Yixiao Song

✉ yixiaosong@umass.edu   •   🌐 yixiao-song.github.io

5th-year Ph.D student at UMass Amherst. Experienced in multilinguality, automatic and human evaluation of LLMs, grammar error correction/explanation, and machine translation evaluation.

Education

Ph.D Student (Computer Science*/Linguistics)
(Advised by Prof. Mohit Iyyer & Prof. Rajesh Bhatt)   3.94/4.0
    University of Massachusetts Amherst   *2019-current*
    In process of conversion

M.A. in Germanic Linguistics   1.1/6.0 (excellent)
    University of Konstanz   *2016-2018*

B.A. in German   3.71/4.0
    University of Shanghai for Science and Technology   *2011-2015*

Publications

VERISCORE: Evaluating the factuality of verifiable claims in long-form text generation
Yixiao Song, Yekyung Kim, and Mohit Iyyer   **EMNLP 2024 Findings**
- Proposed an automatic metric for factuality evaluation of long-form model generations
- The metric effectively distinguishes verifiable and unverifiable claims which earlier metrics are not able to.
- The metric is effectively implemented with either closed or fine-tuned open-weight language models.

GEE! Grammar Error Explanation with Large Language Models   **NAACL 2024 Findings**
Yixiao Song, Kalpesh Krishna, Rajesh Bhatt, Kevin Gimpel, and Mohit Iyyer
- Proposed a two-step pipeline for generating grammar error explanation in natural language
- Utilized atomic edit extraction to guide the GEE generation to increase recall and precision
- Showed the high performance of the pipeline in German and Chinese GEE

A Critical Evaluation of Evaluations for Long-form Question Answering   **ACL 2023**
Fangyuan Xu*, Yixiao Song*, Mohit Iyyer, and Eunsol Choi (*Equal contribution)
- Comprehensively evaluated text generation metrics on long-form open-ended question answering generation
- First expert-annotated long-form question answering dataset

Paraphrasing evades detectors of AI-generated text, but retrieval is an effective defense   **NeurIPS 2023**
Kalpesh Krishna, Yixiao Song, Marzena Karpinska, John Wieting, and Mohit Iyyer
- Introduced a paraphrase generation model DIPPER which leverages context and offers diversity control
- Stress-tested and successfully evaded major AI-generated text detectors (e.g., watermarking, GPTZero)
- Proposed a simple but effective defense that relies on retrieving semantically-similar generations

kNN-LM Does Not Improve Open-ended Text Generation   **EMNLP 2023**
Shufan Wang, Yixiao Song, Andrew Drozdov, Aparna Garimella, Varun Manjunatha, and Mohit Iyyer
- Revealed that interpolation-based retrieval-augmented LMs do not improve open-ended generation quality

SLING: Sino Linguistic Evaluation of Large Language Models   **EMNLP 2022**
Yixiao Song, Kalpesh Krishna, Rajesh Bhatt, and Mohit Iyyer
- A benchmark with 38K minimal sentence pairs in Mandarin Chinese
- Tested 18 publicly available pretrained monolingual and multi-lingual language models
- Showed that the average accuracy for LMs is far below human performance (69.7% vs. 97.1%)
- Revealed the strengths and weaknesses of large language models

DEMETR: Diagnosing Evaluation Metrics for Translation   **EMNLP 2022**
Marzena Karpinska, Nishant Raj, Katherine Thai, Yixiao Song, Ankita Gupta, and Mohit Iyyer
- A diagnostic dataset with 31K English sentences (translated from 10 source languages)
- Evaluated the sensitivity of MT evaluation metrics to 35 different linguistic perturbations
- Found that learned metrics perform substantially better than string-based ones
- Revealed the strengths and weaknesses of learned metrics

Internships

○ **Google**
Google Translate Team *June - September 2024*
Research Intern (mentored by Parker Riley, Dan Deutsch, and Markus Freitag)
Implemented in JavaScript a new template in Anthea for human evaluation
(Ongoing)
Push the limit of fine-grained human evaluation for more insights in model performance
Analyze strengths and weaknesses of evaluation methods to better guide modeling/launch decisions

○ **Quillbot, Learneo, Inc.**
Platform for grammar correction, translation and text rewriting *June - September 2023*
Research Engineer Intern (mentored by Kevin Gimpel and George Wang)
Improved German grammar error correction product
(+5% copy rate & +1.7% 1-day retention compared to previous product)

○ **Cultural Institute of the Federal Republic of Germany (Goethe Institut)**
Non-profit German cultural association, promoting the German language study abroad *August 2014 - February 2016*
Translator, interpreter, project assistant

○ **Friedrich Ebert Stiftung Shanghai Office**
Non-profit German foundation funded by the Government of the Federal Republic of Germany *April-July 2014*
Translator, project assistant

Positions of Responsibility

○ **Reviewer** for NeurIPS 2023 R0-FoMo Workshop, EACL 2024 (Outstanding Reviewer), NAACL 2024, ACL 2024, BEA 2024, ESSLLI 2024, EMNLP 2024

○ **Instructor** of LING201 at UMass Amherst: How Language Works—Introduction to Linguistic Theory (R2)

○ **Research Assistant** advised by Prof. Mohit Iyyer (Summer 2022): Human evaluation of model performance of long-form question answering.

○ **Teaching Assistant** for courses at master level (Univeristy of Konstanz) and (under)graduate level (UMass Amherst)

○ **Proceeding Editor** of NELS50 and SULA11

○ **Student Research Assistant** in Deutsche Forschungsgemeinschaft Project—Questions at the Interfaces (P3 Alternative Questions)

Works and Presentations

Mandarin Chinese Alternative Questions are not Disjoined Polar Questions
Yixiao Song *2021*
○ First qualification paper supervised by Prof. Rajesh Bhatt and Prof. Seth Cable

Early Cue Effects of Chinese Relative Clause Comprehension in Pre-trained Language Model
Yixiao Song *2021*
○ Breadth Paper for satisfying Ph.D. requirements

A Comparative Study of German and Chinese Alternative Questions
Yixiao Song *2018*
○ Poster presentation at Semantics and Philosophy in Europe
○ Talk at the 17th China International Conference on Contemporary Linguistics

Skills

○ **Natural Languages:** Shanghai Wu, Mandarin Chinese, English, German
○ **Programming:** Python (PyTorch, Hugging Face), R, JavaScript, Linux, Perl
○ **Others:** LaTeX, GitHub