

Arnav Thakur

arnav7t.tech | [GitHub](#) | [LinkedIn](#) | arnav19thakur@gmail.com | +91 7742207123

Software Engineer experienced in building production backend systems with a strong focus on **AI infrastructure**, **Kubernetes**, distributed systems, and secure deployments. Comfortable working across backend engineering, platform infrastructure, and applied AI. Currently building an enterprise **air-gapped AI platform** deployed in customer-managed environments on RKE2, with hands-on experience owning the reliability, upgradeability, and operability of production services.

SKILLS

- **Languages:** Python, Go, SQL, Bash
- **Backend Engineering:** REST APIs, FastAPI, Microservices, Distributed Systems, API Design
- **AI Systems:** LLMs, RAG, AI Agents, MCP, ReAct, Tool Calling, Prompt Engineering, Vector Search
- **Cloud, Platform & DevOps:** AWS, Docker, Kubernetes (RKE2, Minikube), Helm, Terraform, Linux
- **CI/CD:** GitHub Actions, Jenkins, Docker Compose, Artifactory
- **Databases:** PostgreSQL, MongoDB Atlas, Qdrant, FAISS, ChromaDB
- **Security:** TLS/mTLS, X.509, Air-Gapped Deployments, FedRAMP, FIPS

EXPERIENCE

Trellix

Jul. 2025 – Present

Software Engineer Apprentice - BasePlatform Team

- **Air-Gapped AI Platform (Trella):** Architected and developed Trella, an enterprise AI platform deployed directly inside customer-managed, air-gapped environments. Built three core microservices with comprehensive **unit** and **Helm integration tests**, delivering a Helm-based deployment on single-node RKE2 with **zero runtime network dependency** and CPU-only inference over customer documentation.
- **AI Infrastructure:** Built the FastAPI orchestration layer for service discovery, request routing, and inference coordination across a six-service AI platform. Built a scikit-learn intent classifier that bypassed the LLM for **40–50%** of conversational turns, reducing inference latency and compute cost, and engineered an **mtime**-based hot-reload mechanism enabling zero-downtime classifier updates across Kubernetes pods.
- **Platform & Security:** Led the proof-of-concept for a **multi-node RKE2** high-availability deployment, implementing and evaluating upgrade and rollback workflows for platform adoption. Contributed to Linux networking, **iptables**-based pod communication, and **FedRAMP**-aligned TLS/X.509 certificate validation for secure inter-service authentication.
- **Developer Tooling:** Built the Windows PowerShell automation client for **SupportRelay**, automating secure connectivity into enterprise appliances over an existing **AWS EKS/Kubernetes** and OpenVPN platform, reducing support turnaround by **60%**.

Coulomb AI (YC S21)

Aug. 2024 – Nov. 2024

Data Analyst Intern

Early-stage YC-backed startup

- Built automated data validation pipelines with PySpark to support scalable ML preprocessing.
- Applied time-series analysis and anomaly detection with STUMPY on EV fleet data.

PROJECTS

k8s-agent | *Kubernetes, ReAct, Prometheus, FastAPI, Docker, Helm, MCP*

- Built a Kubernetes-native AI operations agent that autonomously investigates production incidents using ReAct reasoning, MCP, and tool calling to gather evidence, execute diagnostics, and produce root-cause analyses across Kubernetes, Prometheus, and Grafana Loki.
- Designed a **CNAPP detection engine** implementing graph-based analysis and policy rules to identify cloud and Kubernetes misconfigurations, privilege escalation paths, and AI workload security risks.
- Packaged the platform with FastAPI, Docker, and Helm, exposing agent capabilities through MCP with secure multi-tenant authentication for production deployments.
- Implemented an LLM-as-Judge evaluation framework with regression benchmarking to validate root-cause hypotheses and improve the reliability and observability of automated incident investigations.

ReuseCV | *Go, Next.js, AWS, Terraform, GitHub Actions, Docker, Gemini*

- Built an AI-powered **Career-as-Code** platform that transforms resumes and project documentation into a reusable knowledge base, enabling tailored resumes without repeatedly editing PDFs.
- Developed a production-ready Go backend exposing REST APIs for AI-powered extraction, semantic resume tailoring, schema validation, and automated LaTeX document compilation.
- Provisioned cloud infrastructure on **AWS EC2** and **Amazon S3** using Terraform, enabling scalable document generation, artifact storage, and secure delivery through pre-signed URLs.
- Built CI/CD pipelines using GitHub Actions for cross-platform Go releases, Docker image publishing to GHCR, and automated deployments to AWS.

EDUCATION

Amrita Vishwa Vidyapeetham

Bachelor of Technology in Computer Science Engineering

Bangalore, Karnataka

Sep. 2021 – May 2025