

Akshit Pareta

Bundi, Rajasthan, India • akshitpareta01@gmail.com • +91-9680810188

akshitpareta.com • github.com/akshitpareta • linkedin.com/in/akshitpareta

Re: Application for AI Engineer / Applied AI Engineer

Dear Hiring Team,

I have been building production LLM applications for the past two years — not prototypes, not API wrappers, but systems that real businesses run on every day. I am applying because **[SPECIFIC_THING FROM YOUR JD / PRODUCT]** maps directly onto work I have already shipped.

My flagship system is a multi-tenant legal Retrieval-Augmented Generation (RAG) platform — Gemini Embeddings + ChromaDB over ~15K legal documents (~75K chunks), hierarchical chunking by statute and section, hybrid retrieval (BM25 + dense), and cross-encoder reranking. It serves six attorneys daily and is gated by a 340-case golden-set eval with LLM-as-judge and LangSmith tracing. That eval drove citation precision from 71% to 89% while holding hallucination at ~2–3%, and it saves 15–30 hours per attorney each week.

What sets me apart from most AI Engineer candidates is that I own the infrastructure, not just the API call. I built a self-hosted 2× RTX 5090 cluster (64GB GDDR7) serving quantized Qwen-72B (INT4) via vLLM with continuous batching and paged KV cache — 30–40 tokens/sec, p95 first-token latency under 1s — and I fine-tune with QLoRA (Transformers, PEFT, TRL, bitsandbytes). I also built Swarm OS, a personal multi-agent harness coordinating 50+ concurrent agents across five frontier models with per-agent context budgeting and dependency-cycle detection, cutting multi-step task latency ~40% versus a single-agent baseline.

I am commerce-educated and self-taught as an engineer, which I treat as an advantage: I read the business and the workflow before I write the spec, so my deployments get adopted rather than admired.

I would welcome a 15-minute call to walk through the legal RAG architecture and the eval harness. Worth it?

Best regards,

Akshit Pareta

akshitpareta.com • github.com/akshitpareta