

# JUNJIE XU

+1(814) 996-8596 ◊ University Park, PA

◊ [junjiexu@psu.edu](mailto:junjiexu@psu.edu) ◊ [Homepage](#) ◊ [Google Scholar](#) ◊ [LinkedIn](#)

## SUMMARY

- **Ph.D.** candidate in Machine Learning with **5+** years of R&D experience in the field.
- **Research Interests:** LLM, AI for Science, Graph Learning, Geometric Deep Learning.
- **18 papers** published (NeurIPS, ICLR, ICML, KDD, etc) or under submission, with **900+ citations** as of 11/2025.
- Extensive industry experiences conducting research and developing ML algorithms for real-world applications.

## EDUCATION

<b>Pennsylvania State University</b> <i>Ph.D. candidate</i> in Informatics Advisor: Dr. Suhang Wang & Dr. Xiang Zhang	08/2021 - Present University Park, USA
<b>Huazhong University of Science and Technology</b> <i>B.E.</i> in Software Engineering GPA: 3.91/4.00	09/2017 - 06/2021 Wuhan, China
<b>University of California, Berkeley</b> <i>Exchange student</i> in Computer Science	01/2020 - 06/2020 Berkeley, USA

## SELECTED PAPERS

For the full publication list, please refer to [Google Scholar](#).

- [1] DualEquiNet: A Dual-Space Hierarchical Equivariant Network for Large Biomolecules  
**Junjie Xu**, Jiahao Zhang, Mangal Prakash, Xiang Zhang, Suhang Wang  
**NeurIPS 2025**
- [2] A Comprehensive Survey of Small Language Models in the Era of Large Language Models: Techniques, Enhancements, Applications, Collaboration with LLMs, and Trustworthiness  
Fali Wang, Zhiwei Zhang, Xianren Zhang, Zongyu Wu, Tzuhalo Mo, Qiuhan Lu, Wanjing Wang, Rui Li,  
**Junjie Xu**, Xianfeng Tang, Qi He, Yao Ma, Ming Huang, Suhang Wang  
**TIST, Transactions on Intelligent Systems and Technology**
- [3] Beyond Sequence: Impact of Geometric Context for RNA Property Prediction  
**Junjie Xu**, Artem Moskalev, Tommaso Mansi, Mangal Prakash, Rui Liao  
**ICLR 2025**, also appears in AIDrugX @ NeurIPS 2024, AI for New Drug Modalities @ NeurIPS 2024
- [4] Robustness-Inspired Defense Against Backdoor Attacks on Graph Neural Networks  
Zhiwei Zhang, Minhua Lin, **Junjie Xu**, Zongyu Wu, Enyan Dai, Suhang Wang  
**ICLR 2025 (Oral)**
- [5] LLM and GNN are Complementary: Distilling LLM for Multimodal Graph Learning  
**Junjie Xu**, Zongyu Wu, Minhua Lin, Xiang Zhang, Suhang Wang  
**BigData 2025**
- [6] Geometric Hyena Networks for Large-scale Equivariant Learning  
Artem Moskalev, Mangal Prakash, **Junjie Xu**, Tianyu Cui, Rui Liao, Tommaso Mansi  
**ICML 2025 (Spotlight)**
- [7] LanP: Rethinking the Impact of Language Priors in Large Vision-Language Models  
Zongyu Wu, Yuwei Niu, Hongcheng Gao, Minhua Lin, Zhiwei Zhang, Zhifang Zhang, Qi Shi, Yilong Wang, Sike Fu, **Junjie Xu**, Junjie Ao, Enyan Dai, Lei Feng, Xiang Zhang, Suhang Wang  
ArXiv 2025

- [8] HARMONY: A Multi-Representation Framework for RNA Property Prediction  
**Junjie Xu**, Artem Moskalev, Tommaso Mansi, Mangal Prakash, Rui Liao  
**AI4NA @ ICLR 2025 (Oral)**
- [9] Shape-aware Graph Spectral Learning  
**Junjie Xu**, Enyan Dai, Dongsheng Luo, Xiang Zhang, Suhang Wang  
**CIKM 2024**
- [10] Enhancing GNNs with Limited Labeled Data by Actively Distilling Knowledge from LLMs  
Quan Li, Tianxiang Zhao, Lingwei Chen, **Junjie Xu**, and Suhang Wang  
**BigData 2024**
- [11] HP-GMN: Graph Memory Networks for Heterophilous Graphs  
**Junjie Xu**, Enyan Dai, Xiang Zhang, Suhang Wang  
**ICDM 2022**

## EXPERIENCE

---

**Penn State University** 08/2021 - Present  
**Research Assistant** University Park, PA, US

### *Memory-Augmented LLM Agent with Reinforcement Learning*

- Developed a **memory-editing LLM agent** capable of structured CRUD operations over **long multi-turn dialogues** using **reinforcement learning** (PPO, GRPO, etc.) with per-edit **counterfactual QA rewards** and **evidence-aware retrieval**.
- Built a **VERL-based RL pipeline** with custom environments, trajectory data, multi-GPU rollout, and a unified memory/answering policy, enabling **long-range reasoning** and **persistent memory construction** in LLMs.

