

Sundaram Rajivkumar Rai

+91 9665153365 | sundaramrai02@gmail.com | linkedin.com/in/sundaramrai | github.com/sundaramrai | sundaramrai.vercel.app

WORK EXPERIENCE

Homeville Group

Associate Software Engineer

Mumbai

Jan 2025 – May 2026

- Re-architected the Singularity CreditWorld RM module, resolving stale data and race conditions by consolidating **MySQL** queries and **Firestore** service calls into a unified reactive data layer used concurrently by internal and external lending teams; integrated with the **AI extraction pipeline** to surface parsed document data within the loan origination workflow
- Built the Bharat Housing Network document vault end-to-end: **REST APIs**, **MySQL** schema design and **AWS S3** storage for loan applicant, property and identity documents; shipped an embedded **chatbot** backed by a context API serving real-time journey data to internal teams and external lending companies
- Owned the **EaseBuzz** payment integration end-to-end: server-side webhook for payment verification, **MySQL** status updates, downstream workflow triggers and **Pusher** WebSocket notifications across CRM, LOS and LMS; contributed to the team **Angular 17 to 19** migration that cut bundle size by **83 %** and build time by **64 %**

Zidio Development

Web Developer Intern

Remote

Jun – Sep 2024

- Built a full-stack **MERN** job portal with **JWT** authentication, application-level role-based access control with separate permission scopes for employers and job seekers, and custom **Express** middleware for auth, error handling and request validation
- Implemented compound indexes for multi-field search across job titles, locations and categories with server-side pagination and filtering, and built **REST APIs** covering job listings, application tracking and candidate management

Tata Steel Limited

Software Developer Intern

Mumbai

Aug 2023 – Jan 2024

- Built a centralised **React.js** dashboard backed by **Node.js** REST APIs serving machine-wise and plant-level **OEE** metrics per shift, replacing manual shift reporting across multiple manufacturing units and production lines
- Built a **Python** pipeline using **Pandas** and **NumPy** to parse and normalise machine exports, and applied **ARIMA** models via **pmdarima** auto-order selection and **statsmodels** fitting to forecast OEE values 12 months ahead across 3 plants and 30+ machines

PROJECTS

Vexlo | Self-Hosted Localhost Tunnel [code]

Go · SQLite · WebSocket · TCP · REST API · GitHub Actions

- Built a self-hosted localhost tunnel in **Go** using binary TCP transport with a custom framed protocol for reliable bidirectional request forwarding, serving as a drop-in **ngrok alternative** with zero external runtime dependencies
- Shipped a terminal-style inspection dashboard **embedded directly in the binary** with **SQLite** persistence, request mutation, replay and live **WebSocket** updates; hardened with token-gated tunnel registration, **HttpOnly** cookie auth and **ACME** TLS

Cipheria | Zero-Knowledge Password Manager [site] [code]

Next.js · FastAPI · PostgreSQL · Alembic · Redis · AES-256-GCM · PBKDF2-SHA256 · JWT

- Architected a zero-knowledge model where all encryption runs **client-side via AES-256-GCM** with keys derived using **PBKDF2-SHA256 at 600 k iterations**, so the server stores only ciphertext and the master password is never transmitted
- Hardened the API with **Redis-based per-endpoint rate limiting** and **JWT** rotation featuring short-lived access tokens with refresh token reuse detection, making credential-stuffing attacks computationally expensive

Berkshire Hathaway Intelligence | RAG Chat System [site] [code]

Next.js · TypeScript · Mastra · PostgreSQL · pgvector · Mistral Embeddings · Ollama

- Built an end-to-end RAG system over 48 years of Berkshire Hathaway shareholder letters, designing the full ingestion pipeline from **PDF parsing** through **MDocument** chunking to batched **pgvector** upserts with per-source re-ingestion support
- Wired a **Mastra** agent with **Mistral** embeddings and **pgvector** retrieval for grounded responses, persistent **PostgreSQL**-backed conversational memory across sessions, and a streaming **Next.js** chat UI with clickable source attribution

Juris AI | Legal Query Platform [site] [frontend] [backend]

Angular · Node.js · MongoDB · Google Gemini API · JWT

- Built a legal query platform that grounds every answer in live web sources by chaining **Google Search** retrieval, context injection and **Gemini** generation, reducing reliance on model memory for legal responses
- Preprocessed raw queries to extract **legal intent** before retrieval to improve answer relevance, and structured the **RAG pipeline** to cleanly separate query understanding from context injection and generation

SKILLS

Core Languages: TypeScript, JavaScript, Python, Go, SQL, HTML, CSS

Frontend: Angular, React, Next.js, Tailwind CSS

Backend: Node.js, Express.js, FastAPI

Databases: PostgreSQL, pgvector, MongoDB, MySQL, SQLite, Firestore, Redis

AI and RAG: RAG pipelines, vector search, Mistral embeddings, LLM integration, Mastra, Google Gemini API

Infrastructure & Tools: Docker, GitHub Actions, AWS, Vercel, Figma, Agile/Scrum, DevOps

EDUCATION

Thakur College of Engineering and Technology

B.E. in Computer Engineering

2021 – 2025

CGPA: 9.12 / 10