

Parth Gupta

Columbus, OH | +1-380-250-7402 | haeparth@gmail.com | [LinkedIn](#) | [GitHub](#) | [Portfolio](#)

Professional Summary

AI/Software Engineer with experience delivering end-to-end AI systems and production software, including computer vision pipelines and LLM/RAG applications. Delivered 4.5× faster GPU inference, shipped award-winning React–Python/Flask–PostgreSQL platforms on Cloud, and led production development of transformer-based computer vision models and LLM/RAG chatbots.

Education

Ohio State University

Master of Science, Computer Science

Aug 2024 - May 2026

Medicaps University

Bachelor of Technology, Computer Science and Engineering

2019 - 2023

Experience

The Ohio State University

Graduate Research Associate (Software Engineering & AI Systems), NOWLAB

Aug 2024 - May 2026

Columbus, OH

- Built and maintained Harvest, an NSF-funded drone imagery analysis platform, recognized as “**Most Innovative Software**” at Ohio’s Farm Science Review 2025.
- Developed the **Harvest web app** (Python/Flask + React + PostgreSQL) for job submission, GPU job monitoring, result visualization, and secure downloads, reducing data-access time for growers and researchers while **cutting shapefile(output) size by 60%** and **map rendering time by about 40%**.
- Designed and shipped a **vision transformer–based** computer vision stack in PyTorch that converts raw drone field imagery into structured JSON detections, improving crop/weed detection accuracy and simplifying downstream analytics tools.
- Developed production RAG systems with Chroma/Pinecone and fine-tuned open-source LLMs (Mistral, Zephyr) **via LoRA/QLoRA**, improving retrieval precision by 28% and factual consistency by 45% on internal 1000+ Documents.
- Profiled and optimized GPU inference optimization** using CUDA profiling, batch-size tuning, pinned memory, async data transfers, transfer–compute overlap, and GPU-side preprocessing, **cutting per-image runtime from 0.86 s to 0.19 s** and **enabling near real-time field-scale analysis**.

APenergy

Volunteer Software Engineer

Jan 2025 - Mar 2026

Columbus, OH

- Built an internal dashboard (React, PostgreSQL) to track project status, timelines, and blockers; improved visibility into bottlenecks for faster decisions.
- Designed and shipped Excel-based lead tracking with a lightweight website (Django, PostgreSQL) that automates follow-ups and organizes leads.
- Developed a RAG chatbot (LangChain, OpenAI API) that answered website visitor questions using company docs and policies, handling **~30% of inbound visitors**.

Webonex

Software Engineer

Jun 2023 - Oct 2023

- Implemented backend REST APIs using Java and Spring Boot, including OAuth 2.0 authentication, enabling secure user access the platform.
- Optimized SQL query and database indexing, reducing average response time by **~30–40ms** across high-frequency endpoints, improving throughput.

Indian Institute of Technology

Computer Vision Engineer (Research Intern)

Feb 2023 - May 2023

- Built Computer Vision pipelines for high-resolution satellite imagery preprocessing and feature extraction using Python, OpenCV, and GDAL, improving classification and segmentation **accuracy by 18%** and reducing processing **time by 0.23 sec/tile** by implementing efficient **data structures and algorithms**.

Publications

- Gupta P., Gopinath H. B., Patil V., Dean C., Potlapally A., Subramoni H., Shearer S., Venkatesh R., Fulton J., and Stewart C..Auto-Harvest: An End-to-End Automated Cyberinfrastructure for Precision Agriculture Workflows..*ASABE Annual International Meeting, 2026, Paper No. 2600653*.
- Gupta, P..UAV-Based Weed Detection and Spot-Spraying Prescription Generation: A Scalable Computer Vision and Geospatial Framework.*Master’s thesis, The Ohio State University. OhioLINK Electronic Theses and Dissertations Center. Approved and accepted by The Ohio State University; in process of publication.*

Projects

Monetizer.ai <https://github.com/parthgupta9999/Monetizer.ai>

- A public demo web app where users chat with an AI that answers **in a creator’s voice**, grounded strictly in that creator’s own transcript/caption data (not generic web knowledge), with optional affiliate-link responses and optional voice UX.
- Implemented the RAG layer using **sentence-transformers (all-MiniLM-L6-v2) embeddings** over creator transcripts stored in **Supabase**, with RPC-based vector retrieval when available, cosine-similarity fallbacks, and relevance thresholding to avoid low-confidence/hallucinated answers.
- Delivered the **full-stack + APIs** with a Python/Flask **backend** and **HTML/CSS/JavaScript** frontend; exposed REST endpoints.

Go Rent github.com/parthgupta9999/Vehicle

- Developed a secure vehicle-rental backend with user data recording and OTP verification via Twilio using **Java and Spring Boot**.
- Implemented secure user flows and validations; recognized as the **Most Innovative Project of the University**.

Potato Disease Classification App github.com/parthgupta9999/Potato-disease

- Built and deployed a web app serving a CNN model to classify potato-leaf diseases into three categories with **90% accuracy**.
- Reduced false negatives by 23%** through preprocessing improvements (resizing, augmentation, normalization), improving reliability under field conditions.

Skills and Certifications

- Languages:** Java, Python, SQL, C, C++, HTML, CSS, TypeScript
- ML/AI:** PyTorch, OpenCV, Transformers / ViT, RAG, LangChain, sentence-transformers, Embeddings / Vector search, NumPy
- Backend:** Spring Boot, REST APIs, OAuth 2.0, Hibernate, Servlets/JSP, Flask, Django
- Frontend:** React, HTML5, CSS3
- Databases:** PostgreSQL, SQL, Supabase, Pinecone
- Tools/Systems:** Docker, Git, Linux, Tapis, HPC (OSC), npm, CI/CD, Code Review, AWS, Azure
- Certifications:** AWS Cloud Foundations, Building Cloud Services with Java Spring Framework (Coursera)