HARI PRASANNAA THANGAVEL RAVI

LinkedIn | hariprasannaa2001@gmail.com | 202-743-8633 | Portfolio | GitHub

SUMMARY

ML/AI Engineer (Python, React, AWS, Spark) delivering reliable LLM/RAG copilots for expert workflows (research synthesis, diligence Q&A, internal knowledge retrieval); built semantic search and grounded assistants over 10K+ resources with measurable gains (+10% resolution, 15% fewer zero-result queries).

PROFESSIONAL EXPERIENCE

Data Science Intern | MyEdMaster, *Ashburn, VA* May 2025 – Sep 2025

- Built and shipped a RAG-based citation-grounded knowledge assistant for support LlamaIndex and Pinecone, enforcing source citations and gating quality with Ragas (faithfulness, context precision, citation coverage), improving self-serve resolution by 10%.
- Developed a personalized recommendation system using a hybrid ranking approach (content + interaction signals) by engineering features from metadata and behavioral events with offline training and online serving; improved recommendation CTR by 8%.
- Developed LangChain orchestration services for multi-step prompt workflows (plan, retrieve, answer, verify) integrating MCP tool servers with structured outputs, validation, and retries, improving multi-hop QA by 35% on a 200-query benchmark.
- Improved responsiveness of high-traffic discovery endpoints by adding Redis caching in Django for feed tiles, filters, and recommendations, reducing p95 latency from 800ms to 500ms and lowering DB query load by 7% for core flows.

Data Scientist | Data Science for Sustainable Development, *Washington, DC* Apr 2025 – Present

- Built energy-demand forecasting to support planning and load management, training XGBoost models and serving forecasts via AWS-hosted prediction APIs, with batch scoring and automated retrains, achieving 12% MAPE across 50+ buildings.
- Modernized model delivery by migrating training and scheduled retrains to AWS pipelines (Kubernetes) via Terraform with CI/CD and operational logging/metrics, reducing the model release cycle 40% and enabling reliable cross-team access.

Member of Technical Staff | Facilio Technology Solutions, *Chennai, India* May 2022 – Jul 2024

- Refactored Node.js services using clean architecture and DDD for performance and maintainability. Integrated role-based access and Azure AD into React modules, enhancing security and UX for 3,800+ users.
- Developed backend REST APIs and a query-orchestration layer in Java for high-cardinality analytics, compiling user-selected filters, group-bys, and time windows into optimized ClickHouse SQL with guardrails, reducing p95 dashboard load time from 4.8s to 1.9s.
- Developed time-series forecasting ML for dashboards and alerts using rolling-origin backtests and per-tenant calibration (ARIMA and LSTM) over Kafka-to-ClickHouse telemetry; improved MAPE from 18% to 11% across 8K+ buildings.
- Collaborated with product and customer-facing teams to define KPI semantics and drilldown requirements, translating feedback into config-driven dashboard specs and iterative releases for enterprise tenants.

RESEARCH EXPERIENCE

Dynamical Systems Analysis of LLM Hallucinations Sep 2025 – Present

- Modeled hallucination dynamics as a system using TF-IDF plus cosine clustering and a truth-aware classifier, and implemented transformer components using PyTorch with sampling to quantify time-to-settle into truthful vs false token attractors across prompts.

PROJECTS

LLM Gateway Nov 2025 – Dec 2025

- Built an enterprise RAG gateway (FastAPI, Pinecone) routing via OpenAI API and Hugging Face Inference Endpoints with versioned prompts and write-through caching; improved Recall@5 by 12% on 500-query benchmark and deployed on Baseten.

Personalized LLM Assistant with Custom Fine-Tuning Jun 2025 – Sep 2025

- Developed a personalized LLM assistant using fine-tuned Mistral, LLaMA with RAG; built C# backend, PostgreSQL orchestrating semantic search (Sentence Transformers, OpenAI embeddings) on Azure ML with telemetry, logging, and secure APIs for monitoring.

EDUCATION

The George Washington University Washington, DC

Master's in Data Science, **GPA: 3.92/4.0, Global Leaders Award** Aug 2024 – Expected May 2026

- Coursework: Machine Learning, Reinforcement Learning, NLP, Cloud Computing, Data Warehousing, Data Visualization

TECHNICAL SKILLS

LLM: LangChain, LlamaIndex, MCP, prompt engineering, validation/retries, prompt/version testing | **Model APIs:** OpenAI API, Hugging Face API, Anthropic API | **RAG:** chunking + indexing, hybrid retrieval, re-ranking, grounding + citations, semantic caching, Ragas | **DB:** Pinecone, FAISS, Redis, RedisSearch, ClickHouse, PostgreSQL | **ML:** XGBoost, PyTorch, scikit-learn, A/B testing, TensorFlow/Keras, time-series forecasting, anomaly detection | **Backend:** Python, FastAPI, Spark, Java, GraphQL, SQL, Kafka, Node.js | **Frontend:** JavaScript, React, HTML, CSS, Git | **Cloud/Observability:** AWS, Azure, Docker, CI/CD, Airflow, Datadog, Kubernetes