

Ansh Singhal

AI/ML Engineer — LLM Infrastructure — Backend & Distributed Systems

+91-8929554991 | anshsinghal3107@gmail.com | linkedin.com/in/anshshh-singhal | github.com/AnshSinghal | anshsinghal.tech

EXPERIENCE

AI/ML Engineer

Feb 2026 – Present

CyberUltron Consulting Pvt. Ltd.

Remote

- Architected and shipped an AI security gateway end to end, owning backend, frontend, DevOps, cloud deployment, and scaling; secured LLM, RAG, vector DB, and MCP workflows across 500+ endpoints with authentication, PII detection, prompt-injection mitigation, policy enforcement, and observability.
- Scaled the gateway to 100k+ RPS and reduced unsafe LLM behavior by 99%; built centralized pre/post-inference controls and routing for production AI workloads.
- Built AI-powered vulnerability remediation and CBOM, AIBOM and SBOM workflows, cutting manual triage time by 60%+ and improving platform performance by 40%; also deployed endpoint AI governance protecting 1,000+ enterprise endpoints across 5+ AI tools.

Open Source Engineer

Jan 2025 – Present

IntelOwl Threat Intelligence Platform

Remote

- Merged 13 pull requests across analyzers, integrations, backend reliability, and data integrity in a production Python/Django threat-intelligence platform.
- Added and maintained 5 analyzer integrations spanning DNS, malware, phishing, and OSINT workflows, expanding threat-detection coverage by 23%.
- Resolved CAPA caching, Elasticsearch, API compatibility, error-handling, and data-atomicity issues; added pytest validation and improved migration reliability to 99.9%.

PROJECTS

ZeroDayShield – OSS Zero-Day Discovery | *FastAPI, Next.js, PostgreSQL, Docker*

[Project Link](#)

- Built an OSS zero-day discovery engine spanning 4 ecosystems: npm, MCP servers, Docker images, and Hugging Face models.
- Designed multi-stage static-analysis + LLM triage achieving **100% recall and 93% precision** on the OSSF malicious-packages corpus at a **15x cost reduction**; engineered **50+ custom analyzers** mining **49,522 signals** across **228,020** known-malicious packages.

DHARA – Multi-Agent Legal Research Engine | *FastAPI, LangGraph, Vector Search, AWS*

[Project Link](#)

- Built a multi-agent RAG system over **10,000+ legal documents** achieving **97% retrieval accuracy** with citation-traced answers.
- Implemented query decomposition, reranking, deterministic grounding, and vector retrieval on AWS containerized microservices, achieving **sub-500ms P95 latency**.

Realtime VoiceOps AI | *Kafka, WebSockets, Kubernetes*

[GitHub](#)

- Built a real-time voice streaming backend with **sub-200ms latency** for **100+ concurrent sessions** using WebSockets and Kafka.
- Scaled processing with Kafka consumer groups and Kubernetes HPA across PostgreSQL, MongoDB, and Redis; instrumented consumer lag and P95 latency for distributed performance monitoring.

Flood-GAN – Physics-Aware Flood Mapping | *PyTorch Lightning, Computer Vision*

[GitHub](#)

- Developed a physics-informed GAN for SAR-to-optical image translation, achieving **PSNR 31.25** and **SSIM 0.94**.
- Designed a dual-head U-Net with NDWI supervision for flood-boundary detection and stabilized training with EMA updates and speckle-preservation loss.

TECHNICAL SKILLS

Languages: Python, TypeScript, JavaScript, C++, Java, SQL

AI & Agent Systems: LLMs, Agentic RAG, LangGraph, LangChain, LlamaIndex, MCP, Tool Calling, Embeddings, Vector Search, LLM Evaluation, Hugging Face, PyTorch, TensorFlow

Backend: FastAPI, Django, DRF, Flask, Node.js, Express.js, Celery, REST APIs, JWT

Frontend: React, Next.js, Vite, HTML/CSS

AI Security: Presidio, llm-guard, OPA, Envoy, Prompt Injection Mitigation, PII Detection, OWASP LLM, AI Governance

Infrastructure: Docker, Kubernetes, Kafka, RabbitMQ, Nginx, CI/CD, Prometheus, Grafana

Cloud: AWS, GCP, Azure

Data: PostgreSQL, MongoDB, Redis, Pinecone, Chroma, Milvus

EDUCATION

Bennett University, The Times Group

B.Tech in Computer Science (AI Specialization), CGPA: 9.47

Greater Noida, India

Aug. 2023 – May 2027

LEADERSHIP & IMPACT

- **Rotary International Youth Advisory Council** Sole Asia representative, global youth initiatives
- **Co-Organizer, GDG Bennett University** Led 12+ events, 4,000+ participants, 40% community growth
- **Joint Secretary, Entrepreneurship Cell** Supported 15+ student startup launches and cross-functional initiatives