WALKER CHI

walker.chi.000@gmail.com | github.com/walkerchi | Google Scholar

Foundation Models · GPU Systems · Scientific Computing

Research engineer with experience developing LLMs and VLMs at Moonshot AI and MiniMax. Work spans pretraining data at TB-to-PB scale, multimodal quality modeling, training infrastructure, agent evaluation, and GPU compiler capability. Independent research focuses on differentiable sparse linear algebra and GPU-native simulation.

EXPERIENCE

Moonshot AI

Apr 2026 - Sep 2026

Scaling LLM Researcher | Beijing, China

- **Pretraining data pipeline and scaling:** Led Kimi's end-to-end pretraining data pipeline, spanning acquisition, crawling, procurement, parsing, and processing. Conducted multi-epoch and scaling-ladder experiments to evaluate data strategies and guide adoption in large-scale pretraining.
- **Agent evaluation:** Contributed to Agent Last Exam (ALE) benchmark development through seed-data collection, task construction, and agent-trajectory generation to evaluate planning, reasoning, and execution.
- **GPU compiler capability:** Independently built a MiniTriton kernel-compiler case study to probe Kimi K3 GPU kernel generation. Generated PTX outperformed Triton on selected kernels, including KDA, and achieved comparable performance on others; evaluated compiler reasoning, correctness, and optimization.

MiniMax

Sep 2024 - Apr 2026

VLM Researcher | Shanghai, China

- **M-series foundation models:** Contributed to multiple generations of M-series models through LLM/VLM data construction, pretraining experiments, hyperparameter search, evaluation, and ablation studies.
- **End-to-end data pipelines:** Owned multimodal pretraining data sourcing, processing, quality control, and delivery for M-series development; ran targeted data experiments on world knowledge, spatial understanding, and code.
- **Multimodal quality modeling:** Designed and implemented reward/quality models for pretraining data selection. Built training workflows across model scales and datasets with Hugging Face and Megatron-LM for scoring, filtering, and domain bucketing.
- **Data processing at scale:** Built TB-to-PB-scale LLM/VLM processing pipelines combining rule-based and model-based cleaning, deduplication, quality filtering, and format normalization; produced over **20T tokens** of interleaved multimodal training data.
- **Data experimentation:** Established pretraining-data unit tests and ablation workflows to measure the marginal effects of data quality, mixture composition, and processing choices on model capabilities.
- **Training infrastructure:** Implemented multimodal DPO in Megatron-LM; refactored binary-data construction and packing; contributed to evaluation-platform refactoring, resolved scheduling deadlocks, and improved the training, evaluation, and visualization workflow.

Tencent

Jul 2021 - Sep 2021

Algorithm Intern, TEG Information Security | Shenzhen, China

- **Graph neural network optimization:** Applied sparse computation and JIT compilation to GEN (Graph Estimation Neural Network), reducing memory usage to approximately **1/1,000** of the original implementation and accelerating inference by approximately **15x**.

SELECTED RESEARCH & OPEN-SOURCE PROJECTS

torch-sla
ICML 2026

Differentiable Sparse Linear Algebra with Adjoint Solvers
arxiv.org/abs/2601.13994

- Developed a GPU-native sparse linear algebra library with automatic differentiation and adjoint solvers.
- Introduced Sparse Tensor Parallelism for distributed multi-GPU sparse solving; conducted experiments with up to approximately **400 million degrees of freedom**.

TensorMesh
ICML 2026

tensor-mesh.com

- Developed a high-performance, PyTorch-native finite element framework with automatic differentiation, GPU acceleration, and distributed computing.
- Supported PDE solving, inverse problems, and engineering optimization within a differentiable computational workflow.

TensorGalerkin
ICML 2026

Learning, Solving and Optimizing PDEs
arxiv.org/abs/2602.05052

- Developed efficient Galerkin assembly and differentiable PDE-solving methods to unify PDE learning, forward solving, and inverse optimization.
- Analyzed the numerical properties of weak-form optimization relative to strong-form physics-informed neural networks (PINNs).

EDUCATION

ETH Zurich
2022 - 2024

MSc, Computational Science and Engineering | GPA: **5.5/6.0**

RWTH Aachen University
2021 - 2022

Exchange Student, Computer Science

Huazhong University of Science and Technology
2018 - 2022

BSc (Honors), Computer Science and Technology
Top **3%** within the School of Computer Science and Technology | National Scholarship | Mathematical Contest in Modeling: Finalist