Yiran Xu

Fudan University

✉ yiranxu22@m.fudan.edu.cn github.com/Raizellll 🌐 raizellll.github.io

Research Focus: Emergent Modularity, Representation Geometry, Gradient Dynamics in Transformers

Education

Fudan University **Sep. 2022 – Jun. 2026 (Expected)**
B.S. in Computer Science and Technology *Shanghai, China*

Relevant Coursework: Machine Learning, Natural Language Processing, Algorithms

Selected Research Manuscript

Xu, Y., Dick, R. P. ”Demand-Driven Modularity in Fine-Tuned Transformers: Functional Conflict, Efficiency Bias, and Subspace Collapse.”
Manuscript in preparation (Targeting ICML 2026).

Research Experience

Visiting Scholar **Jun. 2025 – Present**
EECS Department, University of Michigan Supervisor: Prof. Robert P. Dick *Ann Arbor, MI, USA*

Mechanisms of Modularity: Functional Conflict and Efficiency Bias

- Proposed the ”Demand-Driven Modularity” hypothesis:: showed that input-distribution shifts do not induce modularity and that functional conflict is the key driver of physical parameter separation.
- Identified the ”Efficiency Bias” mechanism: Transformers maximize parameter reuse (high neuron overlap) and only separate into distinct manifolds under catastrophic functional interference.
- Reinterpreted the gradient starvation hypothesis by verifying early-layer optimality (L0 Probe Acc = 1.0), indicating that weight stagnation reflects feature sufficiency, not gradient loss.

Undergraduate Researcher **Sept. 2025 – Present**
Alex Reasoning Group, Fudan University Supervisor: Prof. Yixin Cao *Shanghai, China*

Neural Activation Analysis for LLM Evaluation

- Built a full-stack activation-space analysis pipeline to quantify reasoning depth, coherence, and creativity using interpretable low-rank subspaces.
- Identified semantic directions aligned with human rubrics and demonstrated their predictive power for multi-dimensional reasoning quality.
- Connected activation-manifold geometry with model reasoning behaviors across tasks.

Undergraduate Researcher **Feb. 2025 – Jun. 2025**
MEMX Group, Fudan University Supervisor: Prof. Li Shang *Shanghai, China*

Causal RL for Modular Reasoning in LLMs

- Developed and validated a causal-RL framework for compact LLMs using MoE routing to disentangle decomposition, justification, and conclusion roles.
- Discovered and mitigated efficiency-bias collapse in self-training and introduced causal-consistency rewards that restored reasoning depth and stability across math, logic, and commonsense tasks.
- Contributed empirical findings that informed the later NAD interpretability framework.

Industry Experience

Research Intern **Jan. 2025 – Mar. 2025**
Huawei PaaS Lab Mentor: Dr. Yuchi Ma *Shenzhen, China*

LLM Reasoning & Code Generation

- Designed a prompting pipeline for long-horizon code reasoning: decomposition → iterative synthesis → verification.
- Fine-tuned Qwen-2.5-72B on TACO dataset with 20-step reasoning trajectories, boosting symbolic planning accuracy.
- Analyzed reasoning traces to pinpoint bottlenecks and devised process-level correctness metrics.

Honors and Awards

Third Prize in China Mathematical Contest in Modeling (Top 15%, National) **Nov. 2024**
Academic Excellence Scholarship of FDU **Sept. 2024, Sept. 2023**

Technical Skills

Representation Analysis: CKA/CCA, low-rank probing, activation subspaces, clustering
Model Training: PyTorch, HF Transformers, PEFT/LoRA, MoE routing, vLLM
Systems: CUDA, Docker, Linux, JupyterLab
Programming: Python, C++, SQL