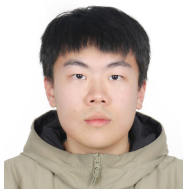

Haitian Zhong

New Laboratory of Pattern Recognition
Institute of Automation
Chinese Academy of Sciences
95 Zhongguancun East Road, Beijing 100190, CHINA

+86-15996295520
haitian.zhong@cripac.ia.ac.cn
jzsawyer0322@gmail.com
GitHub
Homepage

Academic Background

- **Beijing Zhongguancun Academy** *Sept. 2024 - Now*
  Long CoT LLM Reasoning: data, algorithm and system Beijing, China
  – Research Topics: (Multimodal) LLM Reasoning
  – Supervisor: Prof. Bin Dong
  – Co-supervisor: Prof. Wentao Zhang

- **Institute of Automation, Chinese Academy of Sciences (CASIA)** *Sept. 2024 - Now*
  New Laboratory of Pattern Recognition (NLPR) Beijing, China
  – Research interests: (Multimodal) Large Language Models Safety
  – Supervisor: Prof. Tieniu Tan
  – Co-supervisor: Prof. Qiang Liu and Prof. Shu Wu

- **Lanzhou University** *Sept. 2020 - Jun. 2024*
  Mathematics and applied mathematics (China's Top-notch Undergraduate Training Program 2.0) Lanzhou, China
  – GPA: 91.75/100 (See Transcript at Chinese Version OR English Version)
  – Ranking: 2/152
  – CET-4: 656; CET-6: 601 (Oral grading: B+)

- **Nanjing Foreign Languages School** *Sept. 2014 - Jun. 2020*
  Top Class of Science Nanjing, China
  – Selected Awards: (Senior) High School Mathematics Competition, Provincial First Prize; Zhou Enlai's Merit Scholarship; Nanjing Merit Student; Merit Student of NFLS (6-star).
  – TOEFL: 103

Selected Awards

- **Undergraduate Graduate Representative of Lanzhou University** *Jun. 2024*
- **Stars of Tomorrow at Microsoft Research Asia** *Jun. 2024*
- **Provincial Outstanding Graduates** *Mar. 2024*
- **The 14^{th} Chinese Mathematics Competition**,Provincial First Prize(Top 1%, Promoted to National Final) *Apr. 2023*
- **The 13^{th} Chinese Mathematics Competition**,National Second Prize(Top 1%) *Mar. 2023*
- **S.-T. Yau College Student Mathematics Contest(Applied and Computational Maths)**,Excellence award *Jul. 2022*
- **S.-T. Yau College Student Mathematics Contest(Analysis and Differential Equations)**,Excellence award *Jul. 2022*
- **University Merit Scholarship**,First Class(Top 1%) *Nov. 2022*
- **"FLTRP-ETIC Cup" English Public Speaking Contest**,First Prize *Sept. 2022*
- **National English Competition for College Students**,National First Prize(Top 1%) *May 2022*

Publications

- **Representation Extraction And Controllable Tuning to Overcome Overfitting in LLM Knowledge Editing** *Aug. 2025*
  ***Haitian Zhong**, Yuhuan Liu, Ziyang Xu, Guofan Liu, Qiang Liu*, Shu Wu, Zhe Zhao, Liang Wang, Tieniu Tan* **EMNLP 2025**
  – Abstract: Large language model editing methods frequently suffer from overfitting, wherein factual updates can propagate beyond their intended scope, overemphasizing the edited target even when it's contextually inappropriate. To address this challenge, we introduce **REACT** (Representation Extraction And Controllable Tuning), a unified two-phase framework designed for precise and controllable knowledge editing. In the initial phase, we utilize tailored stimuli to extract latent factual representations and apply Principal Component Analysis with a simple learnbale linear transformation to compute a directional "belief shift" vector for each instance. In the second phase, we apply controllable perturbations to hidden states using the obtained vector with a magnitude scalar, gated by a pre-trained classifier that permits edits only when contextually necessary. Relevant experiments on EVOKE benchmarks demonstrate that **REACT** significantly reduces overfitting across nearly all evaluation metrics, and experiments on COUNTERFACT and MQuAKE shows that our method preserves balanced basic editing performance (reliability, locality, and generality) under diverse editing scenarios.

