

Rui Guo



587-588-6800



ruig0401@gmail.com



[linkedin.com/hoinky](https://www.linkedin.com/in/hoinky)

EDUCATION

The University of British Columbia

Master of Engineering in Electrical and Computer Engineering | GPA: 92.9/100

Sept 2024 - June 2026 (Expected)

Vancouver, BC, Canada

The University of Alberta

Bachelor of Science in Computing Science with Honors | GPA: 3.86/4.00

Sept 2022 - June 2024

Edmonton, AB, Canada

Beihang University

Bachelor of Engineering in Software Engineering | GPA: 86.7/100

Sept 2019 - Sept 2022

Haidian, Beijing, China

EXPERIENCES

UBC Deep Structure Learning (DSL) Lab

Research Assistant

Jan 2026 – Present

Vancouver, BC, Canada

- **Full-Stack Environment Setup:** Spearheaded the end-to-end local deployment and reproduction of **DinoV2** + **Llama-2** architectures; engineered robust data pipelines to process complex **RLDS** trajectory formats.
- **Parameter-Efficient Fine-Tuning:** Executed model fine-tuning (50K steps) using **LoRA** ($r = 32$), integrating **BF16** mixed precision, gradient accumulation, and checkpointing to optimize efficiency.
- **Closed-loop Evaluation:** Validated model performance across 10 spatial reasoning tasks in the **LIBERO** simulation environment, achieving an **85% task success rate**; developed automated scripts for ego-view behavioral analysis and visualization.

Huawei Technologies Co., Ltd

AI Software Engineer Intern

July 2025 – Dec 2025

Nanjing, Jiangsu, China

- **SFT Data Engineering:** Designed and implemented four FIM (Fill-in-the-Middle) splitting strategies based on code blocks; Introduced RAG similarity fragments to mitigate training-inference discrepancy; Optimized prompts via dynamic pruning and strong-correlation logic for code context.
- **Advanced Data Ratio Tuning:** Conducted extensive ablation studies on data granularity and ratios, significantly improving the model's offline code generation accuracy from 52.4% to 84.34%.
- **AI-Assisted Quality Control (Qurating):** Designed an automated quality inspection pipeline; Fine-tuned a lightweight model as a quality scorer to filter low-quality code, boosting accuracy of the high-quality subset to 68.39%.
- **DPO & Inference Optimization:** Constructed positive/negative sample pairs for DPO training targeting high-frequency production issues (e.g., bracket asymmetry), resulting in a 4% increase in real-world adoption rate and a 0.1s reduction in TTFT (Time to First Token).

PROJECTS

H-RDT: Heterogeneous Robot Fine-tuning | VLA, Flow Matching, PyTorch

Jan 2025 – March 2025

- Implemented multi-task manipulation for dual-arm robots in the **RoboTwin** simulation platform based on a **2.06B Vision-Language-Action (VLA)** model, achieving a **56.0% average success rate** across 5 complex tasks (e.g., bowl stacking, block tapping).
- Built a distributed training environment using the **Accelerate** framework; integrated **DINO** and **SigLIP** vision encoders with a 2-layer MLP adapter to map 4096-dim T5 text embeddings into the Transformer latent space.
- Adopted the **Flow Matching** generation paradigm, significantly improving action sequence generation efficiency (requiring only 5 inference steps) compared to traditional Diffusion (DDPM) methods.
- Optimized training efficiency by enabling **Gradient Checkpointing** and **BF16 Mixed Precision**, completing 30K steps of fine-tuning on a single RTX 5090; managed **HDF5** expert demonstration data for a 14-dimensional action space.

SKILLS & AWARDS

Languages: Python, Java

Technologies: PyTorch, Accelerate, DeepSpeed, Transformers

AI Frameworks: LLM Fine-tuning (SFT, DPO, LoRA), VLA Multimodal Models

Training Optimization: Mixed Precision (BF16), Gradient Accumulation, Checkpointing, RAG

Awards: Dean's Honor Roll (UofA), International Graduate Excellence Award (UBC), Beihang Freshman Scholarship