Bingda Tang

tangbd2003@gmail.com

Education

2021.09 - 2026.06 **Bachelor of Engineering in Computer Science and Technology**, *Department of Computer Science and Technology, Tsinghua University*

Academic Performance: **3.71 / 4.00**

Research Interests: **Multimodal Learning and Generative Modeling**

Experience

2025.06 - Present **Research Intern**, *Stanford University*, Advised by **Serena Yeung-Levy**

1. **Online RL of Denoising Generative Models for Text-to-Image Synthesis | Lead**
   - Designed variance reduction techniques to unlock the potential of a theoretically sound but highly unstable algorithm that uses ELBO-based objectives as surrogates for model log-likelihoods in GRPO.
   - Achieved state-of-the-art performance and efficiency in post-training text-to-image models while addressing key limitations of prior approaches.
2. **Reasoning for Unified Multimodal Understanding and Generation | Lead**
   - Explored incentivizing unified multimodal models to perform interleaved image-text reasoning for improved text-to-image synthesis.
   - Finetuned BLIP3-o 4B on text-to-image data with long CoT curated from Qwen3 32B.
   - Applied GRPO jointly to both autoregressive and flow matching objectives.

2024.06 - Present **Research Intern**, *New York University*, Advised by **Saining Xie**

1. **Deep Fusion of LLMs and DiTs for Text-to-Image Synthesis | Lead**
   - Proposed a deep fusion architecture that integrates LLMs and DiTs via shared self-attention to leverage the internal dynamics of LLMs for text-to-image synthesis.
   - Developed a scalable SPMD training codebase in PyTorch/XLA for large-scale pretraining on Google TPUv4 infrastructure (code earned 100+ stars on Github).
   - Conducted extensive pretraining experiments to systematically compare architectural variants and evaluate key design choices.
2. **Pretraining for Unified Multimodal Understanding and Generation | Contributor**

2023.07 - 2024.08 **Research Intern**, *Tsinghua University*, Advised by **Ting Chen**

1. **Visual Grounding Multimodal LLMs for Medical Image Analysis | Co-Lead**
   - Co-initiated the research, led the development of the data annotation pipeline, co-designed the model architecture and training procedure, evaluated performance on VQA and report generation, and co-authored the manuscript.
2. **Vision Foundation Models for Medical Image Analysis | Contributor**

Publications

* indicates equal contribution. † indicates corresponding author.

In Submission **V-GRPO: Online Reinforcement Learning for Denoising Generative Models Is Easier than You Think**, ***Bingda Tang***, *Yuhui Zhang, Xiaohan Wang, Jiayuan Mao, Ludwig Schmidt, Serena Yeung-Levy*†

In Submission **Scaling Text-to-Image Diffusion Transformers with Representation Autoencoder**, *Shengbang Tong*, Boyang Zheng*, Ziteng Wang**, ***Bingda Tang***, *Nanye Ma, Ellis Brown, Jihan Yang, Rob Fergus, Yann LeCun, Saining Xie*†

CVPR 2025 **Exploring the Deep Fusion of Large Language Models and Diffusion Transformers for Text-to-Image Synthesis**, ***Bingda Tang***, *Boyang Zheng, Xichen Pan, Sayak Paul, Saining Xie†*

NAACL 2025 **VividMed: Vision Language Model with Versatile Visual Grounding for Medicine**, *Lingxiao Luo*,* ***Bingda Tang****, *Xuanzhong Chen, Rong Han, Ting Chen†*

CVPR 2025 **CORE4D: A 4D Human-Object-Human Interaction Dataset for Collaborative Object REarrangement**, *Yun Liu*, Chengwen Zhang*, Ruofan Xing,* ***Bingda Tang***, *Bowen Yang, Li Yi†*

CIKM 2025 Short Paper **Jailbreaking LLMs Through Alignment Vulnerabilities in Out-of-Distribution Settings**, *Yue Huang, Jingyu Tang, Dongping Chen,* ***Bingda Tang***, *Yao Wan, Lichao Sun, Philip S. Yu, Xiangliang Zhang†*

IMWUT 2024 **CalibRead: Unobtrusive Eye Tracking Calibration from Natural Reading Behavior**, *Chang Liu, Chun Yu†, Xiangyang Wang, Jianxiao Jiang, Tianao Yang,* ***Bingda Tang***, *Yingtian Shi, Chen Liang, Yuanchun Shi*

Preprint **Building Universal Foundation Models for Medical Image Analysis with Spatially Adaptive Networks**, *Lingxiao Luo, Xuanzhong Chen,* ***Bingda Tang***, *Xinsheng Chen, Chengpeng Hu, Yujiang Li, Rong Han, Ting Chen†*

Projects

2024.11 - 2024.12 RISC-V 5-Stage Pipeline CPU on FPGA, with Exception Handling, Clock Interrupt, Virtual Memory, and Cache (Done in **1 week**)

2024.03 - 2024.05 Instant Messaging Web Application Based on Next.js and Socket.io (Teamwork)

2023.09 - 2023.11 Compiler for MiniDecaf (Subset of C) Based on Python-Lex-Yacc

2022.08 - 2022.08 Backend for Online Judge Based on Actix Web and SQLite

2022.08 - 2022.08 Wordle Game Based on EGUI

Skills

Languages Chinese (Native), English (Fluent, **111 / 120** in TOEFL iBT (R: **30 / 30**, L: **29 / 30**, S: **24 / 30**, W: **28 / 30**), **686 / 710** in CET6), Japanese (Beginner)

Programming C / C++ / CUDA, Python, Rust, LaTeX, R, JavaScript / TypeScript, Verilog / SystemVerilog, MATLAB

Highlight Courses Foundation of Object-Oriented Programming: **A**, Programming and Training: **A**, Introduction to Complex Analysis: **A+**, Multivariate Statistical Analysis: **A**, Advanced Topics in Linear Algebra (English): **A+**, Principles and Practice of Compiler Construction: **A**, Physics for Scientists and Engineers B(2): **A+**, Software Engineering: **A**, Fundamentals of Computer Graphics: **A**

Honors

2025.10 Research Excellence Scholarship, Tsinghua University

2020.10 First Prize, Preliminary (Provincial) Round, 34th Chinese Chemistry Olympiad

2019.10 First Prize, Preliminary (Provincial) Round, 33rd Chinese Chemistry Olympiad