### *Multi-modal Knowledge Distillation with LLM for Molecule Property Prediction*

- Investigated the impact of **multi-modal inputs**—including text, molecular diagrams, node features, and graph structures—on molecule property prediction, and **fine-tuned** LLMs to strengthen their ability to encode and integrate these modalities.
- Implemented **knowledge distillation** from Large Language Models to smaller models (e.g., GNNs, MLPs), improving **efficiency** and reducing **inference cost** while preserving strong representation learning performance.

**Pinterest** 05/2025 - 08/2025  
**Machine Learning Intern** Palo Alto, CA, US

### *Cross-domain Ads Sequence Recommendation*

- Integrated ad and organic user sequences into a unified transformer-based model, learning shared latent representations that strengthen **cross-domain modeling** and improve **multi-source recommendations**.
- Explored **LLM-driven generative recommendation**, leveraging preference modeling and alignment techniques to produce personalized item suggestions and strengthen user engagement signals.

**Johnson & Johnson** 05/2024 - 11/2024  
**Research Intern** New Brunswick, NJ, US

### *Geometric Deep Learning for RNA Prediction*

- Generated and refined RNA datasets encompassing **1D, 2D, and 3D** structures; constructed graphs and geometric graphs based on 2D and 3D structures.
- Conducted extensive benchmarking of state-of-the-art models across 1D, 2D, and 3D methods, evaluating model **scalability, robustness to noise, and generalization** under real-world challenges.

- Developed **3D geometric RNA modeling** methods with hierarchical multidimensional GNNs, integrating multi-scale structures to achieve SOTA performance.

**Rice University**  
**Research Assistant**

05/2020 - 06/2021  
TX, US

### *Automated Time-series Outlier Detection System*

- Developed a **full stack and automated** machine learning system with preprocessing, feature extraction, detection algorithms, and human-in-the-loop interfaces.
- Integrated a wide range of algorithms including PyOD. Revisited the definition of the time-series anomalies and proposed a taxonomy for point-wise, piece-wise, and pattern-wise anomalies.
- Implemented AutoML for knowledge-free pipeline construction and automatic optimization of module combinations. Developed GUI to improve usability. [\[Code\]](#) (Github 1.6k+ stars, 200 forks); [\[Website\]](#); [\[Video\]](#).

## SERVICE

---

### Reviewer & Program Committee

NeurIPS (2025-2022); ICLR (2026-2024); ICML (2025-2024); AAAI (2026-2025); ICDM (2024-2022); LoG (2024-2023); KDD (2023-2022); WSDM (2023); CIKM (2023); WWW (2022); TKDD; PAKDD; ICWSM

## TEACHING EXPERIENCE

---

### Teaching Assistant, PSU

ETI 461: Database Management and Administration (Fall 2025); DS 310: Machine Learning for Data Analytics (Spring 2025); IST 597: Machine Learning on Graphs (Spring 2024); HCDD 364W: Methods for Studying Users (Spring 2024); DS 305: Algorithmic Methods & Tools (Fall 2023).

## HONORS & AWARDS

---

NeurIPS Scholar Award	NeurIPS, 2025
ICLR Student Travel Award	ICLR, 2025
IST Travel Award	IST PSU, 2022, 2025
ICDM Student Travel Award	ICDM, 2022
Graham Endowed Fellowship	PSU, 2021
Mitacs Globalink Research Scholarship	China Scholarship Council, 2020
Scholarship for Academic Excellence	HUST, 2017 - 2019

## SKILLS

---

<b>Languages</b>	English (Fluent), Mandarin (Native)
<b>Deep Learning</b>	PyTorch, PyTorch Geometric, DGL, e3nn, TensorFlow, PyTorch Lightning
<b>LLM &amp; Agents</b>	vLLM, HuggingFace, LoRA, RLHF, VERL, RAG, OmniOneRec (generative recsys) & VERL-agent (RLHF pipelines)
<b>Tools &amp; Infra</b>	CUDA, Weights & Biases, Ray, Docker, Linux, Git