Yash Jangid

+91 9079333067 | realyashjangid@gmail.com | yashjangid.dev | LinkedIn | GitHub

TECHNICAL SKILLS

Languages: C++, Python, TypeScript, SQL, HLSL, C#, C, JavaScript, HTML/CSS

Distributed Systems & Infrastructure: Docker, Kubernetes, AWS, Azure, Nginx, Kafka, WebRTC, WebSockets, gRPC

Machine Learning & LLMOps: PyTorch, Transformers, LangChain, Llama.cpp, MCP, RAG, ONNX, OpenCV, FastAPI

Databases & Vector Search: PostgreSQL, Redis, MongoDB, MySQL, Neon, Supabase, Firebase, Pinecone, Typesense, BullMQ

Tools & Frameworks: Node.js, Next.js, Poetry, Linux, Git/GitHub, CI/CD, Shell Scripting, Databricks

Core CS & Systems: Data Structures & Algorithms, System Design, Multi-threading, Memory Management, Distributed Systems, OOPs

PROJECTS & ENGINEERING EXPERIENCE

AirDelivery.site | *WebRTC, WebSocket, Next.js, Cloudflare*   June 2025 – Present
- Architected a P2P file delivery system scaling to **1.2M API requests** with 99.9% availability across global regions.
- Engineered a **zero-copy streaming disk I/O** mechanism to bypass RAM bottlenecks, enabling 200+ Mbps transfers.
- Optimized NAT traversal using custom signaling for **30,000+ users** with zero server bandwidth cost.

Cobbic - Microservices Backend | *Next.js, Redis, Typesense, Docker, BullMQ*   July 2025 – Present
- Developed a microservices social backend achieving **O(1) read access** via a Redis fanout-on-write strategy.
- Built a typo-tolerant search engine using Typesense and semantic ranking via **ONNX vector embeddings**.
- Containerized full-stack deployment via Docker/Nginx on AWS, handling 300+ users in the first week of production.

ProtoPulse-Core | *C++, TCP Sockets, FFmpeg, Multithreading*   Mar. 2025
- Engineered a high-performance C++ streaming protocol with **90% bandwidth reduction** via FFmpeg (H.264).
- Implemented lock-free synchronization and manual memory management for concurrent video/audio/chat streams.

- Optimized packet ordering and raw byte-stream handling at the socket level to minimize transport-layer latency.

Real-Time AI Speech Translation | *FastAPI, Whisper, PyTorch, CUDA, WebSockets*   Nov. 2025
- Optimized **CUDA inference pipelines** for Faster-Whisper, achieving sub-150ms latency for real-time speech processing.
- Integrated neural synthesis with FastAPI WebSockets to support seamless cross-language video interactions.
- Architected the system for low-latency streaming inputs, prioritizing long-context memory retention.

3D Fluid Simulation | *Unity, SPH, HLSL, GPGPU*   Jan. 2025
- Simulated 16K+ particles at **24M operations/sec** using Smoothed Particle Hydrodynamics and Compute Shaders.
- Eliminated CPU overhead by parallelizing physics logic across GPU cores, maintaining a stable 60 FPS.

ACHIEVEMENTS & HONORS

- Awarded both **Best Full Stack Developer** and **Best Coder** distinctions by CSE AIML (2025).
- Certified <u>AWS Cloud Practitioner</u> , <u>Neural Networks (Andrew Ng)</u> , <u>AI for Everyone</u> certifications.
- Scaled **AirDelivery** to **1.2M global requests** across US, India, and Germany within 6 months of launch.

EDUCATION

KIET Group of Institutions   Ghaziabad, UP
B.Tech in CS & Engineering with AI-ML (GPA: 8.8/10.0)   *Sep. 2024 – June 2028*