Zihuan (Ken) Jiang

jiangzihuan123@gmail.com 548-993-3583 https://github.com/zihuanjiang

EDUCATION

University of Toronto | CGPA: 3.88/4.00 *Sep 2021 – May 2026 (Expected)*

Bachelor of Applied Science in Engineering Science | Machine Intelligence

Dean's Honour List (2021F, 2022F, 2022W, 2023F, 2023W, 2024W)

TECHNICAL SKILLS

Programming Languages: Python, C/C++, SQL, RISC-V Assembly.

Libraries: PyTorch, Transformers, bitsandbytes, OpenCV, CVXPY, NumPy, OmegaConf, argparse.

Machine Learning: Diffusion, LLM, RAG, DDP, FSDP, Attention, CNN, RNN, K-means, PCA.

Algorithms: Dynamic programming, Greedy Algorithm, LP, QP, QCQP.

Productivity tools: LaTeX, Microsoft Office, Neo4j.

Languages: Chinese (native), English (Fluent).

WORK EXPERIENCE

Research Assistant *May 2024 – Apr 2025 (Expected)*

Company: Huawei Canada, Markham, Ontario

- Developed a GenerativeUI pipeline by fine-tuning large language models using LoRA techniques for UI code generation, optimizing GPU cluster utilization through Fully Sharded Data Parallel, and resolving UI element conflicts through mixed-integer linear programming.
- Architected and deployed a RAG system incorporating knowledge graph construction and vector embeddings, enabling efficient retrieval of code examples and technical documentation.
- Contributed to Image-to-Video generation research through data collection and preprocessing, assisted with hyperparameter optimization experiments, and conducted comparative analysis against state-of-the-art models.
- Conducted detailed research on SOTA quantization methods (i.e. VPTQ, Quip#) on open-sourced LLMs (i.e. llama, mistral), evaluating algorithm performance via benchmarks including quantization cost and token generation speed, and model performance metrics such as MMLU.

PROJECT EXPERIENCE

Research Workflow Study *Mar 2024 – Apr 2024*

- Simulated a complete research workflow through development of a GAN-based image super-resolution project, gaining hands-on experience in model architecture design, training pipeline development, and technical documentation under practical constraints.

Cross-Dataset Learning with ResNet Models *Feb 2024 – Mar 2024*

- Conducted experimental analysis of ResNet architectures across SDD and DBI datasets, investigating cross-dataset performance patterns and implementing dataset source classification through fine-tuning experiments with varied dropout rates and pre-trained weights.

Additional Technical Projects *Nov 2021 – Apr 2024*

- Developed SQL schemas based on natural language specifications.
- Implemented a three-level copy-on-write page table mechanism in kernel space.
- Developed a linked list-based order management system for restaurant operations.
- Conducted structural engineering analysis through hands-on Matboard Bridge construction, performing systematic stress testing and failure point analysis.