

ZHIYU NI

Email: zhiyuni@berkeley.edu Tel: (+1)323-633-1447

EDUCATION

University of California, Berkeley (UC Berkeley) PhD student, <i>Computer Science</i> , Advisor: Pierluigi Nuzzo	<i>2025.1 - Present</i>
University of Southern California (USC) PhD student, <i>Computer Engineering</i>	<i>Los Angeles, CA</i> <i>2022.8 - 2024.12</i>
University of Science and Technology of China (USTC) B.S. <i>Physics</i> , Outstanding Graduates	<i>Hefei, China</i> <i>2018.8 - 2022.6</i>

PUBLICATIONS

Analyzing Adversarial Vulnerabilities of Graph Lottery Tickets (ICASSP 2024 Oral)
Zhiyu Ni*, Subhajit Dutta Chowdhury*, Qingyuan Peng, Souvik Kundu, Pierluigi Nuzzo

Finding Adversarially Robust Graph Lottery Tickets (TMLR)
Zhiyu Ni*, Subhajit Dutta Chowdhury*, Qingyuan Peng, Souvik Kundu, Pierluigi Nuzzo

Accelerating Grokking via Embedding Transfer from a Weaker Model (ICLR 2025)
Zhiwei Xu*, **Zhiyu Ni***, Yixin Wang[◇], Wei Hu[◇]

RESEARCH EXPERIENCE

Adversarially Robust and Light-weighted GNNs ([github](#))

- Analyzed the robustness of pruned graph neural networks (GNNs) against adversarial attacks.
- Developed self-training techniques and a loss function that improved sparse models' robustness, achieving state-of-the-art (SOTA) robust GNNs with a 90% reduction in computational cost.

LLMs for Anomaly Detection ([github](#))

- Investigated capabilities of LLMs (e.g., ChatGPT, Llama) in anomaly detection and designed in-context learning flows.
- Achieved SOTA precision on GPT-3.5-Turbo compared with GNN-based methods.

Accelerating Grokking via Embedding Transfer from a Weaker Model ([github](#))

- Developed a method to accelerate the grokking phenomenon in neural networks by transferring embeddings from smaller, weaker models to larger target models.
- Demonstrated much faster generalization on target models compared with sota methods.

Defending against Prompt Injection on LLMs

- Examined prompt injection attack mechanics which mislead LLMs to generate unwanted answers.
- Generated positive and negative CoT-based prompt pairs to align LLMs to defend against the attack.

WORK EXPERIENCE

NLP Intern (Iflytek *AI Research Institute*)

- Independently carried out machine translation tasks from English to German and Italian.
- Enhanced translation models to achieve a 15% higher BLEU score compared to Google Translate.

SKILLS

Python, C, PyTorch, Linux