

HETONG WANG

hetong.wang809@gmail.com · <https://erikaawang.github.io/>

EDUCATION

University of Edinburgh, Edinburgh, UK

MSc in Artificial Intelligence, Nov 2023

Thesis: Probing the dynamics of implicit domain alignment in LLMs
advised by Prof. [Pasquale Minervini](#) and Prof. [Edoardo M. Ponti](#)

University of Liverpool, Liverpool, UK

BSc in Computer Science, Jul 2022

First Class with Honours, Highest Distinction (GPA 4.0/4.0, top 1%)

PUBLICATION

Hetong Wang, Pasquale Minervini, Edoardo Ponti. “Probing the Emergence of Cross-lingual Alignment during LLM Training”, In *Findings of the Association for Computational Linguistics*, pp. 12159–12173, **ACL 2024**.

Gabriella Pizzuto, **Hetong Wang**, Hatem Fakhruddin, Bei Peng, Kevin S. Luck, Andrew I. Cooper. “Accelerating Laboratory Automation Through Robot Skill Learning For Sample Scraping”, In *IEEE International Conference on Automation Science and Engineering*, pp. 2103-2110, IEEE CASE 2024.

🏆 **Best Healthcare Automation Paper Award.**

PROJECTS

Boosting LLM Rewarding through Peer Model Ensembles

Work in progress, advised by Prof. [Yulia Tsvetkov](#)

- This project aims to augment feedback signals in both self-training and LLM-as-a-judge settings by leveraging synthetic data generated across an ensemble of LLMs at the 7B parameter scale

Sparse Fine-tuning Towards Multi-task Instruction Following

Group Project at University of Edinburgh, [Paper](#)

- Proposed an approach to avoid catastrophic forgetting by sparsely tuning the parameters with the largest gradient update. Achieved 20% generalization gain over unseen data compared with full-parameter or random-masked tuning

RESEARCH EXPERIENCE

Research Intern, TsvetShop, University of Washington

Nov. 2024 – Present, Supervisor: Prof. [Yulia Tsvetkov](#)

Research Intern, THUNLP Lab, Tsinghua University

Feb 2024 - Oct 2024, Supervisor: [Zhiyuan Liu](#)

Research Assistant, Leverhulme Research Centre, University of Liverpool

Jun 2022 - Sep 2022, Supervisor: [Gabriella Pizzuto](#) and [Andy Cooper](#)

HONORS & AWARDS

Best Healthcare Automation Paper Award, *IEEE CASE*, 2024

British Computer Society Prize, *best academic performance in 3rd year, top 1%*, 2022

Academic Excellence Award, University of Liverpool £5,400/yr, 2020-2022

SKILLS

Research Skills

- Toolkits: Huggingface, Pytorch, Numpy, Keras, Scikit-learn, OpenAI gym
- HPC Cluster: Slurm, SGE(Sun Grid Engine), Shell, Linux

Programming Python, Java, L^AT_EX, R, C/C++, SQL, HTML