

# SHREYAS JAGANNATH

Agentic AI Engineer | Multi-Agent Systems | LLM Platforms

+44 7438 193 743 | London, UK | [shreyasjag@hotmail.com](mailto:shreyasjag@hotmail.com)

[shreyasjagannath.com](http://shreyasjagannath.com) | [linkedin.com/in/shreyasjagannath](https://linkedin.com/in/shreyasjagannath) | [github.com/jaggernaut007](https://github.com/jaggernaut007)

---

## PROFILE

---

Agentic AI Engineer with experience building and shipping production AI systems, from full stack solutions to LLM based multi-agent platforms. Proven track record delivering scalable systems with measurable impact. Experienced in combining deep technical execution with stakeholder-facing delivery. Seeking roles focused on agentic AI systems and end-to-end ownership.

## CORE SKILLS

---

**Agentic AI & LLMs:** Multi-agent orchestration, MCP, RAG, LangGraph, LangChain, LlamaIndex, Google ADK, Evaluations, Observability, Knowledge Graphs

**Engineering & Cloud:** Python, FastAPI, React.js, Node.js, Docker, Terraform, GCP (Vertex AI, Cloud Run, Pub/Sub, BigQuery), AWS (Lambda, ECS, Bedrock, S3, Kinesis, DynamoDB)

**Data & Storage:** PostgreSQL, Neo4j, VectorDBs ( LanceDB, ChromaDB, Pinecone)

## EXPERIENCE

---

### Lead Software Engineer

Dec 2025 – Present

*Move By Vision (Pro bono)*

- Stabilised a failing Flutter + GCP platform by building a guardrail and validation layer across Vertex AI integrations,
- Reduced generation latency by 55% (15s to 7s) through optimised execution flows

### Lead AI Engineer

Aug 2025 – Nov 2025

*Klyft Technologies (Founding Team)*

- Built a multi-agent AI platform for fitness coaching from zero to production on GCP (Vertex AI, Google ADK, Gemini, BigQuery, Cloud Run, Terraform)
- Designed LLM inference pipelines, memory systems, and session management to support scalable agent workflows
- Implemented RAG pipelines, evaluation, and observability frameworks to improve system reliability and output quality
- Delivered the core workout calibration feature and improving agent efficiency by 3x (60s to 20s) and driving user adoption
- Reduced LLM token usage by 10x through prompt optimisation and caching, balancing cost with performance
- Maintained reliable production deployments and ensured the system scaled with user demand

### Lead AI Engineer

Oct 2024 - Aug 2025

*Scan and Sew Inc. (Founding Team)*

- Led AI architecture and hands-on development across two production platforms
- Built a Designer AI Assistant using RAG and vision-language models (VLMs), enabling real-time querying over internal design libraries
- Developed a real-time manufacturing lifecycle platform (React, AWS Lambda, GraphQL, DynamoDB)

- Delivered both systems to production with active usage across design and operational workflows
- Made pragmatic build vs buy decisions (e.g. integrating external 3DGS APIs) to accelerate delivery without over-engineering and understanding trade-offs

### **Lead Software Engineer / CTO**

Jan 2020 - Aug 2023

*AIMAGE Technologies (Seed-Stage, Founder)*

- Built and led a cross-functional team of 10 across engineering, product, and design
- Designed and developed an NLP-driven visual search and recommendation platform on TensorFlow and a visual vector-database RAG knowledge graph
- Built and deployed a virtual try-on engine leveraging OpenCV, PyTorch, and Unity3D
- Delivered both systems to production on AWS (Lambda, ECS, Neptune, DynamoDB, S3, Kinesis) with GraphQL APIs and end-to-end infrastructure ownership
- Deployed on-site with a key retail client, working directly on the shop floor to understand product discovery and try-on challenges
- Iterated on both platforms using real customer feedback, building and validating the product during client sessions
- Improved conversion rates and reduced bounce rates through the visual search

## **EARLIER CAREER**

---

### **Software Consultant**

Oct 2019 - Mar 2021

*Mage Ventures Pvt. Ltd.*

- Delivered production systems including a payment gateway, computer vision based retail checkout solution, and data automation platform

