

EDUCATION

- **University of Georgia, Athens, US**
Ph.D. in Computer Science Advisor: Prof. Ninghao Liu Jan. 2022 - Present
- **North China Electric Power University, Beijing, China**
B.Eng. and M.S. in Renewable Energy Science and Engineering Sep. 2014 - Jun. 2021

RESEARCH INTERESTS

- **Natural Language Processing:** retrieval augmented generation (RAG), model editing.
- **Multimodality:** vision foundation models, multimodal editing.
- **Trustworthy Machine Learning:** foundation models interpretability, malicious ML attack defense.

PUBLICATIONS

Conference and Preprints

- Retrieval-enhanced Knowledge Editing for Multi-hop Question Answering in Language Models.
– Yucheng Shi, et al. (*CIKM*), 2024.
- MKRAG: Medical Knowledge Retrieval Augmented Generation for Medical Question Answering.
– Yucheng Shi, et al. (*AMIA*), 2024.
- Usable XAI: 10 Strategies Towards Exploiting Explainability in the LLM Era.
– Xuansheng Wu*, Haiyan Zhao*, Yaochen Zhu* Yucheng Shi*, et al. (*arXiv*), 2024.
- Quantifying Multilingual Performance of Large Language Models Across Languages.
– Zihao Li, Yucheng Shi, et al. (*arXiv*), 2024.
- Could Small Language Models Serve as Recommenders? Towards Data-centric Cold-Start Recommendation.
– Xuansheng Wu, Huachi Zhou, Yucheng Shi, et al. (*WWW*), 2024.
- Automated Natural Language Explanation of Deep Visual Neurons with Large Models.
– Chenxu Zhao, Wei Qian, Yucheng Shi, et al. (*AAAI, student abstract*), 2024.
- Chatgraph: Interpretable Text Classification by Converting Chatgpt Knowledge to Graphs.
– Yucheng Shi, et al. *ICDM Workshop (ICDMW)*, 2023.
- Black-box Backdoor Defense via Zero-shot Image Purification.
– Yucheng Shi, et al. (*NeurIPS*), 2023.
- GiGaMAE: Generalizable Graph Masked Autoencoder via Collaborative Latent Space Reconstruction.
– Yucheng Shi, et al. (*CIKM*), 2023.
- ENGAGE: Explanation Guided Data Augmentation for Graph Representation Learning.
– Yucheng Shi, et al. (*ECML-PKDD*), 2023.
- Interpretation of Time-Series Deep Models: A Survey.
– Ziqi Zhao*, Yucheng Shi*, et al. (*arXiv*), 2023.

Journal Publications

- Co-supervised learning paradigm with conditional generative adversarial networks for sample-efficient classification.
– Hao Zhen, Yucheng Shi, et al. *Applied Computing and Intelligence*, 2023
- Expected output calculation based on inverse distance weighting and its application in anomaly detection of distributed photovoltaic power stations.
– Yucheng Shi, et al. *Journal of Cleaner Production (IF=11.1)*, 2020.

- New feature extraction method for photovoltaics array output time series and its application in fault diagnosis.
– Honglu Zhu, **Yucheng Shi**, et al. *IEEE Journal of Photovoltaics* (IF=3.57), 2020.
- Uncertainty analysis based on non-parametric statistical modeling method for photovoltaic array output and its application in fault diagnosis.
– Jingna Pan, Weiguo He, **Yucheng Shi**, et al. *Solar Energy* (IF=7.188), 2021.
- Photovoltaic power forecasting based on GA improved Bi-LSTM in microgrid without meteorological information.
– Hao Zhen, Dongxiao Niu, Keke Wang, **Yucheng Shi**, et al. *Energy* (IF= 9.0), 2021.

RESEARCH EXPERIENCE

- **Data Scientist Intern at Harvard Medical School (Remote)** May. 2024 - Present
– *Medical Foundation Models Pre-training and Supervised Fine Tuning.*
- **Research Assistant at University of Georgia** Jan. 2022 - Present
– *Detect and mitigate harmful misalignments utilizing LLMs interpretation.*
– *Member of the NSF Project: Graph-Oriented Usable Interpretation.*
- **Research Assistant at North China Electric Power University** Sep. 2018 - Jun. 2021
– *Anomaly detection for PV arrays.*

TEACHING & TALKS

- **Teaching Assistant at University of Georgia** Jan. 2023 - Present
– *Undergraduate/graduate course: CSCI4380/6380 Data Mining.*
- **Teaching Assistant at University of Georgia** Aug. 2022 - Dec. 2022
– *Undergraduate/graduate course: CSCI4360/6360 Data Science.*
- **Talk at Harvard Medical School AIXMed Seminar** Aug. 2023
– *Topic: LLMs editing with external knowledge graphs for medical QA.*

AWARDS & ACTIVITIES

- **NeurIPS 2023 Scholar Award.** – (10/2023)
- **China National Scholarship** – (09/2019-06/2020)
- **Pacemaker to Graduate Student(top 0.8%)** – (09/2019-06/2020)
- **First-class Scholarship** – (09/2018-06/2019), (09/2019-06/2020).

SERVICE

Reviewer

- ACM Transactions on Knowledge Discovery from Data (TKDD)
- ACM TheWebConf 2024 Conference, 2024
- Neural Networks

External Reviewer

- The 17th ACM International Conference on Web Search and Data Mining (WSDM), 2024
- The 32nd ACM International Conference on Information and Knowledge Management (CIKM), 2023
- European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, 2023

SKILLS

- **Skills:** Python, Linux, Matlab, MySQL