Xinyue (Sherry) Chen

+1 412-251-2831
sherryicss@gmail.com
San Jose, CA 95128
Google Scholar

EDUCATION

Carnegie Mellon University (CMU) MAY 2023

- Master of Computational Data Science, School of Computer Science
- Selected Courses: Machine Learning with Large Datasets, Database Systems, Deep Learning Systems (*Implementing ML framework with C++ and CUDA*), Cloud Computing, Parallel Programming

Shanghai Jiao Tong University (SJTU) JUNE 2020

- Bachelor of Engineering in COMPUTER SCIENCE AND TECHNOLOGY

PUBLICATIONS

1. Revisiting the Role of Language Priors in Vision-Language Models
   Preprint
   Xinyue Chen*, Zhiqiu Lin*, Deepak Pathak, Pengchuan Zhang, Deva Ramanan

2. Hybrid Transducer and Attention based Encoder-Decoder Modeling for Speech-to-Text Tasks
   in Proc. of **ACL 2023** (Outstanding Paper)
   Yun Tang, Anna Y. Sun, Hirofumi Inaguma, **Xinyue Chen**, Ning Dong, Xutai Ma, Paden D. Tomasello, Juan Pino

3. Scalable Multi-Hop Relational Reasoning for Knowledge-Aware Question Answering.
   in Proc. of **EMNLP 2020** | paper | code
   Xinyue Chen*, Yanlin Feng*, Jun Yan, Bill Yuchen Lin, Xiang Ren

4. KagNet: Knowledge-Aware Graph Networks for Commonsense Reasoning.
   in Proc. of **EMNLP-IJCNLP 2019** (Oral Presentation)
   Bill Yuchen Lin, **Xinyue Chen**, Jamin Chen, Xiang Ren

EXPERIENCE

Query Understanding in TikTok Search OCT. 2023 - PRESENT
Machine Learning Engineer *TikTok*

- **Owner of query correction** for TikTok Search across all channels (general search and all vertical searches).
- Enhanced the query correction pipeline, with a diverse array of models, including LLMs, compact neural models, and tree-based models, to deliver improved search experience for *multi-lingual* user interactions in over 20 languages. Employed various data mining techniques.
- Boosting accuracy and reliability of models within online serving resource constraints.

Compositionality Reasoning of Vision-Language Models DEC. 2022 - MAY 2023
Research Assistant, Advisor: Prof. Deva Ramanan, Pengchuan Zhang *CMU, Meta*

- Designed a method that leverages multi-modal generative models for image-text matching that requires compositionality understanding. The proposed method surpasses previous SOTAs on visio-linguistic compositionality benchmarks and serves as a quantitative diagnostic tool for unimodal bias of benchmarks.

Transducer-Based Models for Simultaneous Speech Translation MAY 2022 - AUG. 2022
Research Engineer Intern, Advisor: Juan Pino, Ning Dong *Meta (FAIR)*

- Implemented high-quality transducer-based streaming speech translation systems based on cutting-edge papers from scratch in Fairseq library.
- Benchmarked and significantly improved the performance of the implemented systems both in terms of translation quality and streaming latency.

Relational Reasoning for Natural Language Understanding JULY 2019 - MAY 2020
Advisor: Prof. Xiang Ren *University of Southern California*

- Designed Graph Relation Network, a variant of GNNs capable of performing higher-order message passing over multi-relational knowledge graphs. This enables a system that incorporates LLMs and static knowledge graph, facilitating relational reasoning in natural language question answering tasks. The proposed model surpasses existing knowledge-augmented methods on CommonsenseQA and OpenbookQA datasets.

SKILLS

Programming languages: Python, C/C++, CUDA, Java, Verilog
Libraries: PyTorch, TensorFlow, MySQL, PySpark, NumPy, Pandas, Kubernetes