ZhijieXia

Curriculum Vitae

+(86)13736838512
zhijiexiacs@gmail.com
www.zhijiexia.dev
https://www.linkedin.com/in/zhijie-xia-678b331b5/
https://github.com/zhijie-os

Professional Summary

AI Researcher & Systems Engineer specializing in foundational Reinforcement Learning and training stability. Proven track record in designing adaptive learning approaches to resolve training-inference discrepancies in large-scale models. Deeply focused on advancing compute-efficient continual learning paradigms and streaming environments to build robust, scalable RL algorithms capable of real-world deployment.

Education

09/2019 - 04/2024 **Bachelor of Science in Computer Science (Honours)**, *University of Calgary*, Calgary, Canada

- Honors Thesis: "Android Permission System"
- Research Advisors: Dr. Ryo Suzuki (HCI/AR), Dr. Joel Reardon (Security)
- GPA: 3.92/4.0, Dean's List 2021, 2023
- Relevant: Human-Computer Interaction, Machine Learning, Computer Vision, Distributed Systems

Experience

11/2025 - 03/2026 **AI Infrastructure Engineer**, *Algorithm Team, Lingqu, Ascend Computing, Huawei*, Hangzhou, China

- Developed low-precision quantization (MXFP8/FP8) for Ascend NPU, accelerating LLM inference (Qwen, Deepseek) by 3-4x while maintaining model quality.
- Designed gradient alignment algorithm for quantized RLHF training, resolving inference-trainer pipeline mismatches.
- Co-first authored and submitted "ACRL: Adaptive Control of Training-Inference Discrepancy for Stable Reinforcement Learning" to ICML 2026.

09/2024 - 10/2025 **AI Infrastructure Engineer**, *Infrastructure Team, Lingqu, Ascend Computing, Huawei*, Zhejiang, China

- Engineered RackD middleware for distributed AI clusters, implementing RBAC authentication and pub-sub messaging for secure inter-node communication.
- Developed high-performance HCCL kernels in AscendC: 20+ base operators (AllReduce, AllGather) and fused operators (AllGatherMatmul, MatmulAllReduce).
- Accelerated transformer inference in Ascend Transformer Boost by 2.5x through fused operator integration vs. native PyTorch.

09/2023 - 08/2024 **Software Development Engineer**, *Lucid Vision Lab*, Vancouver, Canada

- Developed CMAS automation system and firmware for GigE camera production line.
- Automated 80% of manual tasks through end-to-end provisioning: JTAG BIOS flashing, firmware deployment, Linux OS installation.
- Engineered C++/Python validation pipeline with self-healing firmware retries, achieving 99.8% first-pass yield.
- Transitioned to a fully remote role from Hangzhou, China starting in May 2024 after concluding physical residency in Canada.

06/2023 - 08/2023 **Software Development Engineer**, *New Technologies(previously Knowd AI)*, Toronto, Canada

- Developed the company's first Minimum Viable Product (MVP), securing $500,000 in pre-seed funding through Entrepreneur First incubation.
- Designed an LLM-powered tool to streamline case study analysis for MBA and business students, improving efficiency and user experience.
- Implemented file parsing for multiple formats (MP3, Word, PDF, Excel, MP4), enabling lexical analysis of parsed text and semantic processing by an AI agent.

04/2022 - 11/2023 **Undergraduate Researcher**, *Programmable Reality Lab at University of Calgary*, Calgary, Canada

- Conducted HCI research in Dr. Ryo Suzuki's Programmable Reality Lab on augmented reality interfaces
- Led end-to-end research for "RealityCanvas" (UIST 2023 first-author): conceptual design, taxonomy analysis of 6 animation techniques, system implementation, and user studies with 20 participants.
- Contributed to "RealityEffects" (DIS 2024) development: implemented object tracking and conducted user evaluations.
- Resulted in 2 peer-reviewed publications at top HCI venues (UIST, DIS).

Technical Skills

HCI & AR	user study design, prototyping, Unity, 8th Wall, OpenCV, MediaPipe
AI & ML	RLHF, model quantization, distributed training, AscendC, CUDA, PyTorch
Programming	C++, Python, JavaScript, React, Node.js
Research	experimental design, literature review, statistical analysis, paper writing

Publications

Conference Proceedings

[1] Liao, Jian, Kevin Van, **Zhijie Xia**, and Ryo Suzuki. "RealityEffects: Augmenting 3D Volumetric Videos with Object-Centric Annotation and Dynamic Visual Effects." In Proceedings of the 2024 ACM Designing Interactive Systems Conference, pp. 1248-1261. 2024.

[2] **Xia, Zhijie**, Kyzyl Monteiro, Kevin Van, and Ryo Suzuki. "RealityCanvas: Augmented Reality Sketching for Embedded and Responsive Scribble Animation Effects." In Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology, pp. 1-14. 2023.

Awards and Honors

2023 **University of Calgary International Undergraduate Award**, *$5,000*
Awarded to high-achieving international undergraduate students at the University of Calgary based on academic merit.

2022 **Owen Family Scholarship**, *$636*
A scholarship awarded to students at the University of Calgary in recognition of academic performance.

2021 **University of Calgary International Undergraduate Award**, *$5,000*
Awarded to high-achieving international undergraduate students at the University of Calgary based on academic merit.

Certificates

2025 **HSDC Professional**, *Huawei*
Huawei Software Development Competency Certification, Professional Level (highest level).

2022 **TCPS 2: CORE 2022**, *Government of Canada*
Research Ethics based on the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans.

References

Ryo Suzuki
Assistant Professor
Director of the Programmable Reality Lab
University of Colorado Boulder, ATLAS Institute
ryo.suzuki@colorado.edu

Joel Reardon
Associate Professor
University of Calgary, Faculty of Science, Department of Computer Science
joel.reardon@ucalgary.ca