WEIJIA XU

weijia@cs.umd.edu
University of Maryland
College Park, MD 20740

RESEARCH INTERESTS

Text Generation, Planning and Reasoning, Human-AI Interaction

EDUCATION

University of Maryland, College Park *2017 - 2022*
Degree: Ph.D. in Computer Science GPA: 4.0/4.0
Advisor: Marine Carpuat

University of Science and Technology of China *2013 - 2017*
Degree: B.Eng. in Computer Science and Technology GPA: 9.0/10

EXPERIENCE

Microsoft Research Redmond, USA
Senior Researcher *Feb 2023 - Present*
· **Manager:** Bill Dolan
· Grounded text generation
· Learning from human-AI interactions
· AI-based agents in game environments

CLIP Lab, University of Maryland College Park, USA
Graduate Research Assistant *March 2018 - 2022*
· **Advisor:** Marine Carpuat
· Low-resource neural machine translation
· Controllable and interpretable text generation

Facebook AI Research New York, USA
Research Intern *June 2021 - December 2021*
· **Mentor:** Jiatao Gu
· **Topic:** Diffusion Models for Text Generation

Microsoft Research Beijing, China
Research Intern *June 2020 - December 2020*
· **Mentors:** Dongdong Zhang, Shuming Ma
· **Topic:** Multilingual Neural Machine Translation

Amazon AI Palo Alto, USA
Research Intern *May 2019 - August 2019*
· **Mentors:** Batool Haider, Saab Mansour
· **Topic:** Cross-lingual language understanding

PUBLICATIONS

[1] Understanding and Detecting Hallucinations in Neural Machine Translation via Model Introspection.
Weijia Xu, Sweta Agrawal, Eleftheria Briakou, Marianna J. Martindale, Marine Carpuat
TACL 2023

[2] Rule-based Morphological Inflection Improves Neural Terminology Translation.
Weijia Xu, Marine Carpuat
EMNLP 2021

[3] Improving Multilingual Neural Machine Translation with Auxiliary Source Languages.
Weijia Xu, Yuwei Yin, Shuming Ma, Dongdong Zhang and Haoyang Huang
EMNLP Findings 2021

[4] How Does Distilled Data Complexity Impact the Quality and Confidence of Non-Autoregressive Machine Translation?
Weijia Xu, Shuming Ma, Dongdong Zhang and Marine Carpuat
ACL Findings 2021

[5] A Non-Autoregressive Edit-Based Approach to Controllable Text Simplification.
Sweta Agrawal, **Weijia Xu** and Marine Carpuat
ACL Findings 2021

[6] EDITOR: an Edit-Based Transformer with Repositioning for Neural Machine Translation with Soft Lexical Constraints.
Weijia Xu, Marine Carpuat
TACL 2021 (Oral at ACL 2021)

[7] Soft Layer Selection with Meta-Learning for Zero-Shot Cross-Lingual Transfer.
Weijia Xu, Batool Haider, Jason Krone, Saab Mansour
MetaNLP at ACL 2021

[8] End-to-End Slot Alignment and Recognition for Cross-Lingual NLU.
Weijia Xu, Batool Haider, Saab Mansour
EMNLP 2020

[9] Dual Reconstruction: a Unifying Objective for Semi-Supervised Neural Machine Translation.
Weijia Xu, Xing Niu, Marine Carpuat
EMNLP Findings 2020

[10] Differentiable Sampling with Flexible Reference Word Order for Neural Machine Translation.
Weijia Xu, Xing Niu, Marine Carpuat
NAACL 2019 (Oral)

[11] Bi-Directional Differentiable Input Reconstruction for Low-Resource Neural Machine Translation.
Xing Niu, **Weijia Xu**, Marine Carpuat
NAACL 2019

[12] The University of Maryland's Chinese-English Neural Machine Translation Systems.
Weijia Xu, Marine Carpuat
WMT 2018

HONORS AND AWARDS

Ann G. Wylie Dissertation Fellowship, University of Maryland, 2022

Rising Stars (Excellent Intern Award), Microsoft Research Asia, 2020

Dean's Fellowship, University of Maryland, 2017-18

Honorable Student Title, University of Science and Technology of China, 2016-17

SERVICE AND LEADERSHIP

Program Chair
· Widening Natural Language Processing (WiNLP)   2021 - Present
The WiNLP workshop aims to foster an inclusive and diverse ACL environment by highlighting the work of underrepresented groups or anyone who self-identifies within an underrepresented demographic.

Reviewer
· Annual Conference of the Association for Computational Linguistics   2020 - Present
· Empirical Methods in Natural Language Processing   2021

TEACHING EXPERIENCE

Guest Lectures
· **CMSC828I** Non-Autoregressive Machine Translation   Spring 2021

Teaching Assistantship
· **CMSC216** Introduction to Computer Systems   Fall 2017
· **CMSC320** Introduction to Data Science   Spring 2018

SKILLS

Programming Languages
Python, C/C++, Java

Tools and Libraries
PyTorch, MxNET Symbol, MxNET Gluon

Human Languages
Mandarin Chinese (native), English (fluent)