HAO ZOU

3204 4th St SE, Minneapolis, MN 55414, United States
zou00080@umn.edu | +1 (631) 809 8931 | haozou-official.github.io

EDUCATION

UNIVERSITY OF PENNSYLVANIA
Research Assistant
2022-2023 | Philadelphia, PA

UNIVERSITY OF MINNESOTA
B.S. in Computer Science
2019-2022 | Minneapolis, MN
Cum. GPA: 3.83 / 4.0

LINKS

HomePage: haozou-official.github.io
Github: z.umn.edu/hz-git
LinkedIn: z.umn.edu/hz-linkedin
Twitter: @haozou_

RESEARCH INTERESTS

Diffusion Models
Federated Learning
Causality in AI
Natural Language Processing
• Controllable Text Generation
• Prompt Engineering
• Open Domain Question Answering
Foundational Deep Learning
Robustness

HONORS/AWARDS

2022	**Google CSRMP**[a] **Scholar**
	Mentor: **Samira Daruki** [b]
2021	**UROP**[c] **scholarship**
2019-2021	CSE Dean's List [d]
2021	Tau Beta Pi [e]

SKILLS

Programming • Python • C/C++
Java • Shell • Javascript
MATLAB • OCaml • MySQL
Tools • Pytorch • Tensorflow
Keras • FedML • FedNLP
Kubernetes • Unix/Linux • Git
LaTeX• MONAI • Docker
Eviews • Wind • Bloomberg
Mathematics • Matrix Theory
Convex Optimization • Probability Theory
Information Theory

[a] CS Research Mentorship Program
[b] CV Reference Contact:
Samira Daruki [sites.google.com/samiradaruki/]
[c] Undergrad Research Opportunities Program
[d] Semester-based honor for GPA>3.7
[e] Invitation only to top 3% of senior class

PUBLICATIONS

[1] **Hao Zou**, Karin de Langis, Dongyeop Kang [a] and Yohan Jo. ***Debiasing Language Models for In-Context Learning Using a Causal Inference-Inspired Method***. Under review of EACL 2023. [link]

[2] Saptarashmi Bandyopadhyay, **Hao Zou**, Chenqi Zhu, Jordan Boyd-Graber [b], et al. ***You Make me Feel like a Natural Question: Training QA Systems on Transformed Trivia Questions***. Submitted to EMNLP 2022. [link]

[3] Saptarashmi Bandyopadhyay, **Hao Zou**, Jordan Boyd-Graber, et al. ***Improving Question Answering with Generation of NQ-like Questions***. Submitted to MRQA 2021. [link]

RESEARCH

UNIVERSITY OF PENNSYLVANIA | Research Assistant
May 2022 – Present | Remote
Foundational Deep Learning; Adversarial Examples

- *Leading* research on **A New Perspective for Adversarial Examples via Local Elasticity**. Studying the reasons for the existence of **adversarial examples** and the properties of their tiny norms. We aim to form a reasonable theory based on the elasticity of labels during neural training, referred to as **Local Elasticity**. (advisor: **Prof. Weijie Su** [c])

UNIVERSITY OF MINNESOTA | Research Assistant
Aug 2020 – Present | Minneapolis, MN
Diffusion Models; Federated Learning; Causality in AI; NLP
Minnesota NLP Group, advisor: **Prof. Dongyeop Kang**

- *Leading* a project to adapt **Diffusion Models** for **Out-of-distribution robustness** which concerns adversarial robustness in the context of domain shifts. Utilizing the controllable generation ability as a preprocessor for zero-shot domain adaptation and the strong sample quality to purify adversarial texts. Analyzing the possibility of augmenting intermediate diffusion samples for **robust training**.
- *Leading* a project about **Covariate Shift Adaptation in Federated Learning for NLP**. We aim to pioneer disentangled representation learning in building robust **federated learning** (FL) models across various NLP problems. We aim to propose a content-style-based FL scheme to outperform baseline FedAvg in **Out-of-distribution generalizations** across different domains.
- *Led* a project on **De-biasing Large Pre-trained Language Models (PLMs) for in-context learning**. Presented a method inspired by **causal inference** to measure the **true causal effect** of the input text on possible labels more accurately and substantially increased the accuracy of PLMs in various tasks and reduces accuracy variance to achieve better robustness.

Data Mining; Political Stance Detection; NLP in Social Media
Data Management Research Group (DMR), advisor: **Prof. Jaideep Srivastava** [d]

- Conducted user-level **stance detection** for Super Bowl **Team Affiliation** using semi-supervised framework and extended binary classification by Valence Scores. Extracted Different target-related topics and eliminated noisy instances through weakly supervised learning of **Political Affiliation** on Twitter.

