

Contact	4902 Forbes Ave Gates Hillman Complex, Pittsburgh, PA	Homepage: https://xuhuiz.com/ ✉ E-mail: xuhuiz@cs.cmu.edu Tel: 206-306-5850
Research Interests	Natural Language Processing: Socially intelligent and equitable NLP, explainable NLP. Other: human-computer interaction, multimodal machine learning, theory of mind.	
Education	Carnegie Mellon University , Pittsburgh, PA PhD in Computer Science (Language Technologies). Aug 2022 University of Washington , Seattle, WA M.Sc in Computational Linguistics. Sep 2019–Jun 2021 Advisor: Noah Smith Nanjing University , Nanjing, China B.Sc in Statistics, Department of Mathematics Sep 2015–Jun 2019 Advisor: Shujian Huang University of California Berkeley (visiting student), Berkeley, CA Aug 2017-May 2018	
Publications (* Equal contribution)	<p>Don't Take This Out of Context! On the Need for Contextual Models and Evaluations for Stylistic Rewriting Akhila Yerukola, Xuhui Zhou, Maarten Sap <i>Preprint, 2023</i></p> <p>Clever Hans or Neural Theory of Mind? Stress Testing Social Reasoning in Large Language Models Natalie Shapira, Mosh Levy, Hossein Seyed Alavi, Xuhui Zhou, Yejin Choi, Yoav Goldberg, Maarten Sap, and Vered Shwartz <i>Preprint, 2023</i></p> <p>Cobra 🐍 Frames: Contextual Reasoning about Effects and Harms of Offensive Statements Xuhui Zhou, Hao Zhu, Akhila Yerukola, Thomas Davidson, Jena D. Hwang, Swabha Swayamdipta, Maarten Sap <i>Findings of ACL, 2023</i></p> <p>Learning to translate by learning to communicate C.M. Downey*, Xuhui Zhou*, Leo Z. Liu, Shane Steinert-Threlkeld <i>Preprint, 2023</i></p> <p>Annotators with Attitudes: How Annotator Beliefs And Identities Bias Toxic Language Detection Maarten Sap, Swabha Swayamdipta, Laura Vianna, Xuhui Zhou, Yejin Choi, Noah A. Smith <i>NAACL, 2022</i></p> <p>Emergent Communication Fine-tuning (EC-FT) for Pretrained Language Models Shane Steinert-Threlkeld, Xuhui Zhou, Zeyu Liu, C. M. Downey <i>ICLR EmeCom, 2022</i> 🏆 Runner-up Best Paper</p> <p>Extracting and Inferring Personal Attributes from Dialogue Zhilin Wang, Xuhui Zhou, Rik Koncel-Kedziorski, Alex Marin, Fei Xia <i>ACL ConvAI, 2022</i></p> <p>Challenges in Automated Debiasing for Toxic Language Detection Xuhui Zhou, Maarten Sap, Swabha Swayamdipta, Noah A. Smith, Yejin Choi <i>EACL, 2021</i></p> <p>Linguistically-Informed Transformations (LIT): A Method for Automatically Generating Contrast Sets Chuanrong Li*, Lin Shengshuo*, Zeyu Liu*, Xinyi Wu*, Xuhui Zhou*, Shane Steinert-Threlkeld <i>EMNLP BlackboxNLP, 2020</i></p>	

Multilevel Text Alignment with Cross-Document Attention

Xuhui Zhou, Nikolaos Pappas, Noah A. Smith

EMNLP, 2020

Evaluating Commonsense in Pre-trained Language Models

Xuhui Zhou, Yue Zhang, Leyang Cui, Dandan Huang

In *Proceedings of AAAI*, 2020

RPD: A Distance Function Between Word Embeddings

Xuhui Zhou, Zaixiang Zheng, Shujian Huang

ACL Student Research Workshop, 2020

Research
Experience

Machine Intelligence @ Apple

Research Intern

Mentors: [Sarthak Garg](#), [Shayne Longpre](#), [Joachim Daiber](#)

March 2021–Sep 2021

Noah's Ark @ University of Washington

Research assistant

Mentors: [Maarten Sap](#), [Swabha Swayamdipta](#), [Nikolaos Pappas](#)

Sep 2019–June 2021

CLMBR @ University of Washington

Research assistant

Mentor: [Shane Steinert-Threlkeld](#)

Jan 2019–Now

Text Intelligence Lab @ Westlake University

Research assistant

Mentor: [Yue Zhang](#)

May 2019–Sep 2019

SUTD NLP @ Singapore University of Technology and Design

Research intern

Mentor: [Yue Zhang](#)

Jun 2018–Aug 2018

DSS @ University of California, Berkeley

Research committee member

Jun 2017–Apr 2018

Service

Workshop Organizing: Theory-of-Mind Workshop at ICML 2023

Program Committee & Reviewing:

Journals & Conferences

- [TLMR](#) 2023
- [ACL](#) 2021, 2023
- [ACL ARR](#) 2021-2023
- [NeurIPS](#) 2023

Workshops

- Workshop on Multimodal Content Moderation (MMCM) at CVPR 2023
- NLP4PI at ACL 2021

Volunteer: EACL 2021, EMNLP 2020, ACL2020, 2018 Singapore Symposium on NLP