

HAO ZOU

(631) 809-8931 \diamond haozou-official.github.io \diamond hz2999@columbia.edu

EDUCATION

Columbia University
MS in Computer Science, Machine Learning Track

New York, NY
Expected Dec 2025

University of Minnesota, Twin Cities
BS in Computer Science

Minneapolis, MN
Sep 2019 - May 2023

(Grad Level) Machine Learning Fundamentals, Topics in Deep Learning, Computer Vision, Biomedical NLP, Engineering Optimization, Artificial Intelligence

PUBLICATIONS

1. Hao Zou, Zae Myung Kim and Dongyeop Kang. *A Survey of Diffusion Models in Natural Language Processing*. Under review of EMNLP 2023. [[arxiv](#)]
2. Hao Zou, Karin de Langis, Dongyeop Kang and Yohan Jo. *Debiasing Language Models for In-Context Learning Using a Causal Inference-Inspired Method*. Under review of EACL 2023. [[paper](#)]
3. Saptarashmi Bandyopadhyay, Hao Zou, Chenqi Zhu, Jordan Boyd-Graber, et al. *You Make me Feel like a Natural Question: Training QA Systems on Transformed Trivia Questions*. Submitted to EMNLP 2022. [[paper](#)]
4. Saptarashmi Bandyopadhyay, Shraman Pal, Hao Zou, Jordan Boyd-Graber, et al. *Improving Question Answering with Generation of NQ-like Questions*. Submitted to MRQA 2021. [[paper](#)]

WORK EXPERIENCE

Google Research
Visual Language Models Intern, mentor: [Dr. Samira Daruki](#)

San Jose, CA
May 2023 - Aug 2023

- Performed benchmarking on ControlNet and T2I-Adapter methods across 3+ multi-model datasets, providing detailed performance metrics and insights to enhance model evaluation and development
- Integrated ControlNet into the OpenFlamingo model, optimizing performance across 3+ multi-model datasets and enhancing overall model accuracy and robustness

Sony Research
Large Language Models Intern

Minneapolis, MN
Jan 2022 - May 2023

- Authored and conducted an in-depth survey on Diffusion Models in NLP, delivering pivotal insights to advance field
- Spearheaded proposal and execution of two groundbreaking strategies for integrating diverse styles into controllable models, resulting in a remarkable 30% enhancement in targeted compositional styles on evaluation set

RESEARCH EXPERIENCE

Duke University
Research Assistant, advisor: [Dr. Enmao Diao](#)

Durham, NC
May 2024 - Present

- Led development of a benchmarking pipeline for various diffusion and sampling algorithms, targeting different prediction strategies in the reverse process (e.g., 'z_noise,' 'x0,' 'x_prev,' 'v-predefined')
- Innovated two scalable diffusion training processes and integrated triangle distribution, surpassing Gaussian, to inspire new advancements in model architectures

University of Minnesota, Twin Cities
Research Assistant, advisor: [Prof. Dongyeop Kang](#)

Minneapolis, MN
Aug 2021 - May 2023

- Developed a method driven by causal inference to measure true causal effect of input text on potential labels, boosted Pre-trained Language Models (PLMs) accuracy across various tasks by up to 22% points
- Applied de-biasing method to reduce accuracy variance in PLMs by up to 15% points, resulting in improved model consistency and reliability

University of Maryland

College Park, MD

Research Assistant, advisor: [Prof. Jordan Boyd-Graber](#)

June 2021 – June 2022

- Proposed and implemented methods to decode convoluted syntax and automatically produce information-seeking questions from longer trivia data, boosting QA system accuracy by 18% in a new, out-of-domain context
- Presented fine-grained analysis on over 2k generated questions on linguistic, grammatical, style and topic dependent features, aiming to under-stand specific attributions to better question generations for desired domain

SKILLS

Programming

Python, C/C++, Java, MATLAB, OCaml, MySQL

Tools

Pytorch, Tensorflow, Keras, FedNLP, Unix/Linux, Git, Docker, L^AT_EX, Eviews