AMAL RAJ

AI Engineer | Applied AI, Data Engineering & LLM Systems

Bengaluru, KA | 8884991063 | obweus@outlook.com | eamal.vercel.app

SUMMARY

Applied AI / Data Engineer with 3+ years of experience at the intersection of data engineering, AI evaluation, and LLM-powered agentic systems. Strong background in building Python-driven evaluation pipelines on GCP (Vertex AI, BigQuery), managing large-scale datasets, and benchmarking LLM and agent architectures for enterprise use cases. Currently an AI Forward Deployed Engineer at Nestingale, deployed to IQ-EQ, building FastAPI and Docker backend services that integrate Azure LLM APIs for loan administration, invoice processing, and document extraction, along with RAG, LangGraph, and MCP-based tooling.

PROFESSIONAL EXPERIENCE

Nestingale (Contractor) – Deployed at IQ-EQ

AI Forward Deployed Engineer | 09/2026 – Present

- Build backend services integrating Azure-hosted LLM APIs into a loan administration application.
- Develop an extraction engine and invoice processing workflows that turn unstructured financial documents into structured data.
- Design and containerize RESTful APIs with Python and FastAPI, deployed with Docker.
- Build RAG pipelines and LangGraph agent workflows, with MCP servers exposing tools to the LLM for grounded, document-based answers.

Emergence.ai – Bengaluru, KA

Associate Software Engineer (AI Evaluation & Benchmarking) | 06/2024 – 04/2026

AI lab specializing in coordinating autonomous agents across enterprise software environments

- Built and maintained Python-based evaluation pipelines for LLMs and agentic AI systems on GCP Vertex AI.
- Evaluated enterprise-grade agent architectures including GenAgents, Web Agents, and Text-to-SQL agents, focusing on correctness, reasoning depth, and robustness.
- Implemented LLM-as-a-Judge frameworks for scalable qualitative evaluation and regression testing.
- Owned benchmarking datasets in BigQuery, designing schemas for evaluation inputs and outputs and enabling structured cross-model and cross-prompt comparisons.
- Analyzed large-scale evaluation data to identify performance bottlenecks, failure modes, and optimization opportunities.
- Partnered with research and engineering teams to align evaluation pipelines with deployment and production needs.
- Improved benchmarking scalability, reproducibility, and reliability across multiple agent types and workflows.

Merlyn Mind – Bengaluru, KA

Annotation Analyst | 12/2022 – 06/2024

- Conducted inter-annotator agreement (IAA) analysis using Python to measure consistency and identify ambiguous cases.
- Implemented majority-vote and consensus logic in Python to resolve annotation conflicts and finalize ground-truth labels.

EDUCATION & CERTIFICATIONS

OPJS University – Churu, RJ

BTech, Chemical Engineering | 04/2018 – 04/2021

TECHNICAL SKILLS

Python, FastAPI, RESTful APIs, Docker, SQL, Google Cloud Platform (GCP), BigQuery, Vertex AI, Azure, LLM APIs, LangGraph, MCP, RAG, Data Engineering Foundations, Data Modeling, AI Model Evaluation, LLM-as-a-Judge, Agentic AI Systems, Prompt Engineering, Data Quality, Git, Excel, Tableau

RESEARCH & PUBLICATIONS

- Deepak Akkil, Mowafak Allaham, Amal Raj, et al. “Emergence WebVoyager: Toward Consistent and Transparent Evaluation of (Web) Agents in The Wild.” arXiv, 2026, arXiv:2603.29020.
- Research Contributor – LLM & Agentic Evaluation Benchmarks: Provided high-quality human-in-the-loop annotations and logic verification for multimodal agentic workflows, contributing to validation datasets of published research including arXiv:2410.00689 and arXiv:2407.13032.