Haiyan Zhao

Tel: +1 2017260517 • hz54@njit.edu • hy-zhao23.github.io

Education

New Jersey Institute of Technology(NJIT) **2023 – Now**
Ph.D. in Computer Science, Advisor: Dr. Mengnan Du *NJ, US*

Huazhong University of Science and Technology(HUST) **2017 – 2019**
M.E. in Control Engineering, Advisor: Dr. Xiaolong Shi *Wuhan, China*
Weighted average score: 90.63/100 (Rank:10/111, Top 9%)

Wuhan University of Technology(WUT) **2013 – 2017**
B.E. in Automation *Wuhan, China*
Weighted average score: 89/100 (Rank: 9/229, Top 3.9%)

Research Interests

- Model Interpretability, Trustworthy AI, Large Language Models

Publications

H. Zhao, H. Chen, F. Yang, N. Liu, H. Deng, H. Cai, S. Wang, D. Yin, and M. Du, "Explainability for large language models: A survey," *arXiv preprint arXiv:2309.01029*, 2023, [**Accepted with minor revision by ACM TIST**].

Z. He, H. Deng, **H. Zhao**, N. Liu, and M. Du, "Mitigating shortcuts in language models with soft label encoding," *arXiv preprint arXiv:2309.09380*, 2023.

C. Chen, J. Xu, L. Ruan, **H. Zhao**, X. Li, and X. Shi, "Dna origami frame filled with two types of single-stranded tiles," *Nanoscale*, vol. 14, no. 14, pp. 5340–5346, 2022.

Q. Liang, N. Bie, T. Yong, K. Tang, X. Shi, Z. Wei, H. Jia, X. Zhang, **H. Zhao**, W. Huang *et al.*, "The softness of tumour-cell-derived microparticles regulates their drug-delivery efficiency," *Nature biomedical engineering*, vol. 3, no. 9, pp. 729–740, 2019.

X. Shi, **H. Zhao**, X. Li, and T. Song, "Isothermal approach to assemble spatial dna nanotubes for drug delivery," *Oncotarget*, vol. 5, 2018.

Invited Talk

2023/10/28: Invited talk @ Talk on MLLM-AI on Explainability in LLMs

2023/10/25: Invited talk @ UberAI on Explainability in LLMs

2023/10/11: Invited talk @ NJIT FinTech Seminar on Explainability in LLMs

Experience

Teaching Assistant(NLP) **Prof. Mengnan Du**
New Jersey Institute of Technology *2023 – 2024*

Research Intern(Remote) **Prof. Jianxin Li**
Deakin University *2022.2 – 2023.2*
- Error-Aware Probabilistic Logical Reasoning Model for Entity Alignment.

Freelancer
2019.9 – 2021.12

- Designed LaTeXtemplate for academic journals.

Honors & Awards

2018 – 2019, **National Scholarship**(Ministry of Education of P.R. China); **Outstanding Graduate, Merit Postgraduate**(HUST)

2017 – 2018, **First-class academic scholarships for postgraduates**(HUST)

2016 – 2017, **Outstanding Graduate Thesis**(Hubei)

2015 – 2016, **First-class University Scholarship**(WUT); **Merit Student Leader**(TA, WUT)

2014 – 2015, **National Scholarship**(Ministry of Education of P.R. China); **Merit Student**(WUT)

2013 – 2014, **Third-class University Scholarship**(WUT)