

Shubham Kumar Prajapati

+91 7080364716 | jackwill021@gmail.com | linkedin.com/in/shubham7080 | github.com/playkashyap
[Portfolio](#) | [macOS Design](#) | [Windows Design](#) | [LUNA Landing Page](#)

EXPERIENCE

Fullstack Developer

Aug. 2025 – Present

Mostedge

Vadodra, India

- Building a remote surveillance platform for C-stores with real-time face recognition, GPU-accelerated video processing, and live streaming via HLS and WebRTC exposed through Nginx.
- Developed RESTful APIs with Node.js (TypeScript) and Python (FastAPI) backed by MSSQL and PostgreSQL; containerized services using Docker and deployed on AWS GPU cloud instances.
- Actively leveraged AI coding tools (GitHub Copilot, Claude Code, Gemini Code Assist) for accelerated development, code review, debugging, and automated test generation.
- Used CUDA, GStreamer, and FFmpeg for GPU-accelerated video encoding, decoding, and real-time stream processing with low-latency HLS and WebRTC playback.

Software Developer (Frontend and Backend)

Jun. 2023 – May. 2025

Cosmotech AI

Noida, India

- Built responsive UIs using React, Angular, Next.js, HTML, CSS, and TypeScript.
- Developed RESTful APIs with Node.js (TypeScript) and MongoDB for real-time data operations.
- Integrated APIs to enhance interactivity and dynamic content.
- Used Git for version control; participated in agile ceremonies including sprint planning and daily standups.

Frontend Developer Intern

Jan. 2023 – Jun. 2023

Cosmotech AI

Noida, India

- Contributed to UI development using React, Angular, HTML, CSS, and TypeScript.
- Assisted in API integration and dynamic content updates.
- Participated in agile processes and collaborated using Git for version control.
- Participated in daily stand-ups and team meetings, actively engaging in sprint planning with project goals and timelines.

PROJECTS WITH COMPANY

WatchGuard | *MediaMTX, GStreamer, Kafka, Wasabi S3, Node.js, Python, InsightFace*

Aug. 2025 – Present

- Designed a microservice-based surveillance architecture from scratch, with services scaling independently per camera and per AI worker; currently processing $2 \times 1080p$ camera feeds at 20fps with AI inference at 20fps (frame-skip of 2) per stream.
- Built a real-time video pipeline using GStreamer and FFmpeg with CUDA acceleration for GPU-optimized encoding/decoding; streamed via HLS and WebRTC exposed through Nginx reverse proxy.
- Implemented an event-driven architecture using Kafka and AWS MSK to handle high-throughput video event logs, face recognition alerts, and metadata across distributed services.
- Integrated InsightFace for GPU-accelerated face detection and recognition; containerized all services with Docker and hosted on AWS GPU cloud instances for scalable deployment.

HRIS - Front-end | *Next.js, React, Tailwind, Zustand, Context API, Typescript*

Aug. 2024 – May. 2025

- Developed a pixel-perfect, responsive UI with a custom sidebar using Tailwind and reusable React components.
- Built a dynamic drag-and-drop dashboard for customizable user workflows and widget arrangements.
- Managed global state efficiently using Zustand and Context API across multiple modules.
- Created custom dynamic organizational chart which can be exported to a pdf.

PMS - Front-end | *Angular, RxJs, Bootstrap, Apex Charts, NgRx, Typescript*

July. 2023 – May. 2025

- Built and maintained the frontend of a Project Management System using Angular with NgRx for predictable state management across complex workflows.
- Developed interactive data visualizations and reporting dashboards using Apex Charts with real-time data binding via RxJs observables.
- Created reusable Angular components and modules with a focus on performance, lazy loading, and maintainability across the application.

