Zhecheng Sheng

516 Delaware Street SE, Minneapolis, MN, 55455

919-519-7176 sheng136@umn.edu github.com/zcsheng95 epsilonzc.github.io

Research Interest

Trustworthy machine learning My current research interests focus on developing methods to ensure trustworthiness of language models and their applications in finding clinical solutions. This involves addressing the presence of sensitive attributes or confounding shifts that may interfere with model performances and build unified evaluation principles. Besides, I am interested in inspecting the inner workings of language models and finding causal mechanism about model's behaviors.

Education

University of Minnesota, Twin Cities Minneapolis, MN *Aug. 2021 – Mar. 2026 (Est.)*

Ph.D. in Health Informatics & Data Science
M.Sc. in Data Science, GPA: 3.99/4.00
Advisor: Serguei Pakhomov

Duke University Durham, NC *Aug. 2017 – May. 2019*

M.Sc. in Biostatistics, GPA: 3.82/4.00
Advisor: Benjamin Goldstein

Selected Publication

Sheng, Z., Ding, X., Hur, B., Li, C., Cohen, T., Pakhomov. S. (2025). Mitigating Confounding in Speech-Based Dementia Detection through Weight Masking. *In Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (ACL). Vienna, Austria.*

Sheng, Z., Zhang, J., & Diao, E. (under review). Toward unifying group fairness evaluation from a sparsity perspective.

Zhang, T.*, **Sheng, Z.***, Lin, Z.*, Jiang, C.*, & Kang, D. (2025). BBScoreV2: Learning time-evolution and latent alignment from stochastic representation. *In Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing (EMNLP). Suzhou, China. (* equal contributions)*

Ding, X., **Sheng, Z.**, Hur, B., Tauscher, J., Ben-Zeev, D., Yetişgen, M., Pakhomov. S. & Cohen, T. (2025). Tailoring task arithmetic to address bias in models trained on multi-institutional datasets. Journal of Biomedical Informatics, 104858.

Sheng, Z.*, Zhang, T.*, Jiang, C.*, Kang, D. (2024). BBScore: A Brownian Bridge Based Metric for Assessing Text Coherence. *In Proceedings of the 38th Annual AAAI Conference on Artificial Intelligence (AAAI)., pages 14937-14945. Vancouver, Canada. (* equal contributions)*

Li, C., **Sheng, Z.**, Cohen, T., & Pakhomov, S. V. S. (2024). Too Big to Fail: Larger Language Models are Disproportionately Resilient to Induction of Dementia-related Linguistic Anomalies. *In Findings of the 62nd Annual Meeting of the Association for Computational Linguistics (ACL). Bangkok, Thailand*

Sheng, Z., Finzel, R., Lucke, M., Dufresne, S., Gini, M.,& Pakhomov, S. (2023). A Dialogue System for Assessing Activities of Daily Living: Improving Consistency with Grounded Knowledge. *In Proceedings of the Third DialDoc Workshop on Document-grounded Dialogue and Conversational Question Answering, pages 68–79. Toronto, Canada. Association for Computational Linguistics.*

Research Experience

Deconfound Deep Transformers Networks *Sep. 2022 – Present*

Mentor: Serguei Pakhomov, University of Minnesota Twin Cities

- Investigated language model's causal reasoning ability in dementia classification results and located weights within the deep network that associated with target labels.
- Proposed model weight filtering strategies that effectively reduce group disparities and improve fairness.
- Developed an probabilistic evaluation framework to quantify distribution shift within subgroups.
- Developed a novel method that uses task vector under soft prompt space to reduce bias from a finetuned model.

Dialogue System for Activity of Daily Living Assessments *July. 2022 – Present*

Mentor: Serguei Pakhomov, University of Minnesota Twin Cities

- Implemented transformer models like RoBERTa for text classification and improved classification accuracy by 10%.
- Fine-tuned Large Language Models such as LLaMA and Vicuna with clinical assessment notes for domain adaptation through parameter efficient methods(LoRA) and improved the usability of the system.
- Built a knowledge base for the dialogue system which reduces hallucinations from large language models and made the system more factual grounded.

Assess Text Coherence with Brownian Bridge *Jan. 2023 – Present*

Mentor: Dongyeop Kang, University of Minnesota Twin Cities

- Introduced a likelihood-based evaluation score motivated by Brownian Bridge for measuring global text coherence.
- Demonstrated the proposed score can improve classification performance from previous SOTA in global artificial tasks without end-to-end training.
- Proposed an likelihood objective for language encoder training that is able to capture latent dependencies.
- Elaborated the proposed score can be sued to compare text with different length both empirically and theoretically.
- Showecased the score can be adapted for diverse downstream tasks, such as distinguishing between human and AI-generated text and detecting different LLM generated text from mixed corpus.

An Unified Machine Learning Fariness Notion via Sparsity *May. 2023 – Present*

Mentor: Enmao Diao, Duke University

- Proposed a novel machine learning fairness notion based on distributional difference and sparsity.
- Proposed an unified fairness evaluation measurements for both classification and regression problem.
- Implemented augmented ADMM in optimization for linear models with designated fairness constraint and provided theoretical guarantees.

Work Experience

Amazon, Inc. *Seattle, WA* *Jun. 2025 – Sep. 2025*

Applied Scientist Intern, Mentor: Kyle Willett

- Developed a multi-agent system to orchestrate and automate a causal estimation analysis pipeline with MCP, improving scalability and reducing manual intervention.
- Built a text to SQL layer in the agentic workflow that transforms unstructured information into query language to streamline the ETL pipeline.
- Post trained open-weight LLMs with LoRA, improving alignment between model outputs and domain-specific reasoning processes.
- Reduced average business deliverable turnaround time by over 50%, from two weeks to under one week, accelerating decision-making and project execution.

Amazon, Inc. *Seattle, WA* *Jun. 2024 – Dec. 2024*

Applied Scientist Intern, Mentor: Kyle Willett & Hengrui Cai

- Developed a Generative AI application using LLM agent framework through AWS Bedrock to orchestrate causal effect estimation workstream for third party seller FBA fee changes.
- Leveraged tool calling ability of LLM to implement chained and recursive tool invocation actions.
- Proposed evaluation frameworks for assessing causal reasoning ability and factual consistency of LLM outputs.
- Helped built an multi-agent framework to recommend, execute and interpret various causal models used in Amazon.

Duke Cancer Institute *Durham, NC* *Aug. 2019 – Jul. 2021*

Bioinformatician I, Mentor: Kouros Owzar

- Designed and implemented an S4 R package for assessing sequencing and mapping quality of RNA-Seq data.
- Operated different bioinformatics pipelines through dockerized containers in remote HPC clusters.
- Delivered reproducible and well-documented reports and code.

Skills & Coursework

Programming Languages: Python, R, SAS, Shell, SQL, Java
Software: Pytorch, Huggingface, AWS, Scikit-learn, Pandas, Tidyverse
Relevant Coursework: Data Structure and Algorithms, Database Management System, Advanced Machine Learning, Artificial Intelligence, Statistical Programming with Big Data, Generalized Linear Model, Causal Inference, Computational Causal Analytics, Natural Language Processing with Deep Learning