RISHABH JAIN

AI Engineer | Full-Stack and Mobile Developer | LLM, RAG and Computer Vision

Bhopal, Madhya Pradesh, India | +91 6263327220 | Rishabhjain26012002@gmail.com | linkedin.com/in/rishabhhjain26
github.com/RJ26012002 | Portfolio: rishabhhjain.com

PROFESSIONAL SUMMARY

AI Engineer with 2+ years of experience in machine learning, deep learning and NLP, building and deploying production AI, mobile and web applications. Built **RAG (retrieval-augmented generation)** chatbots, a **real-time multilingual voice agent**, **face recognition** and **OCR** pipelines, and a **400+ page Next.js ERP** for a battery recycling company. Ships end to end, from requirements and system design to deployment on **AWS** and Windows Server, with **9 live products** used by **500+ people daily**. Holds a published Indian patent and a B.Tech Gold Medal in AI; pursuing an M.Tech in AI at MANIT Bhopal.

TECHNICAL SKILLS

- **Languages:** Python, Dart, TypeScript, JavaScript, SQL, Kotlin, PHP, C++
- **AI, ML and LLMs:** Machine Learning, Deep Learning, NLP, Retrieval-Augmented Generation (RAG), AI agents, LangChain, FAISS vector database, BM25 hybrid search, re-ranking, embeddings, prompt engineering, fine-tuning, model evaluation, model deployment, Ollama, llama.cpp, quantisation, Hugging Face, Gemini API, TensorFlow, Keras
- **Speech AI:** Whisper speech-to-text, text-to-speech, Silero VAD, ONNX Runtime, real-time audio streaming, barge-in handling
- **Computer Vision and OCR:** OpenCV, YOLOv8, DeepFace, ArcFace, RetinaFace, MTCNN, YuNet, SFace, EAST, Tesseract, EasyOCR, PaddleOCR, DBSCAN, liveness detection
- **Data Science:** Pandas, NumPy, Statsmodels, SARIMA, ARIMA, time-series forecasting, backtesting, Matplotlib
- **Backend and Databases:** FastAPI, Django, Django REST Framework, Flask, REST APIs, WebSockets, JWT and OTP authentication, RBAC, MySQL, SQLite
- **Frontend and Mobile:** Next.js, React, HTML, CSS, Tailwind CSS, Zustand, Three.js, Flutter, Android (Kotlin platform channels), Riverpod, Hive, WorkManager, Firebase
- **Cloud and Tools:** AWS, Docker, Windows Server, Vercel, Git, GitHub, pytest, Google Play Console, App Store Connect

PROFESSIONAL EXPERIENCE

AI Developer, VirtuonAI Pvt Ltd, Bhopal | Nov 2024 - Present

- Design, build and deploy AI, mobile and web applications end to end for clients in healthcare, recycling, real estate, logistics and e-commerce.
- Delivered the ReBAT ERP platform (HR, procurement, performance management and production planning) in a two-person development team.
- Published VCall to the **Google Play Store** and **Apple App Store**, and distribute Android apps used by **500+ people daily**.
- Deploy and maintain production services on **AWS** and **Windows Server**, with Firebase Crashlytics and Remote Config for release monitoring.

AI Developer Intern, VirtuonAI Pvt Ltd, Bhopal | Jan 2024 - Oct 2024

- Built OCR for medical records and bills: ran a model evaluation of **Tesseract, EasyOCR and PaddleOCR**, with **EAST** text detection to localise text before recognition.
- Fine-tuned a **YOLOv8** deep learning model to detect key fields in health records on a **10,000-image** labelled dataset, with Python tooling to convert JSON annotations to YOLO format.

PROJECTS

Hinglish RAG Chatbot for Mobile Repair Training, Mooxy | Paid freelance | Live, 100+ users
Tech: *Python, FastAPI, LangChain, FAISS, Ollama, llama.cpp, MySQL*

- Built a domain-specific **RAG** assistant that answers hardware repair questions in Hinglish, with **15+ fault-domain indexes** in a FAISS vector database.
- Improved answer precision with **re-ranking**, a fault ontology and pin mapping; served a **quantised local LLM** via llama.cpp and Ollama.
- Built a FastAPI and MySQL admin API to update knowledge and rebuild indexes without code changes.

Real-Time AI Voice Calling Agent (Hindi, Hinglish, English) | Built
Tech: *Python, FastAPI, WebSockets, Gemini API, Groq Whisper, Silero VAD, ONNX Runtime, Edge TTS, BM25, Twilio, pytest*

- Built a streaming pipeline: **VAD**, speech-to-text, **hybrid RAG** (embeddings plus BM25), streaming LLM and sentence-wise TTS, with **barge-in** over Twilio Media Streams and WebSockets.
- Cut speech-to-text latency from about 2.7 s to **0.8-1.5 s** with Groq Whisper, adding hallucination detection with a Gemini fallback.
- Rejected background speech with **Silero VAD** and a caller-loudness gate; prompt guardrails stopped invented prices; **58 pytest tests**.

