

Rahul Kumar Mall

Full Stack Developer

Github: github.com/Rahul1038402 | LinkedIn: [linkedin.com/in/rahul-kumar-mall](https://www.linkedin.com/in/rahul-kumar-mall)
rahul1038402@gmail.com | +91 7052566746 | Noida, Uttar Pradesh, India

SUMMARY

Computer Science undergraduate with 1 year of professional experience building full-stack systems and AI/ML pipelines. Skilled in backend design, RAG architectures, and shipping end-to-end products across web and inference infrastructure.

EXPERIENCE

Full Stack Developer Intern | Hexoforge LLC

Remote | November 2025 – Present

- Fine-tuned Gemma 4 31B (vLLM) with hybrid RAG retrieval (HyDE + BM25 + MMR) over Qdrant-indexed mortgage document collections; embeddings via Gemini Embedding 2 Preview
- Engineered a multi-layer guardrail system using regex classifiers and semantic triplet retrieval to detect and redirect off-topic or sensitive conversations by severity
- Integrated Mem0 memory (Qdrant vector + Neo4j graph) for cross-session personalization; async pipeline execution via Celery and Redis
- Built and deployed a multi-tenant HRM platform (Next.js 14, FastAPI, PostgreSQL, SQLAlchemy, Alembic) with RBAC, Celery + Redis queues, AWS S3 storage, Docker, and Docploy

SKILLS

Frontend:	React, Next.js, TypeScript, Tailwind CSS, JavaScript, HTML5, CSS3
Backend:	FastAPI, Flask, Node.js, Celery, SQLAlchemy, Alembic, Pydantic, RESTful APIs
Databases & Infra:	PostgreSQL, Supabase, Redis, Qdrant, Neo4j, MinIO, AWS S3
ML / AI:	Python, RAG pipelines, vLLM, HyDE + BM25 + MMR retrieval, Mem0
DevOps & Tools:	Git, Docker, GitHub Actions, Dokploy, Vercel, AWS

PROJECTS

Signal - Automated Tech News Pipeline

[Live Demo](#)

Tech Stack: GitHub Actions, HackerNews API, NewsAPI, Groq, Gemini 2.5 Flash, Supabase, Resend, Next.js

- Built a fully automated daily pipeline that aggregates tech news from HackerNews, RSS, and NewsAPI, ranked for relevance using Groq with Gemini fallback
- Designed a fallback LLM architecture to ensure pipeline reliability; results stored in Supabase and delivered via Resend
- Shipped a companion Next.js 14 PWA with a dark editorial design for browsing ranked stories

Resume AI - Resume Analysis & ATS Optimizer

[Live Demo](#)

Tech Stack: React, TypeScript, Tailwind CSS, Flask, LLMs, Supabase, PostgreSQL

- Built a full-stack tool for resume analysis, building, and ATS match scoring using React + Flask with LLM-powered keyword extraction (85% accuracy)
- Processed 100+ resumes; 30+ active users globally in beta

EDUCATION

2023 - 2027 B.Tech in **Computer Science** at **JSS Academy Of Technical Education, Noida** (GPA: 7.8/10.0)