HAIKANG DENG

+1 919-260-9698
frankdenghaikang@gmail.com

EDUCATION

University of California, Los Angeles Sep 2024 - Present
Ph.D. in Computer Science
Advisor: Nanyun Peng

University of North Carolina, Chapel Hill Aug 2019 - May 2023
B.S. in Computer Science & Statistics GPA: 3.96/4.0
Research Advisor: Colin Raffel, Guanting Chen

PUBLICATIONS

[1] Reward-Augmented Decoding: Efficient Controlled Text Generation With a Unidirectional Reward Model
Haikang Deng and Colin Raffel
*Empirical Methods in Natural Language Processing (**EMNLP**)*, 2023

[2] Large Language Models Struggle to Learn Long-Tail Knowledge
Nikhil Kandpal, **Haikang Deng**, Adam Roberts, Eric Wallace, and Colin Raffel
*International Conference on Machine Learning (**ICML**)*, 2023

RESEARCH & PROFESSIONAL EXPERIENCE

Amazon Dec 2023 - Jul 2024
Software Dev Engineer Bellevue, WA
· Composed strategy to trigger pre-compute of products based on given marketplace, merchant, and store utilization score
· Wrote a tool page for PMs to update up to 1M products' lead time in one go
· Built the Horizonte service for Local Landing Page which displays local products available for pick up

University of North Carolina, Chapel Hill Aug 2022 - Mar 2024
Research Assistant Advised by Prof. Colin Raffel
· Conducted research on human feedback alignment, controlled text generation, and retrieval augmentation
· Ongoing research focuses on expert model routing and tagging with signals from a reward model

Lenovo May 2021 - Aug 2021
Software Engineer Intern Beijing, China
· Trained Encoder-Decoder LSTM for anomaly detection on time series data
· Participated in the design of Control Chart and Anomaly Detection Module
· Performed model tuning and data grouping which improved f1 score from 0.41 to 0.48

PROJECTS

Neurally Decomposed Oracle for Preference Optimization Jul 2024 - Present
· Adapted NADO's training objective to a contrastive one to comply with preference pairs
· Analyzed speed of convergence and model properties as compared to SOTA methods such as DPO
· Interpolate and Extrapolate NADO to explore its potential (to be done)

Routing and Tagging Expert Models Oct 2023 - Mar 2024
· Trained an embedding-based router with tag enhancement to reduce RM noise
· Adapted query tagging to model tagging through router training

Benchmarking Methods for Learning from Human Feedback (LHF) July 2023 - Present
· Incorporated Llama2 reward model into OpenAssistant's code base and trained a Gold RM on RLHF datasets
· Used the Gold RM to generate synthetic data for later experiments on LHF overoptimization

Tracing Model-Generated Wrong Answers Nov 2022 - Feb 2023
· Explored correlation between QA accuracy and the number of relevant documents that contain negative co-occurrences
· Discovered that incorrect generation is not a result of more association of wrong answer and query entities

ACADEMIC ACHIEVEMENTS

Successful Participant, MCM 2021 Feb 2021
UNC Dean's List 19FA, 21FA, 22SP, 22FA
9th Place in China/27th Place Worldwide in AAPT Physics Bowl Apr 2018