

Jonathan Muhire

Software Engineer & ML Researcher

(+1) 572-205-5073 | muhirejonathan123@gmail.com | jonathanmuhire.com | github.com/Jonathan-321

M.S. Artificial Intelligence candidate and Robotics Founder with specialized expertise in Distributed Systems, Computer Vision, and LLM Evaluation. Secured VC funding to architect data infrastructure for embodied AI and engineered adversarial test suites for frontier models at Scale AI. Selected for Google Summer of Code to optimize deep learning pipelines.

EDUCATION

Oklahoma Christian University

B.S. in Computer Science & M.S. in Artificial Intelligence (Concurrent Degrees)

Edmond, OK

Expected Apr. 2026

GPA: 3.89/4.0 | Completing both degrees in 4 years through advanced standing

Status: Completing degree requirements asynchronously; available for immediate full-time employment.

PROFESSIONAL EXPERIENCE

Neotix Robotics

Founder & Lead Engineer

May 2025 – Present

New Haven, CT

- Founded a robotics infrastructure startup and secured \$20,000 in venture funding (Yale Tsai City Ventures) to build scalable data collection systems for embodied AI.
- Architected a distributed data pipeline processing over 500GB of multimodal robotics data (RGB-D, proprioception); implemented MinIO and LakeFS to enable Git-like version control for large-scale datasets.
- Engineered a real-time teleoperation framework using ROS2, integrating SAM2 (Segment Anything) and ORB-SLAM3 to achieve sub-millisecond latency for dual-arm manipulation tasks.

Scale AI

AI Safety & Code Evaluation Specialist

May 2024 – Aug. 2024

San Francisco, CA

- Engineered adversarial test suites for frontier language models (GPT-4, Claude), directly contributing to the safety alignment of code generation capabilities for enterprise partners (OpenAI, Meta).
- Audited 3,000+ lines of generated code across complex algorithmic tasks, identifying critical vulnerabilities including SQL injection, race conditions, and memory leaks.
- Authored technical evaluation specifications for 500+ multi-step coding tasks (async jobs, API integrations), standardizing quality benchmarks for a distributed team of technical reviewers.

RESEARCH & OPEN SOURCE EXPERIENCE

RenAissance: Document Understanding Pipeline

May 2025 – Aug. 2025

Google Summer of Code '25 (Deep Learning Engineer)

- Selected as a top contributor (<5% acceptance) to build a historical manuscript digitization pipeline using LayoutLMv3 Transformers.
- Optimized inference throughput by ~40% via quantization and batching, achieving SOTA performance on layout analysis benchmarks for degraded historical documents.
- Implemented comprehensive logging and error handling, ensuring production-level reliability for large-scale document processing jobs.

Data & ML Systems Research

Graduate Researcher

Jan. 2024 – Present

Edmond, OK

- Designed ETL pipelines in Python and SQL to transform raw sensor logs into curated training tables, enabling reproducible experiments for vision and control models.
- Implemented telemetry for long-running distributed training jobs, reducing debugging time for model convergence issues by providing granular resource usage metrics.

SELECTED PROJECTS

smolVLA: Efficient Vision-Language-Action Models

June 2025 – Aug. 2025

- Distilled a 7B-parameter OpenVLA teacher into a smaller VLA optimized for robotics manipulation, utilizing knowledge distillation and structured pruning.
- Deployed model artifacts to edge devices, maintaining 95% of teacher performance on manipulation success rates while significantly reducing memory footprint.

Bimanual Teleoperation & SLAM Pipeline

June 2025 – Aug. 2025

- Built a dual-arm teleoperation system capturing synchronized video and control signals, creating a foundational dataset for training Diffusion Policies.
- Implemented ORB-SLAM3 for 7-DOF trajectory extraction, enabling precise 3D reconstruction from monocular video streams.

TECHNICAL SKILLS

- Languages:** Python (Advanced), SQL, C++, JavaScript/TypeScript, MATLAB
- AI & Computer Vision:** PyTorch, TensorFlow, Hugging Face, JAX, SAM2, ORB-SLAM3, OpenCV, LLMs
- Robotics:** ROS/ROS2, MuJoCo, PyBullet, NVIDIA Isaac Gym, Real-Time Control
- Cloud & Infra:** AWS (S3, Lambda), Docker, MinIO, LakeFS, Distributed Systems, CI/CD, PostgreSQL