.edu

WORK EXPERIENCE

EXPECT TO START IN AUG. 2023	Research Engineer **Bloomberg AI**, New York, USA

EDUCATION

AUG. 2018–AUG. 2023	Ph.D. student in Computer Science **University of North Carolina at Chapel Hill**, Chapel Hill, USA
SEP. 2015–MAR. 2018	M.E. in Information and Communication Engineering **Beijing University of Posts and Telecommunications**, Beijing, China
SEP. 2011–JUL. 2015	B.E. in Communication Engineering **Beijing University of Posts and Telecommunications**, Beijing, China

INTERN EXPERIENCE

MAY 2022–AUG. 2022	**Bloomberg AI** *Engineering Intern, Supervised by Ozan Irsoy, Steven Lu, Shijie Wu, David Rosenberg, and Mark Dredze* • Language modeling
MAY 2021–AUG. 2021	**Facebook AI** *Research Intern, Supervised by Vishrav Chaudhary and Francisco (Paco) Guzmán* • Multilingual tokenization
JUNE 2020–AUG. 2020	**Microsoft Research** *Research Intern, Supervised by Asli Celikyilmaz* • Email thread summarization
SEP. 2016–JUNE 2018	**CSLT, Tsinghua University** *Research Intern, Supervised by Prof. Dong Wang, Prof. Yang Feng* • Machine translation
JUL. 2017–SEP. 2017	**SAILING Lab, Carnegie Mellon University** *Research Intern, Supervised by Dr. Pengtao Xie, Prof. Eric Xing* • Automatic medical diagnosis

PUBLICATIONS

2023	**Shiyue Zhang**, Shijie Wu, Ozan Irsoy, Steven Lu, Mohit Bansal, Mark Dredze and David Rosenberg. *MixCE: Training Autoregressive Language Models by Mixing Forward and Reverse Cross-Entropies* ACL 2023
2023	Derek Tam, Anisha Mascarenhas, **Shiyue Zhang**, Sarah Kwan, Mohit Bansal, Colin Raffel. *Evaluating the Factual Consistency of Large Language Models Through Summarization* Findings of ACL 2023
2023	**Shiyue Zhang***, David Wan*, and Mohit Bansal. *Extractive is not Faithful: An Investigation of Broad Unfaithfulness Problems in Extractive Summarization* ACL 2023

2023   Swarnadeep Saha, **Shiyue Zhang**, Peter Hase, Mohit Bansal. *Summarization Programs: Interpretable Abstractive Summarization with Neural Modular Trees* ICLR 2023

2022   Xiang Zhou, **Shiyue Zhang**, and Mohit Bansal. *Masked Part-Of-Speech Model: Does modeling long context help unsupervised POS-tagging?* NAACL 2022

2022   Yinuo Hu*, **Shiyue Zhang***, Viji Sathy, A. T. Panter, and Mohit Bansal. *SETSum: Summarization and Visualization of Student Evaluations of Teaching* NAACL Demo 2022

2022   **Shiyue Zhang**, Vishrav Chaudhary, Naman Goyal, James Cross, Guillaume Wenzek, Mohit Bansal, and Francisco Guzman. *How Robust is Neural Machine Translation to Language Imbalance in Multilingual Tokenizer Training?* AMTA 2022

2022   **Shiyue Zhang**, Benjamin Frey, and Mohit Bansal. *How can NLP Help Revitalize Endangered Languages? A Case Study and Roadmap for the Cherokee Language* ACL 2022 Theme Track

2021   **Shiyue Zhang** and Mohit Bansal. *Finding a Balanced Degree of Automation for Summary Evaluation* EMNLP 2021

2021   **Shiyue Zhang**, Benjamin Frey, and Mohit Bansal. *Cherokee-English Machine Translation Demo with Quality Estimation and Corrective Feedback* ACL Demo 2021, [News: The sanctity of Cherokee]

2021   **Shiyue Zhang**, Asli Celikyilmaz, Jianfeng Gao, and Mohit Bansal. *EmailSum: Abstractive Email Thread Summarization* ACL 2021

2021   Zineng Tang, **Shiyue Zhang**, Hyounghun Kim, and Mohit Bansal. *Continuous Language Generative Flow* ACL 2021

2020   **Shiyue Zhang**, Benjamin Frey, and Mohit Bansal. *ChrEn: Cherokee-English Machine Translation for Endangered Language Revitalization* EMNLP 2020

2020   Peter Hase, **Shiyue Zhang**, Harry Xie, and Mohit Bansal. *Leakage-Adjusted Simulatability: Can Models Generate Non-Trivial Explanations of Their Behavior in Natural Language?* Findings of EMNLP 2020

2019   **Shiyue Zhang** and Mohit Bansal. *Addressing Semantic Drift in Question Generation for Semi-Supervised Question Answering* EMNLP 2019

2018   Jiyuan Zhang, Zheling Zhang, **Shiyue Zhang**, and Dong Wang. *VV-Couplet: An open source Chinese couplet generation system* APSIPA ASC 2018

2017   **Shiyue Zhang**, Pengtao Xie, Dong Wang, and Eric P. Xing. *Medical Diagnosis From Laboratory Tests by Combining Generative and Discriminative Learning* arXiv 2017

2017   Yang Feng, **Shiyue Zhang**, Andi Zhang, Dong Wang, and Andrew Abel. *Memory-augmented Neural Machine Translation* EMNLP 2017

2017   **Shiyue Zhang**, Gulnigar Mahmut, Dong Wang, and Askar Hamdulla. *Memory-augmented Chinese-Uyghur Neural Machine Translation* APSIPA ASC 2017

2017   Aodong Li, **Shiyue Zhang**, Dong Wang, and Thomas Fang Zheng. *Enhanced Neural Machine Translation by Learning from Draft* APSIPA ASC 2017

2017   Jiyuan Zhang, Yang Feng, Dong Wang, Yang Wang, Andrew Abel, **Shiyue Zhang**, and Andi Zhang. *Flexible and Creative Chinese Poetry Generation Using Neural Memory* ACL 2017

Professional Services

SESSION CHAIR	AMTA 2022
REVIEWER	ACL Rolling Review
CONFERENCE REVIEWER	ACl 2023, EMNLP 2021/2022, COLING 2022, AAAI 2020/2022, AKBC 2020/2022
JOURNAL REVIEWER	Natural Language Engineering (JNLE)
WORKSHOP REVIEWER	SRW@ACL 2020/2021, Eval4NLP 2020/2021, SRW@EACL 2023

Selected Scholarships and Awards

2021	Bloomberg Data Science PhD Fellowship
2015	Excellent Graduate
2014, 2013	National Scholarship

Skills

Deep Learning Frameworks:	PyTorch, TensorFlow
Programming Languages:	Python
Languages:	Chinese (Native), English