Kumar Harsh

kharsh560@gmail.com | +91 9771140694 | Delhi | Linkedin | Github | Portfolio

A results-driven software engineer with a strong foundation in **Data Structures and Algorithms (C++)**, and **System Design**, with hands-on experience in **MERN-based full-stack development**. Skilled in integrating **AI tools into utility apps for business use cases.** Demonstrates adaptability and ownership by architecting and implementing end-to-end systems in **fast-paced startup environments**.

Experience

SDE Intern @ **Brivio** | [April 2025 – June 2025]

- Engineered a scalable, low-latency, **AI-powered** interview agent, replacing **Vapi** with **LiveKit** to reduce monthly infra costs by **89%**:
  - Built a low-latency SFU-based media pipeline, automated EC2-hosted transcoding, and S3 storage of individual tracks.
  - Orchestrated **STT-LLM-TTS** flow with Deepgram, OpenAI and 11-Labs; optimized voice activity detection with Silero VAD.
- Developed robust, responsive full-stack features in **React.js/Next.js** using **ShadCN** components:
  - Integrated code editor with AI agent, built feedback-system endpoints and responsive UIs with pagination and client-side filtering.
  - Automated the upload of interview recordings from Vapi's S3 to Brivio's S3 as a backup via a **Node.js microservice**.

Full Stack Developer & LLM Engineer Intern @ **Stealth Start-up** | [September 2024 – March 2025]

- Kubernetes & AKS Automation:
  - Experimented with and automated the deployment and interconnection of diverse microservices using **Kubernetes on Docker**, leveraging **Helm** charts, later transitioning to **Azure AKS**, to complete the full deployment automation pipeline.
- Multi-LLM Recommendation Engine:
  - Trained **chatbot** for requirement gathering, integrated a **multi-LLM model**, to generate recommendations for graph, paired with a scoring engine that evaluates LLMs based on custom parameters to identify the top-performing model.
- Web Application Development:
  - Built a Next.js/Express app with Redux state management, **ReactFlow** interactive workflows, custom modals and loaders.
  - Implemented NextAuth with Prisma ORM in Next.js, separate Express backend, and **Socket.io** (for real-time deployment status).

Sync Sphere | [GitHub Repo :- Click Here]

- Social-impact MERN based full-stack platform **integrating Gemini LLM** for personalized tax-benefit guidance.
- Recognized by GDSC India; top-10 finalist at NIT Patna Pitchtember; winner at BPIT Malhaar Ideathon;.

Projects

1) **Content Sharing & Interaction Platform** | [Deployed on VM of DO using Docker + Nginx :- Click Here] | [GitHub Repo :- Click Here]

- Developed a full-stack MERN blog app for my AI & ML department demonstrating robust features and **AI integration**: **TinyMCE** rich-text editor, Redux Toolkit, **custom authentication** (using enrollment number), **session persistence**, **role-based access control**, edit-blog functionality and **Gemini AI integration** to auto-generate summaries, tags, and read-time estimates for the uploaded blogs.
- Architected a secure backend with **JWT** access/refresh tokens, MongoDB aggregation pipelines, and Cloudinary/Multer image uploads.
- Containezed the backend by **configuring Nginx** as a web server, reverse proxy, SSL terminator, and CORS handler with a custom domain (mait-aimldept- blogapp.live), and deployed on a self-hosted Digital Ocean's VM using Docker Compose.

2) **Chat App** | [GitHub Repo :- Click Here]

- Developed a full-stack chat application using Next.js. Integrated **Prisma ORM** for user registration, and **NextAuth** for session management. Architectured a robust server capable of managing both **HTTP** and **WebSocket** requests, enabling real-time messaging.
- Key achievement: Implemented a **custom authentication** flow for secure WebSocket connections: a short-lived access token, obtained via an initial HTTP request, validated the WebSocket connection. This addressed authentication challenges with the 'WS' library.

3) **Scalable Video Streaming Platform Backend** | [Github Repo :- Click Here]

- Engineered a video processing pipeline with **FFmpeg** for multi-resolution streaming and automated thumbnails.
- Implemented Watch History + Continue Watching to enable personalized playback experiences.
- Optimized a **scalable** upload system capable of handling sudden traffic spikes efficiently.
- Developed a Creator Analytics Dashboard with insights on views, watch time, and engagement metrics.

Education

Bachelor of Technology in Artificial Intelligence and Machine Learning — Maharaja Agrasen Institute of Technology, Delhi

SGPA (3rd–6th): 8.85, 8.80, 8.50, 8.96 • CGPA: 8.53 — Nov 2022 – Jul 2026 (Expected)

Skills

Data Structures and Algorithms (Leetcode Link) | React.js | C/C++ | Next.js | JavaScript | TypeScript | AI-tools | Agentic workflow | Backend (Node.js, Express.js, Mongoose) | Tailwind | Redux | Cloud Infrastructure (Docker, Kubernetes, Azure AKS) | Appwrite | Netlify | Github | Git

Co-curricular activities

Technical Executive | Computer Society of India, MAIT chapter (CSI-IW) | [July 2024 – Present]

- Contributed to backend of CSI's mobile app using Prisma, PostgreSQL, Node.js, and Express. Implemented authentication, CRUD, and key features including voting mechanism, event & venue listings, and fest update notifications. Explored communication protocols like Server-Sent Events and Long Polling to optimize voting efficiency. Also contributed in CSI's website using Next.Js, Aceternity UI & Tailwind.
- Selected as one of 11 mentors to evaluate the first three rounds of the "Code Genesis" hackathon under "Uncharted", culminating in a finale at Microsoft Gurugram. Shortlisted 250+ teams, narrowing them down to 16 finalists. Additionally, worked with a team of 50 executives to organize and execute "Zypher", another flagship techno-cultural fest by CSI, successfully managing a footfall of over 1000 attendees.