Pics Picker, AI Event Photo Finder | Live
Tech: *Python, Django, DeepFace, ArcFace, RetinaFace, MTCNN, OpenCV, DBSCAN, NumPy*

- Built a face-search web app: guests scan a QR code, upload a selfie and get every event photo they appear in, matched with **ArcFace embeddings** and **cosine similarity**.
- Raised recall on dark photos with image enhancement, multi-scale detection and **fallback detectors**; grouped faces with **DBSCAN** and cached embeddings for speed.

Face Recognition Attendance System | Built
Tech: *Python, OpenCV, YuNet, SFace, TensorFlow, MySQL*

- Rebuilt a Haar Cascade and LBPH prototype with **YuNet** detection and **SFace embeddings**, adding a rejection threshold for unknown faces.
- Added a head-turn **liveness check** against photo spoofing and 5-frame confirmation; logged attendance in MySQL, with **17 passing unit tests**.

ReBAT ERP: HRMS, Procurement and Production Planning | Live
Tech: *Next.js (App Router), React, TypeScript, Tailwind CSS, Zustand, Axios, jsPDF*

- Built a **400+ page** multi-tenant ERP frontend with **role-based permissions** for a battery recycling company.
- Implemented procurement (indent, quotation, purchase order, GRN, quality check), KRA/KPI reviews with **bell-curve normalisation**, and policy approval workflows.
- Built production planning with BOMs, routings, master scheduling and **multi-level MRP** that handles co-products, by-products and scrap.

VCall and SR Tracker, Call Logging and Sales CRM Apps | Live on Google Play and App Store
Tech: *Flutter, Dart, Kotlin platform channels, Riverpod, Hive, SQLite, WorkManager, Firebase*

- Built a dialer and call logger with **native Kotlin platform channels** and an offline retry queue; moved call-log queries off the UI thread and cut cold start from **29 s to under 6 s**.
- Merged a React Native sales app into the Flutter codebase, shrinking the APK from **114 MB to 20 MB**, with local-first SQLite data and **45+ automated tests**.

CurAyu, AI Health Assistant | Live on Android, iOS and web
Tech: *Flutter, Provider, Hive, Firebase*

- Built one Flutter codebase for AI health chat, medical report analysis and document storage, with Firebase Analytics, Crashlytics, Remote Config and push notifications.

Spark HRMS Field Tracking App | Built for ReBAT
Tech: *Flutter, Android foreground service, Geolocator, SQLite, WebView JavaScript bridge*

- Built duty-hour GPS tracking with a foreground service, 5-minute fused location fixes, a SQLite offline queue and batched uploads, plus a JavaScript bridge into the web HRMS.

Battery Pricing Engine | Built
Tech: *Python, Django, Next.js, Recharts, Gemini API, REST APIs*

- Priced batteries from metal content and live commodity APIs, normalised across currencies and units, with a **Gemini** verdict (underpriced, fair or overpriced) in a Next.js dashboard.

SARIMA Time-Series Price Forecasting | Built
Tech: *Python, Statsmodels, Pandas, NumPy*

- Grid-searched 36 models by **AIC** after ADF testing and Ljung-Box residual checks, reaching **1.17% MAPE** over 8 rolling backtests versus 1.57% for ARIMA.

Also built: Reverse logistics dashboards (HTML, CSS, JavaScript, REST APIs, OTP authentication), Loom Indica e-commerce admin, ReBAT 3D website (Next.js, Three.js), Mcerv and Fit Regimen apps (Flutter), and a Gmail and Gemini AI agent.

EDUCATION

M.Tech in Artificial Intelligence, MANIT Bhopal | 2025 - Present

B.Tech CSE (Hons.) in Artificial Intelligence, Sage University, Bhopal | Graduated June 2024 | CGPA 9.29, Gold Medal

PATENT, CERTIFICATIONS AND ACHIEVEMENTS

- **Indian Patent (Published):** A System to Check the Urine Level in a Catheter (IoT), Application No. 202321006914, Feb 2023; university Innovation of the Year award.
- **NPTEL:** Big Data Computing, IIT Kanpur (Elite + Gold, Top 5%); Data Analytics with Python, IIT Roorkee (Elite + Silver, Top 5%); Industry 4.0 and IIoT, IIT Kharagpur (Elite + Silver); NLP, IIT Kharagpur (Elite).
- **Microsoft Certified:** Azure AI Fundamentals | 3-star on LeetCode, about 400 problems solved.

LEADERSHIP

- School Captain (Class XII) and Vice Captain (Class XI); Sports Club President at Sage University for 3+ years.
- Student Coordinator, IEEE Conference 2022; led 400+ participant events; selected for Yuva Samwad 2022 with the Chief Minister of MP.