

Qin Shi

✉ shi622@purdue.edu • 🏠 shqii1j.github.io

Education

Purdue University

Doctor of Philosophy in Statistics

GPA – 4.0/4.0

Indiana, IN

2023–present

National University of Singapore

Master of Statistics

GPA – 4.67/5.0

Singapore

2022–2023

Shanghai University of Finance and Economics

Bachelor of Science in Data Science

GPA – 3.5/4.0

Shanghai, China

2018–2022

Publications

- [1] **Qin Shi***, Amber Yijia Zheng*, Qifan Song, and Raymond A. Yeh. Knowledge distillation detection for open-weights models. *Advances in Neural Information Processing Systems (NeurIPS)*, 2025.
- [2] Jiawei Du, **Qin Shi**, and Joey Tianyi Zhou. Sequential subset matching for dataset distillation. *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.

Research & Work Experience

Purdue University — Department of Statistics

Research Assistant

Mentor: Prof. Qifan Song

West Lafayette, IN

July 2024 - Present

- Researched knowledge distillation in Image-classification models, Text-to-image diffusion/flow matching models and Large Language Models.
- Applied data distillation to vision-language dataset and developed an efficient algorithm to improve Vision-Language tasks with better understanding.
- Researched understanding of LLMs from a statistical perspective and their security vulnerabilities.

A*STAR's Institute of High-Performance Computing

Research Intern

Mentor: Prof. Joey Tianyi Zhou, Dr. Jiawei Du

Singapore

Sept. 2022 - June 2023

- Proposed a novel and generalized training scheme of data distillation (DD) to improve the performance of the gradient-matching-based DD method.
- Addressed the coupling issue within the synthetic dataset, which previously limited its ability in effectively condense additional high-level features.
- The proposed training scheme is verified to improve several gradient-matching-based DD methods and achieve state-of-the-art in various datasets.

YumChina

Data Scientist Intern

Shanghai, China

Apr. 2022 - June 2022

- Built multi-objective recommendation system in a large-scale dataset and successfully applied it in practice. The model's off-line test set AUC reached 0.73.
- Built deep structured semantic model to generate combination-meal template based on user's food purchasing behavior.

MarcPoint

NLP Algorithm Engineer Intern

Shanghai, China

Sep. 2021 – Jan. 2022

- Proposed an algorithm to remove meaningless comments from social media. The algorithm could achieve 80% filter rate in a human-labeled test set and 4 min/million pieces of speed, which has been put into use.
- Built ACNN-CRF model to implement named entity recognition (NER) on social media data. Compared to Bert-CRF model, the running speed of this model was improved by 11%.
- Participated in the project of "MEMO: Rethinking Named Entity Recognition Evaluation Metrics, Datasets and Models". Proposed a new method of NER dataset partition and a new evaluation measure method for MER models.

Academic Services

Conference Reviewer: European Conference on Computer Vision (ECCV), Neural Information Processing Systems (NeurIPS), International Conference on Computer Vision (ICCV)

Other Skills

Technical Skills: Pytorch, Python, R-Programming, SQL, C Language