[a] CV Reference Contact: Prof. Dongyeop Kang [dykang.github.io]
[b] CV Reference Contact: Prof. Jordan Boyd-Graber [users.umiacs.umd.edu/ jbg]
[c] CV Reference Contact: Prof. Weijie Su [stat.wharton.upenn.edu/ suw]
[d] CV Reference Contact: Prof. Jaideep Srivastava [users.cse.umn.edu/ srivastava]

SERVICES

NEURIPS 2022 | Session Volunteer
Nov 28st - Dec 9th, 2022
EMNLP 2022 | Reviewer
Aug 1st-8th, 2022
NAACL 2022 | Session Volunteer
July 9th-13th, 2022

COURSEWORK

UNDERGRAD LEVEL

Data Structure and Algs (CSCI 4041)
Linux OS (CSCI 4061)
Artificial Intelligence (CSCI 4511W)
Prgm Graphics and Games (CSCI 4611)

GRAD LEVEL

Machine Learning (CSCI 5525)
Deep Learning (CSCI 5980)
Computer Vision (CSCI 5561)
Biomedical NLP (HINF 5610)
ML Fundamentals (CSCI 5521)
Engineering Optmization (MATH 5711)

Note: The levels of courses above are categorized based on https://cse.umn.edu/cs/grad-breadth. An undergrad is allowed to take grad courses under instructor's approval.

SELECTED COURSE PROJECT

CSCI 5525 Machine Learning: Analysis and Methods (Spring 2021)

• **A Survey of Causality in Visual Question Answering** [z.umn.edu/causalvqa], Grade: **A**

HINF 5610 Biomedical NLP (Spring 2021)

• **Text Simplification for Leadless Pacemaker Failure Reports in MAUDE** [z.umn.edu/biomednlp], Grade: **A**

STUDY GROUPS

UMN Café GNNs in NLP
Graph Neural Networks in NLP Reading Group

UMD MARL
Multi-agent Reinforcement Learning Reading Group

UNIVERSITY OF MARYLAND | Research Assistant

June 2021 – June 2022 | College Park, MD
Computational Linguistics and Information Processing (CLIP)
Open-domain Question Answering; Domain Adaptation

- *Co-led* a project about **Improving QA Systems on Transformed Trivia Questions** by transferring knowledge from a new, out-of-domain question answering dataset (QuizBowl). Proposed methods to **decode convoluted syntax** and automatically generate information-seeking questions. Presented fine-grained analysis on **linguistic, grammatical, style and topic dependent features** aiming to understand the specific attributions to better question generations for the **desired domain**. Boosted passage retriever by adapting it to new domains.
  (advisor: Prof. Jordan Boyd-Graber)

UNIVERSITY OF ROCHESTER, NEW YORK | Research Assistant

June 2020 –Oct 2020 | Remote
Signal Processing and Information Systems Laboratory (SPIS)
Signal Processing; EEG signal

- Implemented various Deep Learning models (e.g. LSTM with AutoEncoder, parallel CNN and RNN) on DEAP dataset resulting in better emotional classification based on EEG signals. Compared different methods for data preprocessing including shuffling and considering the base signal.
  (advisor: Prof. Mujdat Cetin [a])

EMPLOYED EXPERIENCE

ZHONGDA INNOVATION VALLEY INC. | Quant Finance Intern

Sep 2018 - Dec 2018 | Shenzhen, CHN

- Finalized analysis of factor performance by Alphalens and Signaldigger in Python and processed data in Ta-lib and JAQS
- Implemented multi-factor combination and visualization to create financial trading report
- Proposed new signal hypothesis by customizing ANN factors and implemented the assumptions to corroborate the efficiency of the hypothesis

AXA HONG KONG | Corporate and Industry Research Intern

Feb 2018 – May 2018 | Hong Kong, CHN

- Deployed decentralized investments to reduce risk and the Markowitz model to calculate optimal industry weights based on historical data
- Created stock price prediction model using CAPM and WACC; developed and presented Star Price Stock report and presented findings
- Engaged in marco-economic research, collecting and sorting global financial news for investment strategy adjustments

TEACHING & TUTORING

UNIVERSITY OF MARYLAND | HS[b] Research Mentoring Assist[c]

June 2021 – Dec 2021 | Remote

- Mentored high school interns on research projects, tutored them with basic research skills such as documentation writing, research question framing, hands on coding, etc.

UNIVERSITY OF MINNESOTA LIBRARY | Peer Tutor

Sep 2020 – Dec 2020 | Minneapolis, MN

- Tutored single-and-multivariable calculus, linear algebra, intro physics, intro stats and some programming in Python and C.

[a] CV Reference Contact: Prof. Mujdat Cetin [rochester.edu/ece/cetin_mujdat]
[b] High School students
[c] Computational Linguistics and Information Processing Lab (CLIP)