Hyderabad
Telangana
India
+91 6371895195
biswajit.jena211@gmail.com
biswajitjenaofficial
Ahrvzihsahst
Portfolio

Biswajit Jena

AI Associate Software Engineer

Professional Summary

AI Associate Software Engineer specializing in multi-agent orchestration systems, GraphRAG implementations, and LLM-powered applications for renewable energy analytics. Experienced in architecting scalable FastAPI backends with LangGraph state machines, PostgreSQL checkpointing, AWS Neptune graph databases, and async Python patterns. Currently building SupervisoryAgentSolar and AutoGraphRagService at GreenBridge.AI—multi-tenant platforms managing 10+ specialized agents and knowledge graph systems for solar plant performance monitoring and fault analysis, serving 500+ daily requests with 80% uptime. Expert in knowledge graph construction, multi-strategy retrieval (agentic, path-based, query-based), prompt engineering, streaming APIs (SSE), and cloud infrastructure (AWS, Azure). Proven track record of optimizing system performance through Redis caching (70% hit rate), reducing API response times from 8s to 2s, and automating graph construction with 90% effort reduction.

Work Experience

July 2024 – Present — **AI Associate Software Engineer**, *GreenBridge.AI*, Hyderabad, India

- Architected multi-agent orchestration system using LangGraph state machines with PostgreSQL checkpointing, managing 10+ specialized agents (RUL analysis, Specific Yield, Loss Factor, QnA) for renewable energy performance analytics
- Built GraphRAG-powered knowledge graph system for solar inverter fault analysis, automating entity extraction from unstructured documentation using KGGen library and AWS Neptune for graph storage with BYOKG-RAG multi-strategy retrieval
- Implemented streaming REST APIs with Server-Sent Events (SSE) delivering real-time progress updates, reducing user wait time perception by 60% and enabling human-in-the-loop interrupts for dynamic data collection
- Built multi-tenant SaaS architecture with AWS Cognito authentication, supporting tenant isolation and both user + M2M token flows, serving 500+ daily requests across multiple facilities
- Optimized system performance with Redis caching layer and urllib3 connection pooling, achieving 70% cache hit rate and reducing API response times from 8s to 2s
- Developed async Python backend using FastAPI with full async/await patterns, PostgreSQL async operations (psycopg), and dynamic LLM routing (OpenAI/AWS Bedrock) for cost optimization
- Engineered conversation memory management with 25-message cap and state persistence, enabling context-aware multi-turn interactions while preventing memory bloat
- Integrated Langfuse observability for LLM tracing and monitoring, tracking 1000+ agent executions daily with detailed performance metrics
- Implemented Docker containerization with Azure DevOps CI/CD pipelines, achieving 80% uptime, zero-downtime deployments to AWS EC2, and automated rollback mechanisms
- Designed custom exception handling framework and standardized error responses across 12+ agents, improving debugging efficiency by 40%

April 2024 – July 2024 — **AI Engineer Intern**, *GreenBridge.AI*, Hyderabad, India

- Designed domain-specific chatbots handling 500+ wind-energy inquiries, reducing response times by 40%
- Extended chatbot capabilities with Llama 3.1 for NLQ-to-SQL translation against a Snowflake backend
- Implemented Python-based Azure Functions to validate data quality rules, catching and resolving 98% of anomalies before downstream consumption
- Skills: Python, LLM, Web Scraping, OCR, Snowflake, Azure Functions, AWS Bedrock, Llama models

January 2024 – April 2024 **Data Science Intern**, *NareshIt*, Hyderabad, India

- Automated Python data cleaning with pandas, NumPy, and SQL to process 1M+ records, reducing time by 30%
- Built and containerized ML models using scikit-learn and TensorFlow (95% accuracy)
- Developed Flask-based UI using JavaScript, HTML, and CSS; deployed and hosted on AWS EC2
- Skills: Python, AWS, Flask, Machine Learning, Data Analysis, Hugging Face

Education

2018 – 2022 **Bachelor of Technology in Computer Science**, *PMEC*, Ganjam, Odisha

