

---

# Juyi Lin

lin.juy@northeastern.edu | (+1) 617-212-8191 | Northeastern University, Boston, US  
GitHub | LinkedIn | Homepage

## EDUCATION

---

**PhD student at Northeastern University, Boston, US** Sept 2024 - Now  
Major in Computer Engineering, GPA: 3.92/4, Advisor: Prof. Yanzhi Wang

**King Abdullah University of Science and Technology** Sept.2022 - May.2024  
Master of Science in Computer Science, GPA: 3.71/4, Advisor: Marco Canini

**Zhejiang University** Sept. 2018 - June 2022  
Bachelor of Engineering in Electronic Engineering, GPA: 3.75/4

## EXPERIENCES

---

**Infrastructure Internship** March 2022 - July 2022  
Beijing VirtAI Tech

- Integrated PaddlePaddle distribution library and Deep Graph Library(DGL) into the performance measurement system.
- Conducted an in-depth performance analysis of PaddlePaddle distributed training.

## PROJECTS

---

**Video Reasoning Benchmark** Sept 2025 – Jan 2026  
VBVR-EvalKit Team, Members are at JHU, CMU, Umich.

- Proposed and implemented VBVR-EvalKit, a benchmark for evaluating reasoning capabilities in video generation models, covering 40 models. Paper: link Code: link. Website: link. HF: link.
- Co-lead a team of 22 researchers, designing the data generation and benchmark framework, defining the development roadmap, and contributing core code and evaluation metrics. Launched an open-source ecosystem including 100+ synthetic data generators.
- Proposed and built a video reasoning dataset, covering 200 curated reasoning tasks and 1M+ video clips (1000× larger than prior datasets).

**Vision Language Action(VLA) Model** Jan 2025 – Sept 2025  
NEU & Funded by EmbodyX

- Proposed **VOTE**, an efficient fine-tuning framework for parallel action prediction in VLA models, reducing computational overhead and accelerating inference. Used by Cisco and EmbodyX. Paper link: VOTE. Code Link: Github.
- Proposed an ensemble voting strategy for the action sampling, improving model performance and enhances generalization across diverse tasks.
- Improved the average success rates of OpenVLA by over 20% across four LIBERO task suites, surpassed 7% average success rate of the state-of-the-art VLA model in SimplerEnv WidowX Robot, and accelerated action generation throughput by 39× on edge device NVIDIA Jetson Orin.
- Pretrained Llama3.2-1B LLM into a vision-language model (VLM) and fine-tuned with VOTE method, surpassing 7B OpenVLA’s average success rate by 16% across four LIBERO task suites.

## PUBLICATIONS

---

The \* sign denotes equal contribution.

VOTE: Vision-Language-Action Optimization with Trajectory Ensemble Voting *Preprint*  
**Juyi Lin**, Amir Taherin, Arash Akbari, Arman Akbari, Lei Lu, Guangyu Chen, Taskin Padir, Xiaomeng Yang, Weiwei Chen, Yiqian Li, Xue Lin, David Kaeli, Pu Zhao, Yanzhi Wang

A Very Big Video Reasoning Suite *Preprint*  
Maijunxian Wang\*, Ruisi Wang\*, **Juyi Lin\***, Ran Ji\*, Thaddäus Wiedemer, Qingying Gao, and others

Cross-Platform Scaling of Vision-Language-Action Models from Edge to Cloud GPUs *Preprint*  
Amir Taherin, **Juyi Lin**, Arash Akbari, Arman Akbari, Pu Zhao, Weiwei Chen, David Kaeli, Yanzhi Wang

---

ALTER: All-in-One Layer Pruning and Temporal Expert Routing for Efficient Diffusion Generation *NeurIPS 2025*  
Xiaomeng Yang\*, Lei Lu\*, Qihui Fan, Changdi Yang, **Juyi Lin**, Yanzhi Wang, Xuan Zhang, Shangqian Gao

**Open-Source Acceleration of Stable-Diffusion.cpp Deployable on All Devices** *Preprint*  
Jingxu Ng, Cheng Lv, Pu Zhao, Wei Niu, **Juyi Lin**, Minzhou Pan, Yun Liang, Yanzhi Wang

---

## PUBLIC SERVICES

---

- AAAI-26 Program Committee, EuroSys'23 Shadow PC Reviewer
- Artifact Evaluation: MLSys'23, EuroSys'24/25, OSDI'24 & ATC'24
- Open Source Promotion Plan 2023

---

## SKILLS

---

**Programming Languages:** Python, C++, Golang, Bash, Matlab, SQL, C

**Tools:** Git, LaTeX, Docker

**Libraries/Frameworks:** PyTorch, LoRA, Transformers, Diffusers, Deep Graph Library, PyTorch Geometric, Open-compass, LLaMA-Factory, Numpy