

# Yuzhou Nie

Department of Computer Science, University of California Santa Barbara, CA, United States 93106

 Personal Website  yuzhounie@ucsb.edu  (+1)7656945966  rucnyz

## Education

<b>University of California, Santa Barbara, CA, United States</b> <i>PhD, Department of Computer Science, advised by Wenbo Guo</i>	<b>Sept.2024–</b>
<b>Purdue University, West Lafayette, IN, United States</b> <i>PhD, Department of Computer Science, advised by Wenbo Guo</i> – <b>GPA:</b> GPA 4.00/4.00	<b>Sept.2023–June.2024</b>
<b>Renmin University of China(RUC), Beijing, China</b> <i>Bachelor of Science, School of Statistics, Major in Data Science and Big Data Technology</i> – <b>GPA:</b> GPA 3.85/4.00(Ranking: 5/52)	<b>Sept.2019–June.2023</b>

## Awards

FAR AI PhD Fellowship	Supported from 2024 to 2025
UCSB Research Excellent Award	2024
National Scholarship	2020
First Prize Scholarship	2019-2020&2021-2022&2022-2023
National Second Prize in China Undergraduate Mathematical Contest in Modeling (Top 4%)	2020

## Publications

\* indicates equal contribution

- Xuan Chen\*, **Yuzhou Nie\***, Wenbo Guo, Xiangyu Zhang, *When LLM Meets DRL: Advancing Jailbreaking Efficiency via DRL-guided Search*, accepted by NeurIPS 2024.
- **Yuzhou Nie\***, Fengjiao Gong\* and Hongteng Xu, *Gromov-Wasserstein Multi-modal Alignment and Clustering*, accepted by The Conference on Information and Knowledge Management (CIKM) 2022.

## Preprint

\* indicates equal contribution

- **Yuzhou Nie\***, Zhun Wang\*, Ye Yu\*, Xian Wu, Xuandong Zhao, Wenbo Guo, Dawn Song, *PrivAgent: Agentic-based Red-teaming for LLM Privacy Leakage*
- Yu Yang\*, **Yuzhou Nie\***, Zhun Wang\*, Yuheng Tang, Wenbo Guo, Bo Li, Dawn Song, *SecCodePLT: A Unified Platform for Evaluating the Security of Code GenAI*
- Xuan Chen, **Yuzhou Nie**, Lu Yan, Yunshu Mao, Wenbo Guo, Xiangyu Zhang, *RL-JACK: Reinforcement Learning-powered Black-box Jailbreaking Attack against LLMs*.
- **Yuzhou Nie**, Yanting Wang, Jinyuan Jia, Michael J De Lucia, Nathaniel D Bastian, Wenbo Guo, Dawn Song, *TrojFM: Resource-efficient Backdoor Attacks against Very Large Foundation Models*

## Working Experience

<b>Virtue AI, San Francisco, US</b> <i>Internship, advised by Bo Li</i> – <b>Code Guardrail Models</b>	<b>April.2025–June.2025</b>
<b>Amazon Web Services AI Lab, Shanghai, China</b> <i>Internship, advised by senior applied scientist Da Zheng, AWS Deep Learning group, United States</i> – <b>Graphical Structure for Electronic Health Records</b>	<b>Aug.2022–Jan.2023</b>
<b>Microsoft Research Asia, Beijing, China</b> <i>Internship, advised by senior researcher Ziheng Lu and senior researcher Hongxia Hao</i> – <b>Graph neural network for molecular and crystal with long-range force.</b>	<b>Jan.2023–June.2023</b>

## Skills

**Proficient:** Python (PyTorch, Scikit-learn, etc.), C/C++, Java, SQL, Git/Terminal, LaTeX, R  
**Familiar:** MATLAB, PySpark, Linux/Unix