# 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 software engineer focused on C++, robotics data pipelines, and ML evaluation. Built dataset storage/versioning with MinIO and LakeFS, prototyped ROS2/C++ teleoperation with perception components, and performed code-generation evaluations at Scale AI.

## EDUCATION

**Oklahoma Christian University**                                                                                                    Edmond, OK
*B.S. in Computer Science & M.S. in Artificial Intelligence (Concurrent Degrees)*                              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**                                                                                                                  May 2025 – Present
*Technical Lead & Founder*                                                                                                              New Haven, CT
- Received $20,000 in non-dilutive support from Yale Tsai City Ventures to fund R&D on robotics data collection tooling.
- Built a data pipeline to store and version ∼500GB of captured RGB-D and geometry logs using MinIO and LakeFS, enabling reproducible dataset snapshots and rollbacks.
- Developed a ROS2/C++ teleoperation and data-capture prototype; integrated SAM2 (segmentation) and ORB-SLAM3 (tracking) into the perception stack.

**Scale AI**                                                                                                                          May 2024 – Aug. 2024
*AI Safety & Code Evaluation Specialist*                                                                                          San Francisco, CA
- Created adversarial and edge-case evaluations for code-generation models, focusing on security-sensitive prompts and robustness failures.
- Reviewed and annotated 3,000+ lines of model-generated code, documenting issues such as injection risks, race conditions, and memory-management bugs.
- Wrote evaluation specifications and scoring rubrics for multi-step programming tasks (async workflows, API integrations) to support consistent review across a distributed team.

## RESEARCH & OPEN SOURCE EXPERIENCE

**RenAIssance: Document Understanding Pipeline**                                                                      May 2025 – Aug. 2025
*Google Summer of Code '25 (Deep Learning Engineer)*
- Contributed to an open-source historical manuscript digitization pipeline using LayoutLMv3; implemented training/inference utilities and evaluation scripts.
- Improved inference throughput by applying quantization and batching; validated speedups on the project benchmark configuration.
- Implemented structured logging, input validation, and failure handling to support large batch processing runs.

**Data & ML Systems Research**                                                                                                  Jan. 2024 – Present
*Graduate Researcher*                                                                                                                         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 by logging resource usage and run metadata to reduce debugging time for convergence issues.

## SELECTED PROJECTS

**smolVLA: Vision-Language-Action Experiments**                                                                      June 2025 – Aug. 2025
- Ran experiments with knowledge distillation and structured pruning for vision-language-action policies, evaluating size/performance tradeoffs in controlled tests.
- Packaged inference artifacts and profiled memory and latency on target hardware for deployment feasibility.

**3D Reconstruction & Teleoperation Pipeline**                                                                        June 2025 – Aug. 2025
- Built a teleoperation data-capture pipeline for synchronized video and control signals to support policy learning experiments.
- Used ORB-SLAM3 for trajectory estimation and 3D reconstruction from monocular video; exported trajectories for downstream modeling.

## TECHNICAL SKILLS

- **Languages:** C++, Python, SQL, JavaScript/TypeScript, MATLAB
- **AI, Vision & Geometry:** PyTorch, TensorFlow, SAM2, ORB-SLAM3, OpenCV, Transformers, 3D Reconstruction
- **Simulation & Robotics:** MuJoCo, PyBullet, NVIDIA Isaac Gym, ROS/ROS2, Teleoperation, Data Capture
- **Cloud & Infra:** AWS (S3, Lambda), Docker, MinIO, LakeFS, Distributed Systems, CI/CD, PostgreSQL