Ayush Madhav Kumar

Ann Arbor, Michigan

ayushmadhav.com | contactayushmadhav@gmail.com | linkedin.com/in/ayushmk | github.com/ayushmk7

Education

University of Michigan, College of Engineering **Aug 2025 – May 2028**

Bachelor of Science in Computer Science; Minor in Business and Mathematics *Ann Arbor, Michigan*

Experience

Cactus (YC S25) **January 2026 – Present**

Core Contributor *Ann Arbor, Michigan*

- Contributing to an open source mobile AI inference engine optimized for smartphones and low power devices
- Implementing ARM specific SIMD kernels for matrix multiplication and attention mechanisms with KV cache quantization
- Developing zero copy computation graphs for efficient on device transformer inference with sub 50ms time to first token
- Building OpenAI compatible FFI APIs for cross platform SDK integration across Flutter, React Native, and Kotlin

NASA CLAWS **September 2025 – Present**

Artificial Intelligence Engineer *Ann Arbor, Michigan*

- Engineering, a dual inference AI assistant for NASA SUITS and RASC AL challenges, integrating lightweight transformers (TinyBERT) on Jetson Orin Nano via ONNX Runtime for sub 350 ms inference, including troubleshooting deployments
- Designing an edge cloud architecture that routes inference based on query complexity, latency, and network state
- Implementing DSPy based orchestration and Pydantic validated NLP pipelines within Unity Inference Engine
- Integrating WebSockets and Whisper based preprocessing to synchronize telemetry in high noise environments

Patch **June 2025 – August 2025**

2025 Cohort Founder *Dublin, Ireland*

- Built a startup in 7 weeks as part of the Patch Summer Accelerator, leading end to end technical full stack development
- Offsited at Stripe Dublin and Developed the mobile frontend using React Native, optimizing UI responsiveness and UX
- Implemented a scalable backend with Django and PostgreSQL to manage health data, authentication, and notifications

GIIT Solutions **June 2024 – July 2024, April 2025**

Associate Software Intern *Remote*

- Maintained CI/CD pipelines on AWS, automating deployments across 12+ microservices and improving reliability by 28%
- Developed Python scripts to streamline backend processes, reducing manual overhead by 40% and saving 15+ hours weekly
- Collaborated with senior devs to optimize cloud infrastructure, achieving 35% cost reduction through resource management
- Deployed cloud native solutions with AWS CI/CD tools, improving system scalability and achieving 99.7% uptime

Projects/Contributions

Machine Learning Model Serving Microservice | *Python, Flask, RabbitMQ, MongoDB, Docker*

- Developed a backend microservice to host ML models through Flask REST APIs, for model loading, inference, health checks
- Handled inference requests using RabbitMQ for scalable task queuing, enabling reliable processing of high-volume workloads
- Persisted model outputs and user queries in MongoDB for analytics and logging, and containerized the system using Docker

NeRF Based 3D Modelling | *Neural Radiance Fields, Machine Learning, PyTorch, Computer Vision*

- Built an end-to-end Neural Radiance Fields (NeRF) pipeline for photorealistic 3D reconstruction from multi-view images.
- Implemented volumetric ray marching, positional encoding, and differentiable density/color prediction for view synthesis.
- Designed a modular training pipeline with configurable hyperparameters to support multiple 3D reconstruction datasets.

Technical Skills

- **Programming Languages:** Python, JavaScript, TypeScript, Java, C, C++, C sharp, Ruby, PHP, SQL, HTML, CSS
- **AI and ML:** PyTorch, TensorFlow, JAX, HuggingFace Transformers, ONNX Runtime, Scikit learn, OpenAI APIs, LLMs, NLP, Computer Vision, Reinforcement Learning, Edge AI deployment, Model optimization
- **Software Development:** React, Next.js, React Native, Node.js, Django, Flask, FastAPI, REST, GraphQL, Flutter, Dart, WebSockets, Progressive Web Apps, Vue, Git, GitHub, GitHub Actions, GitLab, APIs
- **Databases and Data Engineering:** MongoDB, PostgreSQL, Redis, MySQL, Data modeling, ETL pipelines
- **DevOps and Cloud Infrastructure:** Docker, Kubernetes, CI CD pipelines, RabbitMQ, Linux, AWS EC2, AWS Lambda, AWS S3, Google Cloud, Edge deployment, Serverless architecture, Scalable ML inference orchestration
- **Cybersecurity:** Nmap automation, PCAP forensics, TLS protocol analysis, API and web vulnerability analysis, reverse engineering, YARA based malware triage, cryptographic weakness detection, AWS IAM and S3 security
- **Core Computer Science:** Data Structures, Algorithms, Time and Space Complexity Analysis, Object Oriented Design, Memory Management, Concurrency and Multithreading, Operating Systems Basics