

BRIJ VAGHANI

AI SOFTWARE ENGINEER

☎ +91 6359724811 — ✉ VAGHANIBRIJ222@GMAIL.COM — 🌐 LINKEDIN.COM/IN/BRIJ2003 — 🐙 GITHUB.COM/BRIJ2003

Education

Indian Institute Of Information Technology Dharwad (IIIT Dharwad)

Dharwad, Karnataka

Bachelor of Technology in Computer Science: GPA-7.27

Professional Summary

Senior Applied AI Engineer with 3 years experience focused on multi-agent orchestration, enterprise AI platforms, and scalable cloud-native architectures. Skilled at designing and deploying advanced agent-based systems that support long-context interactions and efficient agent memory, fine-tuning large language models, and implementing guardrails for safe AI outputs. Demonstrated success building multi-agent solutions using LangChain, AWS, and modular MCP servers to enable real-time workflow automation and task orchestration across multiple industries.

Professional Experience (3 Years)

Torvalds

Applied AI Engineer

Present

- Built an efficient, context-engineered system for airline operations automation using multiple interacting AI agents and an upgraded permission-based Mem0 memory/context solution.
- Designed and Engineered a complex legal validation systems, built from the ground up and comparable to Jeppesen and AIMS—to verify flight legality and ensure regulatory compliance before takeoff.
- Architected an enterprise AI operations platform for airline clients, reducing manual workflows and replacing with intelligent agent-based automation and vision-powered document processing.
- Designed and deployed a multi-layer orchestrated workflow system integrating AI agents with tool-calling capabilities to interface with airline APIs for real-time data retrieval, updates, and automated work order generation for ground staff.

Cloudbloom

Software Engineer

2023 - 2025

- Developed and deployed GenAI-powered backend systems leveraging AWS Lambda, DynamoDB, and Cognito, boosting product scalability and reliability for diverse AI workloads.
- Designed, integrated, and optimized microservices to serve agent-based and LLM-powered workflows (using technologies like LangChain, LlamaIndex and LangGraph).
- Led development and automation of intelligent pipelines for event-driven AI use cases, including semantic search, conversational AI, and real-time data processing.
- Encouraged Agile collaboration to build, iterate, and refine AI features in cloud-native environments, aligning product delivery with evolving GenAI strategies.

MadAlgos

Software Engineer Intern

2022

- Collaborated on AI-driven UI/UX enhancements and contributed to backend systems enabling real-time, collaborative learning and code assessment platforms using live video and coding environments.
- Designed Azure pipelines for continuous integration and orchestrated feature deployments for intelligent EdTech solutions.
- Developed a real-time collaborative coding IDE and integrated live video chat functionality using WebSockets to simulate a realistic interview environment.

Projects

Enterprise AI Airline Operations Platform - dCortex

- Architected a declarative, DSL-driven regulatory compliance engine with typed data contracts, encoding crew duty and flight-time rules across multiple regimes (EASA, FAA Part 117, ICAO, DGCA) as versioned, independently testable rule sets decoupled from agent logic.
- Built an airline-agnostic Flight Time Limitation (FTL) legality validator evaluating duty and rest compliance across 24-hour, 7-day, and 28-day rolling windows, with configurable per-carrier rule objects that handle airline-specific overrides without code changes.
- Separated probabilistic LLM reasoning from deterministic compliance checking, so every legality verdict traces to an explicit, auditable rule rather than model output—critical for safety-regulated, regulator-facing aviation decisions.

- Integrated vision-language models to parse crew rosters, NOTAMs, and operational PDFs into structured, schema-validated data feeding the rules engine and downstream agents.
- Engineered a permission-scoped, namespaced agent memory layer (extending Mem0) with lifecycle-managed context, enabling long-horizon multi-agent workflows with strict per-tenant data isolation.
- Exposed compliance and operations capabilities as portable MCP servers and tool-calling endpoints, letting agents dynamically attach to any workflow and interface with airline APIs for real-time data retrieval and automated work-order generation.

GenAI based Learning & Interaction Platform - Hitwit

- Designed and implemented a RAG (Retrieval-Augmented Generation) pipeline to power chatbot conversations, podcast discussions, and knowledge base interactions using LLMs.
- Integrated LLMs with third-party tools like Zoom and Slack, enabling interactive assistant features within real-time collaboration platforms.
- Built a dual-portal web platform (admin/user) for tracking, managing, and analyzing chatbot data in real-time.
- Developed intelligent content generators: dynamic podcast dialogue, auto-course creation, and AI-generated mind maps.
- Built backend using AWS Lambda, DynamoDB, SQS, and SES to ensure high availability and performance.
- Developed a Chrome extension to summarize web pages and generate insights, visuals, etc. instantly using contextual AI.
- Built in-site widgets for ai powered user guidance and seamless navigation using site-map concept, enhancing the overall UX in site navigation.

Agentic Human-Resource Management System - TalentFlow

- Developed a GenAI-driven recruitment SaaS featuring automated candidate screening, intelligent resume parsing, smart assessments, and AI-powered communication flows.
- Engineered end-to-end microservices architecture with secure, multi-tenant management, JWT authentication, and granular RBAC.
- Built real-time dashboards and analytics leveraging GenAI for performance prediction, skill matching, and recruitment insights.
- Integrated AWS S3, LinkedIn, Gmail, and Outlook APIs for rich document, search, and communication workflows.

AI-Powered Chat-Based Event Management System - Flo360

- Engineered an agentic AI workflow using n8n and LLMs to interpret natural language commands from chat platforms (e.g., WhatsApp, Telegram).
- Implemented a full CI/CD pipeline using Docker and AWS EC2, ensuring automated, seamless deployment of backend services.
- Automated event creation, booking, and user invites through GenAI agents orchestrated in n8n, with full CI/CD for seamless cloud deployments.

Multi-tenant LMS Platform - Arivu

- Built a custom AI orchestration engine to interface with multiple LLM endpoints for processing dynamic queries and generating insights.
- *Architected scalable knowledge base systems aggregating and managing large collections of enterprise documents; built custom RAG pipelines that empower LLMs to provide precise, up-to-date answers by retrieving and synthesizing multi-document context.*
- Engineered microservices for authentication (AWS Cognito) and authorization, ensuring secure, scalable access control.
- Architected and contributed to a multi-tenant frontend system, allowing scalable onboarding and role-based feature access
- Enabled fine-grained access control via role management microservice across tenants, improving data separation and customization flexibility for enterprise clients.

Technical Skills

Programming Languages: C/Cpp, Python, HTML/CSS, TypeScript, JavaScript, SQL, Go

AI/ML & Agent Orchestration: LangChain, LangGraph, LlamaIndex, RAG Pipelines, Multi-agent Systems, Tool-calling Frameworks, MCP Servers, LLM Fine-tuning, Knowledge Base Integration

Frontend: ReactJS, NextJS, ReactNative, PHP

Backend: NodeJS, ExpressJS, Serverless, Flask, FastAPI, Golang

Tools: Azure, AWS (Bedrock, SES, SQS, S3, RDS, Lambda, Cognito, Amplify, IAM, EC2, API Gateway, etc.), Docker, Git, Github, N8N, MCP's, Cursor, Firebase

Databases: NoSQL(MongoDB, DynamoDB), SQL(MySQL, PostgreSQL, Aurora RDS, SQLite), VectorDB (pgvector, Pinecone, Qdrant)

Tooled Libraries: MUI, Redux, Huggingface, BeautifulSoup, Selenium

Achievements

- Flipkart Grid 2023 Semi-finalist
- Runner-up in an inter-college startup hackathon