Yuzhi Fan

Seattle, Washington |fdfxfanyuzhi@gmail.com | LinkedIn | website

Education

Bachelor of Science in Computer Science, **Sep. 2022 – Jun. 2026**

University of Washington Seattle. Seattle, WA.

GPA: 3.86 (Cumulative)

Selected Coursework: High-Performance Computing, AI&ML, Robotics & Embodied AI, Embedded Systems, Cybersecurity

Selected Publications

Ge Yan*, Jinghao Liu***, Yuzhi Fan****, Lei Cai, Minwen Liao, Jesse Zhang, Dieter Fox. *"Flex-π: A Multi-Stream World-Action Model with Compute Flexibility."* arXiv preprint arXiv:2608.10860, 2026.

Ge Yan, Shun Iwase, ***Yuzhi Fan***, Dieter Fox, Sergey Zakharov, Katherine Liu. *"UniFlow: A Unified Multi-Modal Policy for Robot Manipulation."* Conference on Robot Learning (**CoRL**), 2026.

Selected Skills

Microsoft Azure, Amazon AWS, Python, Java, C/C++, JavaScript, React, Next.js, TensorFlow, PyTorch, Machine Learning, Embedded Systems, Linux Shell, Networking, Docker, Kubernetes, MongoDB, PostgreSQL.

Experiences

Research Assistant | UW Robotics and State Estimation Lab | Flexible World Action Model **Nov. 2025 – Now**

- Drove the end-to-end research cycle for robot manipulation policies, from establishing real-world task setups and data collection methodologies to managing continuous train-and-deploy iterations.
- Participated in the core model design for multi-modal imitation learning, optimizing how policies process and fuse varying visual representations during real-world execution.

Full-Stack Software Engineer Intern | Xpectrum AI | Seattle, WA. **Mar. 2025 – Aug. 2025**

LLMs, Agentic AI, Retrieve-Augmented Generation, Cloud Computing, AWS, Docker, Python, JavaScript, Mongodb

- Redesigned cloud infrastructure by decoupling monolithic deployment into microservices, adding load balancing and autoscaling to improve scalability and fault tolerance.
- Led development of fallback customer service platform enabling seamless handoffs from AI agents to human representatives, architecting full-stack solution with Next.js frontend, FastAPI backend, and MongoDB database.
- Built full CI/CD pipeline using GitHub Actions, AWS ECR/ECS to automate build, test, and deployment workflows, significantly reducing release time.

Research Intern | Fudan University | The impact of AI on the labor market in China. **Jun. 2024 – Jan. 2025**

- Built an end-to-end NLP pipeline to process 40,000+ job postings (2015–2021) from Zhilian Recruiting
- Applied LLMs to classify ~16k AI-related job requirement sentences into ~300 unique AI-related skills.
- Implemented semantic matching with Sentence-BERT, mapping job descriptions to AI skills with threshold-based labeling.
- Built pipelines to measure AI skill intensity, enabling comparison across industries and changes over time within sectors.
- Demonstrated expertise in NLP, LLM-based text classification, and textual analysis within economics research context.

Data Analysis Intern | MetroData Tech Co. Ltd | Shanghai, China. **Aug. 2023 – Sep. 2023**

- Use Stata and Python Pandas to clean and select 250,000 restaurant data entries in Shanghai & Chengdu from Dianping (a Yelp-like platform).
- Generated models that predict the agglomeration of different types of restaurants in different locations.

Projects

Vision-Based Pedestrian Button Recognition | Deep-Learning research | UW **Apr. 2025 – Jun. 2025**

- Designed and implemented ML models to detect and classify pedestrian crosswalk buttons in urban environments.
- Created a custom dataset of 1200 real-world images; applied data augmentation to improve model robustness.
- Benchmarked three object detection pipelines (YOLO + CNN, unified YOLO, custom PyTorch model) for pedestrian button detection and direction classification; evaluated trade-offs in accuracy and task interference in multi-task setups.

Autonomous Vehicle Competition | Huawei Developer Competition 2022 |Huawei **Aug. 2022 – Sep. 2022**

- Create a dataset of traffic signs/scenarios of over 8000 pictures; label data and train a TensorFlow semantic segmentation model. Deploy the model onto a model car with a LiDAR, enabling autonomous detection based on lane markings and 10 different traffic signs, and pedestrians.
- Write Python ROS code to process and response to detection data, enabling auto driving on track.