

MANNAN SHUKLA

Tri-State Area / Bay Area • US Citizen

(609) 575-0436 • mannanshukla@icloud.com • linkedin.com/in/mannanshukla • github.com/mannanshukla

TECHNICAL SKILLS

CUDA, Slurm, DCGM, nvidia-smi, NCCL • Kubernetes, Docker, Helm, ArgoCD, Terraform • Go, Python, Bash, JavaScript, TypeScript • AWS (EC2, ECS, RDS, S3, DynamoDB, CloudWatch, IAM, SNS) • Prometheus, Grafana • PyTorch, Hugging Face, LangChain • Auth0, JWT, OAuth, RBAC, GitHub Actions

EXPERIENCE

Software Engineer

July 2025 - Present

Voltage Park / Lightning AI

Redmond, WA

- Collaborated with senior engineers on architecture and design of the **VP Cloud** platform; contributed to **DynamoDB** data model design for fleet and node management with composite keys, org-scoped GSIs, and event-sourced history tables
- Designed and implemented a **P2C** load balancer for the storage control plane, improving request distribution across backend instances — involved in both the system design and production rollout
- Built the **VP Cloud Python SDK** end-to-end: **OpenAPI** code generation pipeline, **Auth0** M2M authentication with datacenter-aware token exchange, and automated CI/CD publish workflows via **GitHub Actions**
- Integrated **Auth0** into the VP Cloud auth flow with the internal **IAM** service, implementing JWT validation, RBAC, and org-scoped multi-tenant access control
- Served as on-call engineer for production **H100** clusters; debugged **WireGuard** VPN, **CoreDNS**, **MPI Operator**, and **GPU Operator** issues across **Kubernetes** environments; set up **Prometheus** and **Grafana** monitoring for fleet observability
- Built **Slurm** prolog/epilog GPU health check suite (ECC, PCIe AER, NCCL stalls, tiered thermal protection) with automatic node draining across client clusters including Reflection AI

Independent Researcher — SuperAgent

Apr 2024 - Apr 2025

Rutgers University

New Brunswick, NJ

- Built **LLM**-based agent recommendation system under Prof. Yongfeng Zhang; achieved 83% accuracy vs. **GPT-4** ground truth using a 1B model via **LoRA** fine-tuning with **Unsloth**

Partnerships + Business Development

Feb 2025 - May 2025

Nas.io

New York, NY

- Drove influencer partnerships and client onboarding for a Series A startup, generating \$2K GMV in one month while 50x'ing outreach efficiency via a **Python/Klaviyo** automation pipeline

PROJECTS

AdBlockIRL — AR ad-blocking proof-of-concept for Meta Quest

May 2025

- Trained **YOLOv8** on **Roboflow** to detect 100+ concurrent ads at 5ms latency, applying real-time mosaic censoring via **OpenCV**; won \$2000 at Bain Capital Ventures AI Hackathon

SlopGen — Automated short-form video generator

June 2024 - Sep 2024

- Built end-to-end pipeline using **yt-dlp**, **WhisperX**, and **llama3** via **ollama** achieving 92% segmentation accuracy (15x faster than manual), with custom TTS voices via **edge-tts**

EDUCATION

Rutgers University - New Brunswick

Bachelor of Science - Computer Science • Cum Laude

- Linux User Group (President), Road to Silicon Valley Program