YIMING TANG

Personal Website: realyimingtangible.github.io/
Email: yiming [at] nus [dot] edu [dot] sg ⋄ Phone: +65 89546908 ⋄ WeChat: YimingTangible
Google Scholar: scholar.google.co.uk/citations?user=0WYAYQ8AAAAJ

EDUCATION

Doctor of Philosophy August 2024 - Present
College of Design and Engineering
National University of Singapore
Supervisor: Prof. Dianbo Liu

Bachelor of Science September 2019 - June 2024
School of Mathematical Sciences
Peking University
GPA: Ranked 3rd in Major

High School Diploma September 2016 - June 2019
Chengdu Jinjiang Jiaxiang Foreign Languages High School
Gaokao Score: 700/750, Provincial Rank: Top 200
Chinese Mathematical Olympiad Gold Medal, National Rank: 74

RESEARCH INTERESTS

Primary Focus	Mechanistic Interpretability, Vision-Language Models
Interested Topics	AI Interpretability, Computer Vision, Machine Learning Theory, AI Safety, Biomedical Engineering, Medical Imaging

RESEARCH PHILOSOPHY

My primary aspiration is to conduct solid research that advances the intuitive understanding of intelligence as a scientific problem, encompassing both theoretical foundations and algorithmic innovations. I consider mechanistic interpretability as one promising approach and focus on concept-based interpretability. I also develop practical applications that leverage this understanding to solve various downstream tasks, with medical diagnostic models standing as a typical example where improved model interpretability directly translates to clinical trust.

SKILLS AND LANGUAGES

Programming	Python, C, C++, MATLAB, LaTeX, Linux, HTML, Github, PyTorch, TensorFlow, Hugging Face, Transformers, and etc.
Languages	English, Chinese

AWARDS AND HONORS

- **Chinese Mathematical Olympiad (CMO) Gold Medal**, National Rank 74 2018
- **Outstanding Student Award**, Chengdu Jiaxiang Foreign Languages High School 2019
- **The Chinese Mathematics Competitions First Prize** 2020
- **Mathematical Contest in Modeling (MCM) Honorable Mention** 2020

PROFESSIONAL DEVELOPMENT

- **Peer Review**, Active reviewer for NeurIPS, CVPR, ECCV, and other ML conferences.
- **Mentorship**, Mentored five interns in Cognitive AI for Science lab, with developed mentoring skills.
- **Community Contributions**, Released LanSE models, datasets, and evaluation suite for community use.

PROFESSIONAL EXPERIENCES

Research Associate 2024 - Now
Cognitive AI for Science Lab, National University of Singapore

- Worked with Prof. Dianbo Liu's group on mechanistic interpretability and AI for medicine.
- One paper under review at Science Advances, proposing Language-Grounded Sparse Encoders.
- One paper under review at CVPR, proposing Matryoshka Transcoders as one SOTA-level interpretability tool.
- One paper under review at ICLR, benchmarking LLM's repository refining skills.
- Leading six on-going projects about mechanistic interpretability encompassing diverse aspects of MI.

Research Intern 2023 - 2024
Beijing International Center for Mathematical Research , Peking University

- Collaborated with Prof. Bin Dong's group on prompt engineering and AI for education.
- One paper published in JML, developing the first theoretical framework for prompt engineering.
- One AI tutor application adopted officially by Peking University, as one of the first AI teaching facilities in China.
- Two first-author papers preprinted, proposing IntFF and Demonstration Notebook.

Research Intern 2023 - 2024
Human Machine Intelligence Lab, Peking University

- Worked with Prof. Shanghang Zhang's group on computer vision, dataset analysis, and PEFT.
- One paper published in ECCV, first to propose to utilize LLMs to analyze datasets.
- One paper published in ICML, improving PEFT with neuroscience inspiration.

Research Intern 2022 - 2023
Hyperplane Lab, Peking University

- Collaborated with Prof. Hao Dong and Prof. Wei Pan (from TU Delft) on physics-informed neural networks.
- Learned about some fundamental techniques and applications for robotics and PINN.

PUBLICATIONS AND PREPRINTS

Y. Tang, A. Lagzian, S. Anumasa, Q. Zou, Y. Zhu, Y. Zhang, T. Nguyen, et al., "*Human-like Content Analysis for Generative AI with Language-Grounded Sparse Encoders*", arXiv:2508.18236, 2025

Y. Tang, A. Sinha, D. Liu, "*How does My Model Fail? Automatic Identification and Interpretation of Physical Plausibility Failure Modes with Matryoshka Transcoders*", arXiv:2511.10094, 2025

Y. Tang, W. Zhong, R. Shah, D. Liu, "*CXR-LanIC: Language-Grounded Interpretable Classifier for Chest X-Ray Diagnosis*", arXiv:2510.21464, 2025

Q. Zou, H.H. Lam, W. Zhao, **Y. Tang**, T. Chen, S. Yu, et al., "*FML-bench: A Benchmark for Automatic ML Research Agents Highlighting the Importance of Exploration Breadth*", 2025

Y. Tang, B. Dong, "*Demonstration Notebook: Finding the Most Suited In-Context Learning Example from Interactions*", arXiv:2406.10878, 2024

Y. Luo, R. An, B. Zou, **Y. Tang**, J. Liu, S. Zhang, "*LLM as Dataset Analyst: Subpopulation Structure Discovery with Large Language Model*", **European Conference on Computer Vision (ECCV)**, 235-252, 2024

G. Dai, C.K. Fan, **Y. Tang**, Q. Zhang, Y. Gan, Z. Zhang, C.C. Tseng, S. Zhang, T. Huang, "*Discovering Long-Term Effects on Parameter Efficient Fine-tuning*", arXiv:2409.06706, 2024

G. Dai, C.K. Fan, **Y. Tang**, Q. Zhang, Y. Gan, Z. Zhang, C.C. Tseng, S. Zhang, T. Huang, "*SAN: Hypothesizing Long-Term Synaptic Development and Neural Engram Mechanism in Scalable Model's Parameter-Efficient Fine-Tuning*", **International Conference on Machine Learning (ICML)**, 2024

Y. Luo, **Y. Tang**, C. Shen, Z. Zhou, B. Dong, "*Prompt Engineering Through the Lens of Optimal Control*", **Journal of Machine Learning** 2(4), 241-258, 2023

Y. Tang, "*The Integrated Forward-Forward Algorithm: Integrating Forward-Forward and Shallow Backpropagation with Local Losses*", arXiv:2305.12960, 2023