Hemanth Jammigumpala

480.319.3688 | Gmail | LinkedIn | GitHub | Portfolio

Education

The University of Texas at Dallas | **B.S. in Computer Information Systems and Technology** **May 2026**
Courses: Applied AI & ML, Database Fundamentals, OOP (Python), BI & Analytics | *Certification: Google Cybersecurity*

Technical Skills

Languages & LLMs: Python, SQL | Claude, OpenAI, Gemini, Groq
AI/LLM & Agents: LangChain, LangGraph, LlamaIndex, MCP, Claude Agent SDK, Google ADK, Prompt Engineering
RAG & Vector Stores: Hybrid Search, Reranking, ChromaDB, Pinecone, Upstash Vector, PostgreSQL
Eval & Apps: RAGAS, LangSmith, Pytest, FastAPI, REST APIs, Next.js, Streamlit
Data & Tools: Pandas, Tableau, Alteryx, Snowflake, Excel, Git, Vercel, Claude Code

Professional Experience

State Farm (UTD Senior Capstone) | **AI Solutions Analyst** **Feb 2026 – May 2026**

- Built **RetentionIQ**, an AI-powered HR workforce intelligence agent on Stack AI using Claude Sonnet 4.5, connecting cleaned and structured HR data through a RAG pipeline (chunking, embeddings, vector database).
- Designed an LLM evaluation framework benchmarking **4 instruction-prompt configurations** across varying question difficulty, measuring accuracy, consistency and hallucination rate; **raised user satisfaction from 59% to 88%**.
- Reduced hallucination frequency using guardrails, custom system instructions, and step-wise, context-rich prompting, identified as the most effective strategy across benchmarks.
- Delivered a reusable AI prototype, evaluation framework and prompting best-practice guide to help State Farm HR offset a projected talent shortage, collaborating in a **6-person team** with client stakeholders.

Technical Projects

Event Planner | **Multi-Agent AI System** | *Python, LangGraph, Claude, MCP, FastAPI, Next.js, PostgreSQL, LangSmith*

- Architected a **4-agent LangGraph system** (orchestrator + Venue, Logistics, Invites) with parallel fan-out that plans an event end to end from a single natural-language request.
- Built a custom FastMCP server exposing **5 Gmail and Google Calendar tools** (OAuth 2.0) plus **4 native tools** (Tavily search, weather, budget, memory), with Pydantic structured outputs and a **human-in-the-loop interrupt**.
- Persisted shared state and long-term memory in PostgreSQL for multi-turn follow-ups, streamed agent steps to the UI over SSE, and traced every agent and tool call in LangSmith.

PriorAuth Copilot | **Prior Authorization Chatbot** | *LangChain, Groq, ChromaDB, FastAPI, RAGAS, LangSmith*

- Built a hybrid RAG chatbot that checks doctor's notes against **3 real Medicare policies** and **4,700+ ICD-10 codes**, citing the exact note sentence and policy page for billing staff.
- Engineered retrieval with BM25 + embeddings, reciprocal rank fusion and cross-encoder reranking, scoped to each procedure's policy.
- Reduced hallucinations by verifying every quote in code against the source and exact-matching diagnosis codes; reached **100% valid structured outputs** with JSON-schema mode.
- Added prompt-injection blocking, PHI redaction, RAGAS evaluation, LangSmith tracing and **51 automated tests**.

SparkSight | **EV Charging Infrastructure Analytics for DFW** | *Alteryx, ETL, Tableau*

- Led and developed a BI project analyzing EV charging gaps across DFW, building an Alteryx ETL workflow that cleaned, geocoded and spatially joined **3 data sources** (stations, 311 complaints, district boundaries).
- Designed KPIs such as demand-supply ratio and suitability score to rank **12 districts**, and built Tableau dashboards covering **1,891 chargers**, showing fast chargers were only **22%** of the network.
- Created and delivered the final presentation to a stakeholder judging panel, translating the analysis into a phased expansion roadmap prioritizing Dallas, Irving and Arlington.

AI Product

Ask Hemanth | **Conversational AI for Recruiters** | *Next.js, Gemini, Upstash Vector, Vercel*

- Designed and shipped a conversational RAG product on my portfolio site that answers recruiter questions about my experience, skills and projects with instant, grounded responses in my own voice.
- Added LLM function calling so the assistant takes actions (e.g., opening LinkedIn or GitHub), validated by a **46-case eval suite** covering hallucination, privacy, prompt injection and tool use.
- Production-ready at **$0/month** with guardrails, rate limiting, answer caching and **automatic failover across 3 LLMs**.
- Built AI-natively with Claude Code, using spec-first planning, test-driven development and frontend design skills; versioned on GitHub.