2016 – 2018 **Associate Degree in Science**, *Talanagar College*, Balasore

2016 **High School (PCM-CS)**, *GSREM*, Bhubaneswar

Technical Skills

Programming Languages	Python, JavaScript, HTML/CSS
Backend & APIs	FastAPI (async/await, streaming SSE), PostgreSQL (async psycopg), Redis caching, RESTful API design, Multi-tenant architecture, Connection pooling (urllib3)
AI/ML & LLM	LangChain, LangGraph (state machines, checkpointing, human-in-the-loop), Multi-agent orchestration, Prompt engineering, Dynamic LLM routing, GraphRAG, Knowledge Graph extraction (KGGen), AWS GraphRAG Toolkit (BYOKG-RAG), Entity linking, Agentic retrieval, Multi-hop reasoning, Triplet extraction, OpenAI GPT-4, AWS Bedrock (Claude), Langfuse observability, TensorFlow, scikit-learn, Hugging Face
State Management & Databases	PostgreSQL (sync & async operations), State persistence & checkpointing, Redis caching strategies, Conversation memory management, AWS Neptune (graph database), Gremlin queries, Graph traversal, Knowledge Graph construction, Entity-relationship mapping, Snowflake, SQL
Cloud & Infrastructure	AWS (EC2, Cognito, Bedrock, Neptune, S3 bulk loading), Azure (Functions, DevOps, CI/CD pipelines), Docker containerization, Zero-downtime deployments, boto3 SDK, Gremlin Python client
Architecture Patterns	Singleton pattern, MVC architecture, Supervisor orchestration pattern, Human-in-the-loop patterns, Event-driven architecture (SSE), GraphRAG, REST APIs, Microservices
Tools & Technologies	Git, Azure DevOps, Llama models, Web Scraping, OCR
Data Processing	Pandas, NumPy, SQL, Data pipeline automation

Key Projects

SupervisoryAgentSolar - Multi-Agent Renewable Energy Analytics Platform

Tech Stack: FastAPI, LangGraph, PostgreSQL, Redis, AWS Bedrock, Docker, Azure DevOps

- Led development of intelligent agent orchestration system for solar plant performance monitoring and analytics
- Architected supervisor-agent pattern managing 10+ specialized agents for domain-specific tasks: RulCustomAgent (Remaining Useful Life analysis), SyInverterAnalysisAgent (Specific Yield optimization), LossFactorAgent (Performance loss detection), QnA Agent (Natural language query interface), and additional agents for EBA, OpEx Analysis, SCB Analysis, Inverter Efficiency, Clipping Detection
- Implemented PostgreSQL-based state checkpointing for conversation continuity across sessions
- Built streaming API architecture with SSE for real-time progress updates during multi-step agent workflows
- Designed multi-tenant authentication system with AWS Cognito supporting user and M2M token flows
- Integrated Redis caching layer reducing response times by 75% for frequently accessed analytics
- Impact: Processing 500+ daily queries, 80% system uptime, 70% improvement in data retrieval accuracy

AutoGraphRagService - Knowledge Graph RAG for Solar Inverter Fault Analysis

Tech Stack: FastAPI, AWS Neptune, AWS GraphRAG Toolkit, KGGen, Gremlin, S3, OpenAI/Bedrock

- Developed end-to-end knowledge graph ingestion and query system for solar inverter fault documentation analysis
- Implemented LLM-powered entity extraction pipeline using KGGen library to automatically identify components, faults, subcomponents, and recommendations from unstructured text
- Architected automated CSV transformation service converting extracted knowledge graphs into Neptune bulk loader format (vertices/edges/relationships)
- Built asynchronous job processing system with status tracking for parallel ingestion of multiple fault documents
- Integrated AWS Neptune graph database with Gremlin for efficient graph traversal and complex relationship queries
- Developed BYOKG-RAG (Bring Your Own Knowledge Graph) implementation using AWS GraphRAG Toolkit with: Agentic Retriever (LLM-guided triplet extraction), Path Retriever (Multi-hop reasoning through relationship chains), Graph Query Retriever (Direct Gremlin query generation), Entity Linker (Fuzzy string matching for question-to-graph entity mapping)
- Designed fault diagnosis query engine performing iterative multi-strategy retrieval (5 iterations, max 100 triplets) and generating structured recommendations with supporting evidence
- Implemented S3 integration for bulk CSV uploads with fallback local storage for development/testing
- Built RESTful endpoints for KG ingestion, job monitoring, and intelligent querying
- Impact: Automated fault knowledge extraction from 50MB+ documentation, enabled natural language queries over graph data, reduced manual graph construction effort by 90%

Real-time Hand Sign Recognition System

- Implemented a real-time hand sign recognition system using Python, OpenCV, Keras, and TensorFlow
- Utilized computer vision techniques for real-time processing and accurate gesture recognition

Customer Churn Analysis

- Developed a predictive model using Python, scikit-learn, and data visualization tools
- Implemented advanced analytics for proactive customer churn prediction

Certifications

Google Advanced Data Analytics Professional Certificate

Machine Learning Specialization

Data Science & AI Professional Certificate

Python Nano Degree

Google Data Analytics Professional Certificate

Languages

English	Professional
Odia	Native
Hindi	Professional