Shenqi Zhou

shenqi_zhou@zju.edu.cn | shenqi-zhou.github.io

Education

Zhejiang University Sep. 2023 - Jun. 2027 (expected)

Bachelor of Engineering in Biomedical Engineering

- GPA: **4.11/4.3** | Average: **91.12/100**

Research Interests

My research interests focus on AI for healthcare and the evaluation of LLMs and agents. My current project, Auto-MD-bench, studies whether a single performance metric is sufficient to guide autoresearch agents toward models for real-world medical use.

Manuscript Under Review

Shenqi Zhou[*], Qiran Zou[*], Wenhao Zhao, Zhenyu Liao, and Dianbo Liu. *Auto-MD-bench: Evaluating Autoresearch Agents for Real-World Medical Tasks.* Under review. arXiv:2610.25069. [*]Equal contribution.

Research Experience

Auto-MD-bench: Evaluating Autoresearch Agents for Real-World Medical Tasks

Research Intern, ***National University of Singapore*** */ Adviser:* ***Prof. Dianbo Liu*** *Aug. 2026 - Present*

- Led an independent research project investigating whether optimizing a single performance metric is sufficient to guide autoresearch agents toward models for real-world medical use.
- Designed and built Auto-MD-bench, covering 15 tasks across survival analysis, classification, and sequential prediction, with external validation across patient populations.
- Evaluated four autoresearch agents beyond primary task performance, analyzing risk calibration, performance across patient populations, and the agents' modeling choices.
- Found that primary task performance alone was insufficient to assess autonomous clinical modeling: agents achieved competitive scores, yet calibration differed substantially at similar scores, and performance varied across patient populations.

Evaluating Chinese Sentence Understanding in Large Language Models

Undergraduate Researcher, ***Zhejiang University*** */ Adviser:* ***Prof. Nai Ding*** *Apr. 2025 - Mar. 2026*

- Led a team of three students to investigate how well LLMs understand Chinese sentences and how their sentence spaces compare with those derived from human judgments.
- Built a dataset of 370 Chinese sentences and collected entailment judgments from human participants and three LLMs.
- Evaluated the resulting sentence spaces using three Sentence-Space Metrics: sparsity, clustering of semantically related sentences, and similarity to human sentence spaces.
- Found that LLMs agreed closely with humans on semantic relatedness but showed much weaker agreement on entailment direction.

Relevant Coursework

Mathematics and statistics: Probability and Mathematical Statistics (94); Linear Algebra (A) (93); Calculus (A) II (93); Complex Variable Functions & Integral Transformation (92); Error Theory and Data Processing (99).
Computing: Advanced Programming (100); Python Programming (H) (98); Lab & Fundamentals of C Programming (95); Introduction to Pattern Analysis and Machine Learning (93); Data Structures and Algorithm Analysis (90).
Biomedical engineering: Biomedical Signal Processing (96); Biomedical Imaging Technology (95); Engineering Physiology (92).

Honors and Awards

Zhejiang Provincial Government Scholarship | 2026
Bioer Scholarship, Zhejiang University | 2024-2025
Chinese Mathematics Competitions, Provincial Second Prize | 2024
Zhejiang College Physics Innovation Competition, Second Prize | 2024
Chinese Mathematical Olympiad (Preliminary), First Prize | 2022

Skills and Languages

Programming: Python, C.
AI tools: Proficient with coding agents.
Languages: Chinese (native); English (IELTS Academic: 7.0/9, July 2026).