

WEI SHENG

1016 166th Pl NE, Bellevue, WA ◊ (+1) 206-390-9720 ◊ wshenguw@gmail.com

🌐 wsheng24.github.io ◊ [🌐 linkedin.com/in/weisheng0220](https://www.linkedin.com/in/weisheng0220) ◊ [🎓 Google Scholar](https://scholar.google.com/citations?user=wshenguw)

EDUCATION

Ph.D. in Computer Science

Purdue University - West Lafayette

08/2023 - 05/2028

Advisor: Rohan Paleja

B.S. Double Major in Computer Science & Mathematics

University of Washington - Seattle

08/2019 - 06/2023

GPA: 3.84/4.0

PUBLICATIONS

Workshop Papers, Symposium Papers, and Doctoral Consortia -

(W1) Sheng, W., & Paleja, R. (2026). "Influence-Salient Coordination Shaping for Scalable Cooperative MARL." In Proceedings of the ICLR Workshop on AI for Mechanism Design and Strategic Decision Making (AIMS).

(W2) Sheng, W., & Paleja, R. (2026). "Beyond Partner Diversity: An Influence-Based Team Steering Framework for Zero-Shot Human-Machine Teaming." arXiv preprint arXiv:2605.15400.

WORK EXPERIENCE

Software Engineer Intern | Fesco & Adecco, Shanghai, China

05/2024 – 08/2024

Skills: Java, Spring Boot, RESTful APIs, Microservices, LLM Evaluation

- ◆ Built an automated LLM evaluation pipeline, including test set construction, scoring, and regression tracking, to compare Ernie Bot, DouBao, and Qwen and support model selection decisions.
- ◆ Designed offline metrics and an online feedback loop, then analyzed evaluation logs with SQL/Pandas to slice results by cohort and error type, identify regressions, and prioritize prompt iterations to reduce failure modes.
- ◆ Developed a Spring Boot-based microservices backend and built RESTful APIs to automate candidate data evaluation, improving recruitment efficiency by 75%.

RESEARCH PROJECTS

Human-AI Collaboration & Influence Shaping in Multi-Agent RL

07/2025 – 05/2026

Skills: Python, PyTorch, Stable-Baselines3, LLM APIs, MARL

- ◆ Designed an influence-based RL framework for zero-shot human-machine teaming, enabling AI agents to guide interaction toward higher-performing coordination modes with unfamiliar partners.
- ◆ Extended cooperative RL to support mixed human-AI gameplay, including 1 Human + 2 AI settings, in a GPU-accelerated Overcooked environment with reproducible multi-seed evaluation.
- ◆ Conducted real human-subject studies and synthetic LLM-partner evaluations, with human-study results showing up to 44.7% higher task score in dyadic HMT and 13.3% higher task score in group HMT over learned-agent baselines.

Trajectory Prediction in Free-Walking VR Environments

01/2025 – 09/2025

Skills: Python, PyTorch, Transformers, OpenCV, Unity, Multi-Modal Sequence Modeling

- ◆ Developed a multi-modal predictive model with an encoder-decoder architecture for VR user trajectory forecasting, where the encoder captures temporal motion embeddings and the decoder integrates gaze direction and spatial scene layout via 2D floor plan masks.

- ◆ Achieved 15–20% reduction in Average Displacement Error (ADE) on the LocoVR dataset, demonstrating improved spatial-temporal accuracy over established baselines.
- ◆ Leveraged predicted destinations for intent inference, enabling downstream large language models to perform goal recognition in immersive environments.

LLM Safety Alignment via Soft Prompt Adversarial Training

01/2025 – 05/2025

Skills: Python, PyTorch, CUDA, Transformers, Pandas

- ◆ Designed a novel safety training framework combining universal adversarial soft prompting and Direct Preference Optimization to enhance the adversarial robustness of LLMs without sacrificing performance.
- ◆ Trained a soft prompt jailbreak generator to expose vulnerabilities in base models, achieving up to +25% increase in harmful activation rate during attack simulations.
- ◆ Improved model defense rate by over 70% across diverse adversarial attacks, while preserving performance on reasoning and instruction-following tasks.

Mars Ice Cap Layer Reconstruction

08/2024 – 12/2025

Skills: Python, C++, PyTorch, Scikit-learn, Open3D, VTK, CUDA, Teem, SHARAD Radar

- ◆ Built a scalable ETL-to-model pipeline for SHARAD-derived Mars ice-layer reconstruction, processing 123M+ triangle elements and converting raw geometry into model-ready training tables using VTK/Open3D.
- ◆ Implemented a modified ridge extraction algorithm based on Schultz’s Hessian analysis method, achieving 92% continuity of valid surface segments and reducing spurious detections by over 65% compared to Marching Cubes.
- ◆ Designed a feature-based classification model using Random Forest and geometric descriptors, improving region-level segmentation precision by $\sim 3\times$ in post-visual inspection benchmarks.
- ◆ Developed a computer vision-based segmentation pipeline using convolutional neural networks on radargram imagery, reducing spurious detections by over 60% and improving continuity of valid ice surfaces beyond classical feature-based methods.

SKILLS

Programming Languages: Python, Java, C++, JavaScript, SQL, Bash, C#, MATLAB, PHP, Ruby

Frameworks & Libraries: React, Node.js, Spring Boot, PyTorch, TensorFlow, Stable-Baselines3, Pandas, Scikit-learn, OpenCV, Unity

Systems & Tools: UNIX/Linux, Git, Docker, RESTful APIs, MySQL, PostgreSQL, Postman, AWS, CUDA, Slurm

Machine Learning & AI: Reinforcement Learning, Multi-Agent Reinforcement Learning (MARL), CTDE, RLHF, Preference Learning, Human-AI Collaboration, LLMs, Transformers, Sequence Models, Agent-Based Simulation

Computer Vision & Visualization: Deep Learning, CNNs, Vision Transformers, Generative Models, Image Segmentation, 3D Reconstruction, Scientific Visualization