

Nima Kelidari

Los Angeles, CA | kelidari@usc.edu | (425) 520-4390 | kelidari.com

linkedin.com/in/nima-kelidari | github.com/Nikelroid

Education

University of Southern California - Viterbi School of Engineering

Los Angeles, CA

Master of Science in Computer Science - Artificial Intelligence | GPA: 3.55/4.0

Dec 2024 – Dec 2026

- Coursework: Natural Language Processing, Multi-modal Probabilistic Learning, Machine Learning, Deep Learning, Robotics

Sharif University of Technology

Tehran, Iran

Bachelor of Science in Computer Science | GPA: 3.37/4.0

Sep 2019 – Feb 2024

- 3.5/4.0 across 15+ CS/AI courses: Computer Vision, Deep Learning, Statistical Learning, Algorithm Design, Numerical Analysis, Image Processing

Experience

Memorial Sloan Kettering Cancer Center - Department of Medical Physics

New York, NY (Remote)

Computational Optimization Intern

May 2026 – Present

- Accelerated the convex quadratic-programming solver behind **IMRT radiotherapy treatment planning** – a path-following ADMM warm-starting a Mehrotra interior-point method – on clinical cases up to **~380K variables** / **~570K constraints**, GPU-accelerated (CuPy, mixed precision) on V100 nodes
- Benchmarked first-order methods on clinical lung and prostate cases: PDHG/PDQP (JAX/MPAX), ADMM with over-relaxation and adaptive penalties (residual-balancing, OSQP, spectral/ARADMM), Fast-ADMM restart, and safeguarded Anderson acceleration
- Designed a combined ADMM variant (aggressive spectral penalty recovery + safeguarded Anderson mixing) that cut iterations to a usable warm start by **~2×** (up to **~4×** from a mis-set penalty) – the only method to reach the tight KKT tolerance on both cancer sites
- Built a PDQP→interior-point hybrid reaching a **~100×** **tighter optimality gap** than the production solver ($7 \times 10^{-5}\%$ vs $7 \times 10^{-3}\%$), plus a matrix-free warm-started conjugate-gradient GPU engine and a SLURM harness with verified CPU/GPU parity

University of Southern California - Viterbi School of Engineering

Los Angeles, CA

Graduate Student Researcher

Aug 2025 – Present

- Inference-Time Steering for Mixture-of-Experts LLMs** (Profs. Jia, Ma): designed a 50K-parameter learned low-rank head redirecting MoE routing with zero weight changes, trained contrastively on SQuAD, HotpotQA, and FEVER; **+0.101 mean steering gain** on six faithfulness benchmarks – **4×** **the SteerMoE baseline** (ICLR 2026) – transferring across three MoE families (7B–47B); manuscript in preparation
- Adversarial Co-Evolution of RL and LLM Agents** (Prof. Liu): built a distributed PPO/TRPO pipeline with a rising-opponent curriculum over **100+ controlled runs**, reaching a **99% win rate vs. baseline agents**; engineered a master-worker LLM serving stack (Llama-3, Gemma, GPT-OSS; 14 GPU workers, ~32 queries/s, **62×** **faster model loading**) to bring LLM opponents into the training loop; preprint: arXiv:2607.06854
- Speech Emotion Interpretability** (Prof. Soleymani, USC ICT): trained sparse autoencoders (**93% variance explained**) on ~940K frame activations across six speech encoders, isolating **61 confound-controlled emotion features** (60/61 acoustic); built the SLURM probing harness (6 encoders × 8 datasets) and FastAPI dashboard

Sharif University of Technology

Tehran, Iran

Research Assistant - Computer Vision

Sep 2022 – Aug 2023

- Conducted ten months of research with Dr. Tefagh on a 4-member team, building generative-AI pipelines with Vision Transformers and 3D Convolutional Neural Networks on university HPC/GPU clusters

Sharif University of Technology

Tehran, Iran

Teaching Assistant and Tutor

Sep 2021 – Jul 2023

- Served as TA for 4 semesters across 6 courses (Basic/Advanced Programming, Computer Fundamentals, Machine Learning): presented lectures, designed assignments, and graded exams and projects for **300+ students** alongside Head TAs
- Built and delivered TensorFlow workshops and diagnostic sessions for 60 students in the ML course, with a **65% attendance rate**

Projects

Hybrid Anime Recommender System

[Link to GitHub](#)

- Built a hybrid neural collaborative-filtering recommender in TensorFlow – **128-dim user/anime embeddings** feeding a 4-layer deep network with layer normalization and bias terms – over **80M+ interactions** (300K users × 14K titles)
- Versioned datasets and checkpoints with DVC on Google Cloud Storage, tracked every run in Comet ML, and served ranked recommendations via FastAPI
- Automated an **8-stage Jenkins CI/CD pipeline** (DVC pull, Docker build, GCR push, GKE deploy) behind a Kubernetes LoadBalancer

Multi-Modal Sentiment Classification

[Link to GitHub](#)

- Fused RoBERTa, ViT, and wav2vec2 in a three-class sentiment network for image-text-audio dialogue – **69% accuracy** on MSCTD
- Refactored notebook research into a typed, modular pipeline with one-command data ingestion and graceful missing-modality fusion
- Queued training on SLURM with per-batch Weights & Biases logging; served real-time predictions from a FastAPI app with an interactive upload UI

Image Compression via SVD & FFT

[Link to GitHub](#)

- Designed a lossy compression pipeline: per-RGB-channel **singular value decomposition** with adjustable rank plus FFT Gaussian low-/high-pass filtering
- Cut file size **~4×** at minimal perceptual loss by retaining top singular values; packaged U/S/V factors and config into a custom compressed archive format
- Authored a written analysis of rank-truncation vs. reconstruction-quality trade-offs (Python/OpenCV, MATLAB)

Skills

Programming: Python (fluent), C++, Java, SQL, R | **ML/DL:** PyTorch, JAX, TensorFlow, Hugging Face, vLLM, Stable-Baselines3, scikit-learn, OpenCV | **Optimization:** convex QP, ADMM, interior-point & first-order methods, Anderson acceleration | **GPU & HPC:** CUDA, CuPy, SLURM, distributed training | **MLOps & Cloud:** Docker, Kubernetes, GCP, AWS, MLflow, DVC, Jenkins, Weights & Biases, FastAPI | **AI-Assisted Engineering:** AI coding agents (Claude Code) & LLM APIs for reviewable development and CI/CD | **Robotics:** ROS, Kalman & particle filters, motion planning