### **AI Consultant (Founding Team)**

Oct 2018 - Oct 2021

*Cellstrat Inc.*

- Built AI/ML POCs for enterprise clients (Airbus, Volvo, Target) and delivered workshops for 100+ professionals

### **Full Stack Engineer (Founding Team)**

Jul 2017 - Jul 2018

*EndGate Global Pvt. Ltd.*

- Developed a full-stack food delivery platform, focusing on order processing and automation

### **AI Research Intern**

Nov 2016 - Apr 2017

*ISRO, Indian Space Research Organisation*

- Conducted research on computer vision-based depth detection for lunar rover navigation

### **Software Developer Intern**

Oct 2014 – May 2015

*WebMobi Technologies Pvt. Ltd.*

- Worked on a dynamic web and cross-platform mobile application for events using HTML5, JavaScript, Bootstrap, PHP and Apache Cordova

## **NOTABLE OPEN-SOURCE PROJECTS**

---

### **Nexus-MCP: Hybrid Code Intelligence Server**

- MCP server enabling token-efficient code understanding for AI agents (15–30% reduction/ 10s to 30s reduction per session)
- Implemented hybrid search (vector + BM25 + code graph) with re-ranking and semantic memory
- Designed for local-first usage, supporting multi-language parsing across 25+ languages

### **AI Coding Pipeline: Multi-Agent App Generator**

- A multi-agent system that generates functional mini-applications from user intent
- Designed multi-agent workflows covering UI generation, code writing, review, and testing
- Deployed on GCP using OpenAI APIs, React, and Vercel AI SDK, enabling real-time iteration and refinement

### **JobScout: Agentic Job Discovery Platform**

- Built an agentic job discovery platform aggregating 1500+ companies across multiple ATS platforms and APIs
- Implemented a LangGraph-based pipeline for embedding, similarity search, deduplication, and LLM-based ranking
- Developed an MCP server to enable structured agent interaction with job discovery tools

### **Health Intelligence Engine: Agentic Wellness Intelligence Platform**

- Built a production-ready agentic framework for scientifically-grounded wellness guidance using LangGraph, Neo4j Knowledge Graphs, and multi-agent orchestration
- Architected a knowledge graph integrating products, ingredients, biological mechanisms, symptoms, and clinical research (PubMed, NIH DSLD) to provide evidence-grounded recommendations
- Engineered a consultative Discovery Agent that dynamically generates clarification questions to resolve ambiguous user intent before retrieval and recommendation
- Implemented multi-layered safety guardrails including intent classification, jailbreak detection, contraindication validation, and fail-closed escalation workflows
- Developed deterministic orchestration pipelines with Pydantic-validated contracts, bounded retry loops, uncertainty-aware response handling, and DeepEval-based quality gates

## **EDUCATION**

---

### **MSc Artificial Intelligence (Distinction)**

2023 - 2024

*University of Surrey*

- Dissertation: Human avatar reconstruction using Gaussian Splatting for edge compute
- Built a local LLM token-classification system, improving rendering performance by 2x
- Winner, UKSEDS IOSM Competition 2024; Most Innovative Business Award

### **Master of Computer Applications (First Class)**

2013 - 2017

*Christ University*

- Co-authored research on AI and RL-based path planning
- Developed an IoT-based autonomous rover as capstone project

### **Bachelor of Computer Applications (First Class)**

2010 - 2013

*Bangalore University*

- Built a healthcare management system to streamline patient workflows
- Led Global Shapers (WEF) initiatives engaging with industry leaders

## **CERTIFICATIONS**

---

LangGraph & LangSmith (LangChain Academy) | Multi AI Agent Systems with CrewAI (Coursera) | Agentic RAG with LlamaIndex (Coursera) | Knowledge Graphs for RAG (Coursera) | Deep Learning & Computer Vision (Stanford Online) | Entrepreneurship: Idea to Launch (UC Berkeley) | Technology Entrepreneurship (Harvard Online)