Haoyan Yang

haoyan-yang | homepage | hy2847@nyu.edu | 201-744-0731

Research Interest

My research focuses on enhancing the reliability of large language models (LLMs), addressing issues such as hallucinations, privacy leakage, and overconfidence. Additionally, I am interested in better fine-tuning and alignment methods to improve LLM performance and reliability. Speeding up inference to enhance LLM efficiency is also one of my research interests.

Education

New York University

Master of Science in Data Science, GPA: 3.86/4.0 — *Sep. 2023 – May 2025*

Beijing Normal University-Hong Kong Baptist University United International College

Bachelor of Science in Data Science, GPA: 3.84/4.0 (1/94) — *Sep. 2019 – Jun. 2023*

Publications

Haoyan Yang, Zhitao Li, et al. (2023). "PRCA: Fitting Black-Box Large Language Models for Retrieval Question Answering via Pluggable Reward-Driven Contextual Adapter". In: *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pp. 5364–5375.

Yuxuan Chen*, **Haoyan Yang*** et al. (2024). *BURExtract-Llama: An LLM for Clinical Concept Extraction in Breast Ultrasound Reports*. Accepted as an oral paper at ACM MM 2024 Workshop.

Haoyan Yang (2023). "Multimodal Stock Price Forecasting Using Attention Mechanism Based on Multi-Task Learning". In: *Asia-Pacific Web (APWeb) and Web-Age Information Management (WAIM) Joint International Conference on Web and Big Data*. Springer, pp. 454–468.

Haoyan Yang, Hongjiu Zhang, et al. (2023). "Swarm Intelligence Optimization of UAV Routing with Simultaneously Stochastic Pick-up and Delivery during COVID-19". In: *2023 8th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA)*. IEEE, pp. 579–587.

Haoyan Yang, Yixuan Wang, et al. (2024). "Can We Trust LLMs? Mitigate Overconfidence Bias in LLMs through Knowledge Transfer". In: *arXiv preprint arXiv:2405.16856*.

Haoyan Yang*, Zhitao Li* et al. (2024). "PFID: Privacy First Inference Delegation Framework for LLMs". In: *arXiv preprint arXiv:2406.12238*.

Haoyan Yang, Ting Hua, et al. (2024). *Dynamic Noise Preference Optimization for LLM Self-Improvement via Synthetic Data*. Under Review by ICLR 2025.

Research Experience

Reward-driven Adapter for Enhanced Retrieval Question Answering (ReQA)

- Proposed the Pluggable Reward-driven Contextual Adapter (PRCA) method to treat LLMs as black boxes in retrieval-augmented modes, addressing fine-tuning constraints of limited local computational resources.
- Approved that PRCA enhanced average ReQA performance by 3%, 6%, and 9%, and up to 20% on three QA datasets, and robustly adapted to various configurations of retrievers and generators.
- Authored and published a paper on this work at the EMNLP 2023 conference.

Mitigating LLM Overconfidence Bias with Knowledge Transfer (KT)

- Introduced a KT method that leverages larger models with advanced reasoning capability to transfer knowledge through chain-of-thoughts (CoT), fine-tuning smaller ones to correct their overconfidence bias.
- Demonstrated that KT achieved average improvements of 55.3% and 43.1%, respectively, over vanilla and QA models with serious overconfident bias, in metrics of accuracy, ROB, and ECE.

Privacy-Preserving in LLMs Using Model Sharding

– Proposed a Privacy First Inference Delegation (PFID) framework for preserving privacy within LLMs by localizing user data via model sharding, reducing the need to share data with central servers.

– Showed that PFID achieved average drops of 1.14% and 8.07% in COMET and BLEU scores, respectively, compared to the original pipeline, significantly outperforming the drops in the scenario of data interception (28.16% and 51.58%), highlighting the model's ability to maintain performance while enhancing data privacy in machine translation tasks.

Table-Logic Sequential Prompting for TableQA Performance Analysis

– Introduced a prompting flow, Table-Logic Sequential Prompting, to analyze performance disparities between big and small LLMs in TableQA tasks.

– Revealed that the proposed method improved large models by 7.8% in accuracy for HybridQA tasks, but declined by 11% in smaller models, illustrating performance discrepancies and guiding model refinement.

Professional Experience

Samsung Research America *May 2024 - Aug. 2024*

Researcher Intern (NLP Focused); Supervisor: Ting Hua; Mountain View, US

– Researched methods for utilizing synthetic data generated by LLMs to facilitate model self-improvement.

– Proposed Dynamic Noise Preference Optimization to achieve LLM self-improvement via self-generated data.

NYU Langone Health *Mar. 2024 - Now*

Research Assistant (NLP for Healthcare Focused); Supervisor: Artie Shen; Remote, US

– Built a pipeline that integrates LLMs and vision models for breast ultrasound report generation.

– Fine-tuned LLaMA-3-8B using 4k breast report data, achieving a 10% improvement compared to few-shot learning and performance comparable to GPT-4 under human-annotated labels.

Ping An Technology | *Subsidiary of the leading insurance company in China* *Mar. 2023 - Jun. 2023*

Algorithm Engineer Intern (NLP Focused); Supervisor: Zhitao Li; Shenzhen, China

– Conducted self-motivated research on LLMs and published one paper as the first author under the supervision.

– Developed an intelligent QA system based on the LLM "Phoenix" tailored for the insurance and finance sector.

– Reduced manual customer service workload by 20% with the new intelligent QA system, achieving cost savings.

SF Express | *Leading logistics company in China* *Aug. 2021 - Sep. 2021*

Data Analyst Intern; Shanghai, China

– Managed logistics data and analyzed operations for 600+ SF Express sites, boosting efficiency by 10%.

Tongcheng Travel | *Leading online travel service company in China* *Jun. 2021 - Jul. 2021*

Data Analyst Intern; Suzhou, China

– Analyzed app data and built dashboards, aiding in 6 tourism marketing strategies.

Awards

2022 - 2023 AY	Scholastic Award (Top 2%), Outstanding Academic Poster Award
2021 - 2023 AY	Three Consecutive First Class Scholarships (Top 4%)
2020 - 2021 AY	Student Internship Scholarship Award
2019 - 2020 AY	Second Class Scholarship (Top 12%)

Skills

Programming Languages	Python, R, Java, C, SQL, MATLAB, HTML/CSS, JavaScript
Technologies/Frameworks	NLP, Machine Learning, Deep Learning, Time Series Analysis, Data Analysis, Regression, Big Data, Tableau, Pytorch, Tensorflow