Daking Rai

Fairfax, Virginia, USA

drai2@gmu.edu | https://dakingrai.github.io/ | https://github.com/Dakingrai

Education

George Mason University — *Fairfax, VA*

Ph.D. student in Computer Science — *Aug 2021 – Aug 2026 (Expected)*

Research Advisor: Dr. Ziyu Yao

Tribhuvan University — *Kathmandu, Nepal*

Bachelors in Computer Science and Information Techology — *Aug. 2012 – Aug. 2016*

Awarded **"Outstanding Student of the Batch 2012"** Award

Publications

Daking Rai, and Ziyu Yao. **A Practical Review of Mechanistic Interpretability for Transformer-Based Language Models**. *(Arxiv pre-print)*, 2024.

Daking Rai, and Ziyu Yao. **An Investigation of Neuron Activation as a Unified Lens to Explain Chain-of-Thought Eliciting Arithmetic Reasoning of LLMs**. *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2024.

Daking Rai, Bailin Wang, Yilun Zhou and Ziyu Yao. **Improving Generalization in Language Model-based Text-to-SQL Semantic Parsing: Two Simple Semantic Boundary-Based Techniques**. *Annual Meeting of the Association for Computational Linguistics (ACL)*, July 2023.

Daking Rai, Yilun Zhou, Bailin Wang and Ziyu Yao. **Explaining Large Language Model-Based Neural Semantic Parsers** . *AAAI Student Abstract and Poster Program*, February 2023.

Research & Work Experience

Graduate Research Assistant — *Fairfax, VA, USA*

George Mason University NLP Lab — *May 2022 - Present*

- Advisor: Dr. Ziyu Yao
- **Ongoing Projects:**
  - **Mechanistic Interpretability of Code LLMs:**
    This research project aims to form an understanding of how a Code LLM (e.g., CodeLLaMA, DeepSeek Coder, etc.) stores and recalls programming-related facts (e.g., syntactic vs semantic facts).

  - **Calibrating Trust in Human-Machine Interactions with Algorithm Transparency:**
    We're conducting a user study to examine how algorithm explanations at different transparency levels influence human trust and task performance. Notably, we study this research problem under the intriguing task setting of text-to-code generation, particularly focusing on users who lack coding expertise. This study aims to gain a deeper understanding of "trust" in secure human-machine interaction settings. Additionally, we also want to showcase how calibrating human trust can make a state-of-the-art semantic parser more effective and secure in practice.

Graduate Teaching Assistant — *Fairfax, VA, USA*

George Mason University — *Aug 2021 - April 2022*

- Teaching Assistant for Computer Systems and System Programming (CS531) & Essentials of Computer Science (CS110).

Machine Learning Engineer — *Kathmandu, Nepal*

Infodev Pvt Ltd — *April 2019 - June 2020*

- Led a research and development team for the integration of semantic search enhanced by named entity recognition (NER) within an e-commerce platform. Additionally, worked on projects involving face recognition, facial aliveness detection, object detection, and localization for various prototypes.

Lecturer/Instructor — *Kathmandu, Nepal*

Islington College — *Oct 2018 – Feb 2019*

- Taught "Introduction to Artificial Intelligence" course with over 70 undergraduate students.

Services

- Served as a reviewer for **AAAI'25, ARR June'24, AAAI'24**.
- Served as a secondary reviewer for **NeurIPS'23**.
- Volunteer for "The 10th annual Mid-Atlantic Student Colloquium on Speech, Language and Learning".
- Volunteer for "The third ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '23)".
- Executive committee member (volunteer) for Rose Foundation Nepal, an NGO working to raise awareness about different types of cancers, especially breast cancer in Nepal.
- Organizer and mentor on 10+ machine learning and deep learning workshops conducted by AI Developer Nepal, an AI community based in Kathmandu.

Awards

- Graduate Student Travel Fund (GSTF) for ACL'24 and ACL'23.
- Honorary mention, GMU CS Research Symposium (Poster Presentation).
- "Outstanding Student of the Batch 2012" Award. Bachelors in Computer Science and Information Technology, Tribhuvan University (TU), Kathmandu.

References

Ziyu Yao (ziyuyao@gmu.edu)

Assistant Professor
Dept. of Computer Science (CS)
George Mason University
4400 University Dr, Fairfax, VA 22030