Aaron Leevord Fernandes

AWS Certified Developer Associate — +1 (930) 333-4174 — aaron.fernandes.ml@gmail.com
aaronferns.com — github.com/Aaronaferns — linkedin.com/in/aaron-leevord-fernandes

SUMMARY ML Engineer with experience developing computer vision, generative AI, and computational imaging systems using PyTorch, DeepSpeed, and scalable GPU infrastructure. Skilled in diffusion models, image synthesis, representation learning, and high-performance training and inference pipelines for visual AI applications. Experienced in building AI imaging workflows, surrogate modeling systems, and computer vision pipelines for autonomous driving, scientific imaging, and reinforcement learning applications. Strong programming background in Python and C/C++ with experience in GPU-accelerated computing, Linux environments, and scalable ML systems.

EDUCATION

Indiana University Bloomington — M.S. Intelligent Systems Engineering — **GPA: 3.9/4.0** May 2025
- Research Areas: Meta-Optics, medical Image Processing, Reinforcement Learning, Generative Models (Diffusion)
- Coursework: Computer Vision, Deep Learning Systems, Reinforcement Learning, NLP, Cloud Computing, Multivariate Statistics

Goa College of Engineering — B.S. Electronics and Telecommunications Engineering — **GPA: 3.7/4.0** Sept 2021
- Research Areas: Computer Networking, IoT

TECHNICAL SKILLS

Computer Vision & Imaging: Computer Vision, Computational Imaging, Image Processing, Diffusion Models, Generative AI, Representation Learning, Vision Transformers, Self-Supervised Learning, Image Synthesis, Visual AI

AI Frameworks & Deployment: PyTorch, TensorFlow, ONNX, CoreML, HuggingFace, vLLM, Mixed Precision Training, Inference Pipelines

Distributed Training & GPU Systems: DeepSpeed (ZeRO 1/2/3), DDP, Multi-GPU Training, GPU Acceleration, SLURM, MLflow, W&B, High-Performance Computing (HPC)

Systems & Infrastructure: Linux, Docker, FastAPI, REST APIs, AWS, PostgreSQL, Git, CI/CD

Languages: Python, C/C++, C, Java, JavaScript, SQL

EXPERIENCE

Graduate Research Assistant — Synthetic Driving Data for Autonomous Driving (Industry Sponsored) — Indiana University Mar 2026 – Present
- Evaluated and benchmarked diffusion-based generative models and driving world-model video generators for synthetic data generation in autonomous driving applications.
- Designed distributed multi-GPU workflows for large-scale inference, benchmarking, data processing, and experiment orchestration.
- Curated multimodal driving datasets and developed preprocessing pipelines to support training and evaluation of generative and perception models.
- Analyzed model behavior, controllability, robustness, and downstream performance under long-tail and distribution-shift scenarios.

Machine Learning Engineer (Computer Vision) — Leung Research Group Apr 2024 – Dec 2025
- Developed self-supervised computer vision and computational imaging pipelines for OCT video-to-histology translation using topology-constrained U-Net registration and spectral-domain modeling techniques.
- Engineered GPU-accelerated optical simulation workflows for metasurface imaging systems using large-scale HPC infrastructure (SLURM/Linux).
- Built optimized U-Net surrogate models to approximate computationally intensive FDTD simulations, reducing inference latency from ~**3 mins to sub-second execution** while preserving structural image fidelity **(SSIM $>$ 0.85)**.
- Improved image reconstruction throughput and model evaluation efficiency through scalable GPU training workflows and optimized data-processing pipelines.

Machine Learning Engineer (Reinforcement Learning) — Indiana University Jan 2025 – Jun 2025
- Implemented goal-conditioned reinforcement learning algorithms including Hindsight Experience Replay (HER) and curriculum-based policy optimization methods for sparse-reward robotic navigation tasks.
- Built scalable RL experimentation infrastructure using Hydra, MLflow, and W&B to orchestrate **100+** multi-seed distributed training runs and automated hyperparameter optimization workflows.
- Improved training reproducibility, experiment tracking, and evaluation efficiency across reinforcement learning pipelines.

Software Engineer — Tata Consultancy Services (HDFC Bank) Jan 2022 – Jan 2023
- Developed and maintained Spring Boot microservices supporting **100K+** daily financial transactions in a regulated enterprise banking environment.
- Managed production release cycles, CI/CD workflows, backend services, and cross-functional engineering coordination for distributed enterprise systems.

PROJECTS

Multimodal Image Generation (Diffusion & Rectified Flow) — PyTorch, DiT, DeepSpeed Sept 2025 – Present
- Implemented diffusion and rectified-flow transformer architectures in PyTorch for multimodal image generation using CLIP conditioning and classifier-free guidance.
- Built end-to-end training and inference pipelines incorporating EMA parameter tracking, Euler ODE sampling, mixed-precision training, and FID-based evaluation.
- Scaled distributed training across **8× L40S GPUs** and conducted large-model experimentation on **2×L40s** GPUs using DDP, bf16 optimization, and SLURM orchestration.

- Profiled single and multi-GPU workloads, analyzing forward/backward execution, memory consumption, communication overhead, and data-loading bottlenecks in distributed training environments.
- Developed high-throughput preprocessing and streaming pipelines for a **300+ GB** LAION-Aesthetics image-text dataset supporting large-scale generative model training.

LLM-Assisted Qualitative Analysis and Retrieval System (NICC Brussels Collaboration) — Python, Neo4j, LlamaIndex Oct 2024 – Dec 2024

- Built LLM-assisted retrieval pipelines for qualitative interview analysis across **200+** transcripts using vector embeddings and graph-backed retrieval systems.
- Implemented scalable document ingestion, graph extraction, and source-traceable querying workflows using Neo4j, Pinecone, and LlamaIndex.
- Worked with **External Collaborators** (Researchers from NICC Brussels).

Efficient Model-Based Reinforcement Learning for Robotic Planning — MPC, World Models Aug 2024 – Dec 2024

- Designed model-based reinforcement learning pipelines using learned world models and Model Predictive Control for long-horizon robotic planning tasks.
- Implemented a learned world model based algorithm, planning with Cross-Entropy Method (CEM) and Model Predictive Control (MPC) for trajectory optimization; achieved **2× sample efficiency** vs SAC baseline on HalfCheetah and Reacher

Unsupervised Industrial Defect Detection — PyTorch, Vision Transformers Dec 2024 – Jan 2025

- Designed a ViT-B/16 autoencoder with Gaussian Mixture Model anomaly scoring, improving BTAD AUROC from **0.78 to 0.94** over ResNet-50 baselines.
- Optimized representation learning and anomaly detection performance for industrial computer vision inspection workflows.