

Atharva Gupta

(216) 559-9392 | atharva.jgupta@gmail.com | [Portfolio](#) | [GitHub](#) | [LinkedIn](#) | Cleveland, OH

Software Engineering | Distributed Data, Retrieval & AI Systems

EDUCATION

Case Western Reserve University

B.S. Computer Science & Applied Mathematics; GPA: 3.6

Cleveland, OH

Expected December 2027

EXPERIENCE

AI Engineering Intern

Cisco

Jun 2026–Present

San Jose, CA

- Built a Kubernetes deep-research system coordinating planner, retriever, and synthesis agents with MCP tools to deliver evidence-grounded, source-cited answers.
- Built PySpark/PyArrow ETL processing 28K+ external records per quarterly cycle, live feeds, and Common Crawl into a Blob-backed evidence lake and adaptive semantic index; orchestrated updates with Azure Functions.
- Checkpointed agent state by session, sandboxed runtimes in Podman, and streamed LangSmith traces to evaluate grounding, tool use, latency, failures, and cost.
- Designed Cisco/Splunk’s external-signal churn platform, selected among 7+ public/third-party data providers, and partnered with vendor engineers to join external evidence with internal telemetry.

Full-Stack Software Engineer Intern

PulseImaging.AI

Sep 2025–May 2026

Cleveland, OH

- Built a TypeScript/Electron DICOM viewer with three-plane MPR (axial, sagittal, coronal) and 8-slice concurrent loading for interactive study navigation.
- Engineered DICOM ETL for university-hospital studies, validating the three-level study/series/instance hierarchy and normalizing metadata and pixels into reproducible training and inference datasets.
- Served PyTorch U-Net segmentation through NVIDIA Triton/FastAPI on AWS, achieving sub-100-ms batch latency and under 2 minutes per full scan.

Software Engineering Intern

CASFER

Jun 2025–Aug 2025

Cleveland, OH

- Wrote concurrent, distributed Go crawlers with a shared manifest to coordinate nationwide ingestion across 2,229 HUC-8 watersheds and multiple geospatial APIs.
- Built PySpark/PyArrow ETL on CWRU’s Slurm cluster to normalize an estimated 20M+ nationwide nitrate/phosphate raster cells from GeoTIFF/NetCDF into Geohash-partitioned source, sink, and river-network datasets; recovered 22 HUC-8s lost to identifier mismatches.
- Prepared 60K+ geolocated AFO/CAFO image records (40K+ AFO; 20K+ CAFO) with PySpark on AWS Glue; transfer-learned YOLOv8 for bounding-box detection and facility classification.

SELECTED PROJECT

Robodex | [GitHub](#) | [Docs](#) | **Program:** YC Startup School (2026)

Apr 2026–Present

- Built a Next.js/TypeScript interface for a local-first robot-design agent that converts task specifications into ranked, constraint-validated designs with deterministic post-processing and revisioned human review.
- Measured a bounded 100-run grammar-loop slice at 81% compile safety, 0.835 mean compiler validity, and 0.736 mean task alignment; retained LangSmith traces for 97 returned runs.

RESEARCH & PUBLICATION

- **O. D. Akanbi, A. Gupta, et al. “Towards Circular Nutrient Economies.”** *Resources, Conservation and Recycling* 226 (2026), 108697. Second author; curated and validated data for a national study spanning 2,229 watersheds. [DOI](#)

TECHNICAL SKILLS

Languages: Python, Go, TypeScript/JavaScript, SQL

Systems & Data: distributed systems, concurrency, Unix/Linux, information retrieval, PySpark, PyArrow, ETL, object storage/data lakes, Slurm

Agent & ML Systems: deep-research agents, HGM, LangGraph, MCP, LangSmith, LLM-as-a-judge evaluation, PyTorch, NVIDIA Triton

Cloud & Applications: Azure, AWS, Kubernetes, Docker/Podman, Next.js, Electron, FastAPI, PostgreSQL