Rate System - Front-end | *React, Vite, Bootstrap, Apex Charts, Context API, Javascript*

Jan. 2023 – May. 2025

PROJECTS

- LUNA** | *Python, Whisper, IndexTTS2, InsightFace, Llama.cpp, LM Studio, Ollama* 2026 – Present
- Built a full-stack emotion-aware AI assistant that integrates speech, vision, and language capabilities into a unified real-time pipeline.
 - Implemented Speech-to-Text using OpenAI Whisper and Text-to-Speech using IndexTTS2 for natural, expressive voice interaction.
 - Integrated InsightFace for real-time facial recognition and emotion detection, enabling context-aware, personalized responses based on the user's emotional state.
 - Connected local LLM backends (Llama.cpp, LM Studio, Ollama) for fully offline LLM inference, ensuring privacy and low-latency responses without cloud dependency.
- HandFlow** | *Python, OpenCV, Mediapipe, CNN, TensorFlow, NumPy* Aug. 2022 – Apr. 2023
- A proposed Framework for Controlling Basic Computer Functionalities with Hand Gesture
 - Created a real-time video program with 96% accuracy for static gestures and 97% accuracy for dynamic gestures, employing a robust process encompassing data collection, model training, and real-time implementation.
 - Executed diverse computer functions based on recognized gestures, including actions like zooming, saving, and minimizing open windows.
- Image Processing based Brain Tumor Detection** | *Python, SVM* Aug. 2021 – Jan. 2022
- Applied advanced algorithms, including pre-processing, segmentation, and SVM classification, to enhance accuracy in brain tumor detection
 - Published a paper titled "Image Processing based Brain Tumor Detection" at ICFIRTP-2022.

TECHNICAL SKILLS

Languages: JavaScript, HTML/CSS, Typescript, Python, MongoDB, GraphQL, SQL, MSSQL, PostgreSQL
Frameworks: ReactJs, Node.js, Express.js, FastAPI, Tailwind, Bootstrap, Material-UI, Angular, NextJs
AI Tools: GitHub Copilot, OpenAI Codex, Claude Code, Gemini Code Assist, Ollama, LM Studio, Llama.cpp
AI/ML: InsightFace, OpenAI Whisper, IndexTTS2, Llama.cpp, Local/On-Device LLM Integration, CUDA
LLM & Agentic AI: Prompt Engineering, RAG, Function Calling, Structured Outputs, Embeddings, Vector Search, Agentic Pipelines, Multi-Agent Systems (A2A), MCP (Model Context Protocol), Prompt Caching, KV Cache, Chain-of-Thought, Few-Shot Prompting, LLM Orchestration, Quantization (INT4/INT8/GGUF)
DevOps & Infra: Docker, AWS, Nginx, Git, MS Azure, MongoDB Atlas, Postman
Streaming: GStreamer, FFmpeg, HLS, WebRTC, MediaMTX, RTSP
Libraries: MUI, Rxjs, Axios, Body Parser, bCrypt, JWT, Moment, Angular Material, ThreeJs, ShadCn UI, Tanstack

CERTIFICATIONS

- Leadership and Team Effectiveness** | *NPTEL* | ID: NPTEL23MG28S53610062 Apr. 2023
- Software Testing** | *NPTEL* | ID: NPTEL22CS61S5367040710050680 Oct. 2022
- Python (Basic) Certificate** | *HackerRank* | ID: 18819e72407e Jul. 2022
- Introduction to Industry 4.0 and Industrial IoT** | *NPTEL* | ID: NPTEL21CS66S13510169 Oct. 2021
- Managing Change in Organizations** | *NPTEL* | ID: NPTEL21MG22S11480034 Mar. 2021
- Learn Python Programming Masterclass** | *Udemy* May. 2020

EDUCATION

- Sharda University** Gr Noida, India
Bachelor of Computer Science and Technology Aug. 2019 – Aug. 2023
- Columbus International School** Ballia, India
Higher Secondary, 81% Mar. 2016 – Apr. 2018
- Columbus International School** Ballia, India
Secondary ,10 CGPA Mar. 2014 – Mar. 2016