

# Wenjun Li

[wjli.2020@phdcs.smu.edu.sg](mailto:wjli.2020@phdcs.smu.edu.sg) • +65-89412807 • [homepage](#)

## EDUCATION

**Singapore Management University, Singapore, Singapore**

**May 2025**

PhD in Computer Science, Supervisor: Professor Pradeep Varakantham

**University of Southern California, Los Angeles, CA, U.S.**

**May 2020**

MS in Electrical Engineering, GPA 3.6/4.0

**China Jiliang University, Hangzhou, China**

**June 2018**

BEng in Opto-electronics Information Science and Engineering, GPA 3.7/4.0

Excellent Graduate; 1<sup>st</sup> Class Scholarship for Academic Excellence

## WORK EXPERIENCE

**Intern at Huawei Noah's Ark Lab, Singapore**

**June 2024 – Dec 2024**

- **Adaptive Tool Use in LLMs:** Introduced a novel algorithm, MeCo, for adaptive tool use in LLMs. MeCo captures emergent representations of high-level cognitive phenomena that quantify the LLM's meta-cognitive scores, guiding decisions on external tool usage. [ACL 2025 [2]].

**Intern at Geely, Hangzhou, China**

**June 2019- Aug 2019**

- Applied the Kanade-Lucas-Tomasi (KLT) algorithm for moving object detection (MOD) within the Active Parking Assist (APA) framework. Designed an algorithm to filter noisy flows, mitigated the impact of moving vehicles, and employed clustering methods to identify and extract information about moving objects.

## RESEARCH EXPERIENCE

**Large Language Models**

**March 2023 – Present**

- **Critical Analysis and Replication of R1-Zero Training:** investigated the impact of base models and RL on LLM reasoning, identifying pretraining biases and optimization inefficiencies in GRPO. Proposed Dr. GRPO to improve token efficiency and developed a minimalist R1-Zero recipe achieving 43.3% accuracy on AIME 2024 with a 7B model, setting a new state-of-the-art. [Preprint [1]]
- **Adaptive Tool Use in LLMs:** Introduced a novel algorithm, MeCo, for adaptive tool use in LLMs. MeCo captures emergent representations of high-level cognitive phenomena that quantify the LLM's meta-cognitive scores, guiding decisions on external tool usage. [ACL 2025 [2]].
- **Enhanced Fine-tuning Strategies:** Extensively explored the impact of fine-tuning on the planning capabilities of LLMs. Developed the Clustering-based Maximum Diversity Sampling (CMDS) algorithm to select diverse and representative data in the embedding space, significantly improving the sample efficiency for fine-tuning. [NAACL-findings 2025 [4]].

**Non-Expert Human Training with Environment Generation Approaches**

**Sept 2023 – Present**

- **Developed 3D Emergency Response Training Game:** Conceptualized and created an immersive 3D emergency response training game using Unity. Designed scenarios that challenge players to acquire and apply emergency response knowledge, including the use of diverse medical equipment for patient rescue.
- **Designed Adaptive Human Training Algorithm:** Developed an adaptive training algorithm based on the similarity and transferability of training tasks, enhancing training efficiency and overall outcomes. [Preprint [3]].

**Environment Generation for Training Generalizable RL Agents**

**June 2021 – May 2024**

- **Enhanced Environment Generation with Novelty Quantification:** Employed the GMM model to dynamically compute state-action space coverage and quantify the novelty score of environments. Developed a novelty-driven framework to improve existing environment generation algorithms. [Accepted as **Oral at NeurIPS 2024** [6]].
- **Integrated Meta-Reinforcement Learning Techniques:** First, encode student policy with validation performance on environments selected by Quality Diversity (QD). Then, integrated meta-reinforcement learning with environment generation to train a versatile teacher agent capable of generalizing across student agents with diverse learning properties. The student policy is encoded with Quality Diversity. [Accepted to AAAI 2024 [7]].
- **Developed Novel Distance Metric:** Introduced a distance metric based on the occupancy measure of agent

\* Equal contribution

trajectories to quantify environment distance. Utilized this metric to measure the diversity of training environments, achieving state-of-the-art generalization performance. [Accepted to IJCAI 2023 [8]].

- **Proposed a Novel Reward Signal:** Designed a reward mechanism based on marginal benefits to guide teacher agent in generating “right” training environments, resulting in improved computational efficiency and leading performance among RL-based environment generation algorithms. [Accepted as **Oral at AAAI 2025** [5]].

#### Wildlife Conservation Trust by Google

March 2021 – Jan 2022

- **Hierarchical CNN Model for Conflict Prediction:** Designed a hierarchical CNN model to predict human-wildlife conflicts. Conducted extensive experiments on the WCT dataset and co-authored a paper [9].
- **Addressing Data Sparsity with Sequential Decision-Making:** Applied sequential decision-making methods to determine optimal spatial granularity and address data sparsity issues in prediction. Integrated a predictive model and clustering technique with the CNN model. Conducted experiments and validated the proposed model’s performance on the WCT dataset.

#### Adversarial Training

Aug 2020 – March 2021

- **Ensemble Adversarial Training:** Proposed an ensemble adversarial training method that leverages data difficulty levels, achieving state-of-the-art performance under white-box attacks.
- **Risk-Averse Methods in Adversarial Training:** Implemented the Conditional Value at Risk (CVaR) approach in adversarial training and validated that the risk-averse method is not suitable for supervised adversarial learning.
- **Exploring GAN-Structured Pipeline:** Investigated the application of a GAN-structured pipeline in adversarial training.

#### SELECTED PUBLICATIONS

- [1] Liu\*, Chen\*, **Li Wenjun\***, et al. Understanding R1-Zero-Like Training: A Critical Perspective. [preprint]
- [2] **Li Wenjun**, et al. Adaptive Tool Use in Large Language Models with Meta-Cognition Trigger. [**ACL 2025**]
- [3] **Li Wenjun\***, Sidney Tio\*, Varakantham P. RL as a Surrogate: Teacher Algorithms for Human Learning. [preprint]
- [4] **Li Wenjun**, Chen Changyu, Varakantham P. Unlocking Planning Capabilities of Large Language Models with Clustering-Based Maximum Diversity Sampling. [**NAACL-findings 2025**]
- [5] Li, **Li Wenjun**, Varakantham P. Marginal Benefit Driven RL Teacher for Unsupervised Environment Design. [**AAAI 2025 Oral**]
- [6] Teoh Jayden\*, **Li Wenjun\***, Varakantham P. Improving Environment Novelty Quantification for Effective Unsupervised Environment Design. [**NeurIPS 2024 Oral**]
- [7] **Li Wenjun**, Varakantham P. Unsupervised Training Sequence Design: Efficient and Generalizable Agent Training. [**AAAI 2024**]
- [8] **Li Wenjun**, Varakantham P, Li D. Generalization through Diversity: Improving Unsupervised Environment Design. [**IJCAI 2023**]
- [9] Ghosh S, Varakantham P, Bhatkhande A, **Li Wenjun**, Milind Tambe. Facilitating human-wildlife cohabitation through conflict prediction. [**AAAI 2022**]

#### LANGUAGES

- English, Mandarin, German

\* Equal contribution