

OM JEE MISHRA

Senior AI Engineer · Agentic Systems, RAG & Full-Stack

Gurugram, Haryana, India

📞 +91 9411097707 ✉ omjeem558@gmail.com [in LinkedIn](#) [GitHub](#) [Portfolio](#)

Professional Experience

HexaHealth

July 2025 – Present

Senior Software Developer

Gurugram, India

- Architected **HealthGpt** ([Live](#)), HexaHealth's AI gateway – a single **Mastra**-based agent runtime powering a streaming text chatbot (**SSE**) and a realtime voice assistant (**WebSocket** ↔ **Gemini Live**) from one shared brain, with namespaced **Pinecone RAG** and persistent **PostgreSQL** memory hydrated from live **CRM/EHR** context.
- Owned a server-authoritative **booking engine** that never fabricates a lead; hardened both channels with **prompt-injection defense** and rate/IP guards, cut a **2–3s** voice-assistant startup stall via non-blocking memory writes, and instrumented end-to-end **Langfuse** tracing with **LLM-as-judge** medical-safety scorers.

Wasserstoff RJ Innovations Private Limited

June 2023 – July 2025

Software Developer

Gurugram, India

- Engineered the **backend** for **Litt Law** ([Live](#)), an AI-powered Indian-law contract-drafting platform – **Node.js/Express** microservices over **gRPC** with **PostgreSQL**, **Redis**, and **Drizzle ORM**; shipped **40+ REST APIs** with JWT auth at **99.9% uptime** and **2.3× throughput**.
- Built **RAG pipelines** with custom embeddings on **Pinecone** for legal contract generation (**48% latency reduction**, **1,500+ concurrent users**, **95%+ test coverage**), plus real-time collaborative editing over **SSE** and **Razorpay/Cashfree** payments at **100% reliability** across **500+ monthly signings**.

Open Source

C.O.R.E. | *Agentic AI Memory OS for LLM Agents* | [RedPlanetHQ](#), [YC S23](#) · [1.9k★](#) | [Contributions](#)

9 PRs Merged

- Shipped **16 read-only MCP tools** – a full **Postman integration** and **skills-as-tools** ([#850](#), [#886](#)) letting **Claude Code**, **Codex**, **Cursor** discover skills on demand; unblocked self-hosted **Ollama** inference end-to-end ([#885](#)) and added **62 unit tests** ([#905](#)) – **+4.1K lines** across **9 merged PRs**.

Town | *Multiplayer AI-NPC Pixel World* | [RedPlanetHQ](#) · [Live](#) | [Contributions](#)

7 PRs · 5 Merged

- Added **Ollama** as a third LLM provider across app, CLI, and billing ([#37](#), [#45](#), [#47](#)) so Town runs free with no paid API key; designed visitor-granted **CORE** integrations for NPCs via per-person token routing ([#72](#)) and bootstrapped the repo's first test suite – **61 unit tests** ([#34](#)).

Projects

PostMind AI | *LinkedIn & Twitter Content Intelligence Platform* | [Live](#) | [Github](#)

- Architected a **6-agent Mastra pipeline** generating **5–7** personalized content briefs via a **provider-agnostic LLM factory** (**Gemini/Ollama Cloud**).
- Engineered a deterministic **Writing-DNA engine** – **15+ voice metrics** at zero LLM cost, 100% reproducible – feeding a recency-decayed feedback loop.

AI Learning Lab | *Browser-Based, Zero-Chatbot AI Education Platform* | [Live](#) | [Github](#)

- Built a gamified course teaching how **LLMs** work by operating one – live embeddings/attention/gradients via **transformers.js** & **WebLLM**, 22 chapters.
- Pure **TypeScript** engines (zero React/DOM deps), **732 offline tests** (~1s); Worlds 1–5 run **zero-network** via Cache Storage + IndexedDB.

Education

GL Bajaj Group of Institutions, Mathura

November 2020 – June 2024

Bachelor of Technology in Computer Science Engineering (CGPA: 8.01)

Technical Skills

Languages & Frameworks: TypeScript, JavaScript (ES6+), Node.js, Express.js, React, Next.js (SSR), Tailwind CSS, Redux.

AI & Agentic Systems: LLMs, Agentic & Multi-Agent Systems, RAG Pipelines, Vector Search & Embeddings, MCP (Model Context Protocol), Mastra AI Framework, Prompt Engineering, Gemini Live API, Ollama, Pinecone, LLM Evals, Memory Management.

Databases & Infrastructure: PostgreSQL, MySQL, MongoDB, Redis, Drizzle ORM, Docker, Google Cloud Platform (GCP), CI/CD (GitHub Actions).

Tools & Integrations: Git, WebSocket, SSE, gRPC, Langfuse, OpenTelemetry, Vercel AI SDK, Razorpay, Cashfree, JWT/OAuth2, Zod.