– Published in EMNLP 2025, see the paper at arXiv:2505.18933

- **VLKEB: A Large Vision-Language Model Knowledge Editing Benchmark** *Sep. 2024*
  *Han Huang†, **Haitian Zhong†**, Qiang Liu*, Shu Wu, Liang Wang, Tieniu Tan* **NeurIPS 2024**
  – Abstract: Recently, knowledge editing on large language models (LLMs) has received considerable attention. Compared to this, editing Large Vision-Language Models (LVLMs) faces extra challenges from diverse data modalities and complicated model components, and data for LVLMs editing are limited. The existing LVLM editing benchmark, which comprises three metrics (Reliability, Locality, and Generality), falls short in the quality of synthesized evaluation images and cannot assess whether models apply edited knowledge in relevant content. Therefore, we employ more reliable data collection methods to construct a new Large **V**ision-**L**anguage Model **K**nowledge **E**diting **B**enchmark, **VLKEB**, and extend the Portability metric for more comprehensive evaluation. Leveraging a multi-modal knowledge graph, our image data are bound with knowledge entities. This can be further used to extract entity-related knowledge, which constitutes the base of editing data. We conduct experiments of different editing methods on five LVLMs, and thoroughly analyze how do they impact the models. The results reveal strengths and deficiencies of these methods and hopefully provide insights for future research.
  – Published in **NeurIPS 2024**, see the paper at arXiv:2403.07350, codes at Github:VLKEB, dataset at Huggingface:VLKEB

- **PTransIPs: Identification of phosphorylation sites enhanced by protein PLM embeddings** *Mar. 2024*
  *Ziyang Xu†, **Haitian Zhong†**, Bingrui He, Xueying Wang, Tianchi Lu** **IEEE J-BHI**
  – Abstract: Identification of phosphorylation sites is an important step for understanding the molecular mechanisms of SARS-CoV-2 infection and the changes within the host cells pathways. In this study, we present PTransIPs, a new deep learning framework for the identification of phosphorylation sites. PTransIPs utilizes protein pre-trained language model (PLM) embeddings to achieve SOTA performance, with AUCs of 0.9232 and 0.9660 for S/T and Y sites, respectively. PTransIPs is also a universal framework for all peptide bioactivity tasks.
  – Published on **IEEE Journal of Biomedical and Health Informatics**. Codes at Github:PTranIPs

Experiences

- **Microsoft Research Asia** *Mar. 2024 - Jun. 2024*
  Research Assistant at Social Computing Group, supervised by Prof. Xing Xie and Fangzhao Wu Beijing, China
  – Project: In the rapidly advancing field of artificial intelligence, ensuring the safety and ethical integrity of Large Language Models (LLMs) is paramount. These models, while powerful and versatile, have the potential to generate harmful outputs, including biased, misleading, or offensive content. We focus on developing methods to identify and mitigate such harmful outputs, aiming to make LLMs more society-friendly and aligned with ethical standards. By implementing robust post-processing techniques and incorporating comprehensive interpretable frameworks, we strive to enhance the reliability and trustworthiness of LLMs, ensuring their positive and responsible integration into various applications.

- **CRIPAC, CASIA** *Oct. 2023 - Feb. 2024*
  Undergraduate Researcher, supervised by Prof. Qiang Liu and Prof. Shu Wu Beijing, China
  – Project: VLKEB: A Large Vision-Language Model Knowledge Editing Benchmark

Integrated Skills

My skills consist of Mathematics knowlegde, Computer Programming and excellent language ability. Nevertheless, I am a fast learner of new tools and a fanatic lover of self-learning.

- **Mathematics**: Optimzation, Analysis, PDE, Numerical Analysis, Statistics
- **Programming Languages**: Python, R, C/C++, Mathematica, MATLAB, LaTeX
- **Artificial Intelligence**: PyTorch
- **English**: Very fluent in oral English; Proficient in English writing and reading English papers