Shreyas Jagannath

+44 7438193743 | shreyasjag@hotmail.com | London

LinkedIn | Github | Website

PROFILE

Senior AI Engineer with over 9 years of experience in developing and deploying production software and ML systems. Specialises in creating reliable, maintainable AI and software solutions within cross-functional teams. Expertise in Agentic AI, NLP, and Computer Vision. Skilled in delivering AI solutions aligned with defined product and business requirements. Proficient in software architectures and cloud infrastructures, with a focus on integrating AI and software engineering to deliver solutions that yield measurable business value.

EMPLOYMENT HISTORY

Aug 2025 — Nov 2025

Lead AI Engineer (Founding Team), Klyft Technologies Ltd.

- Delivered a multi-agent AI coaching platform for fitness, nutrition and behavioural change using Python, Google ADK, OpenAI and Gemini APIs on Vertex AI, delivering highly relevant workout plans with high completion and low abandonment in a 20-user pilot.
- Engineered a scalable GCP infrastructure (Cloud Run, Cloud Build, Pub/Sub, Vertex AI Search, BigQuery, Docker, Terraform), achieving low-latency responses, 100% pilot uptime and a 10x reduction in serving costs while keeping recommendation quality stable.

Oct 2024 — Aug 2025

Lead AI Engineer (Founding Team), Scan and Sew Inc.

- Implemented a generative AI assistant platform for 3D product designers using OpenAI LLM/VLM APIs and RAG and delivered a novel, production-ready design support tool.
- Research on Gaussian Splatting and PyTorch3D to enable spatially aware CAD 2D–3D workflows
- Built a real-time, scalable product manufacturing lifecycle tracking platform, integrating Gaussian Splatting–based 3D model creation with ReactJS, AWS Lambda, GraphQL, DynamoDB, Amazon Q and Amplify, delivering the end-to-end supply chain solution from concept to full deployment, iteratively delivered within the product roadmap.

Jan 2020 — Jul 2023

Chief Technology Officer, AIMAGE Technologies Pvt. Ltd.

- Designed and implemented an AI/AR virtual try-on engine using SLAM and 3D geometry, building 2D-to-3D conversion pipelines in OpenCV, PyTorch3D, Mediapipe and Unity3D to accurately capture facial and body motion across Android, iOS and web clients.
- Designed and developed an NLP-driven search and recommendation platform for e-commerce, combining TensorFlow/PyTorch models with a vector-database-backed visual RAG knowledge graph and automated product ingestion on AWS (API Gateway/OAuth2, DynamoDB, Neptune/Gremlin, Lambda, Beanstalk, Fargate, Amplify/GraphQL) and GCP Cloud Functions to deliver higher-relevance, low-friction product discovery for retailers.

Oct 2019 — Mar 2021

Software Consultant, Mage Ventures Pvt. Ltd.

- Implemented a secure cross-platform payment gateway for Walmart-owned PhonePe across multiple client apps, building Node.js/React services on AWS EC2 to reduce checkout friction and enable fast, reliable in-app payments.
- Built an AI-powered automated grocery checkout engine for Xplorazzi using computer vision (image segmentation), TensorFlow and FastAPI, improving checkout speed and consistency while reducing manual intervention.
- Created an automated data collection platform for Varthana Finance with Selenium, Python and Pandas, enabling streamlined data capture from 1,000+ schools.

Oct 2018 — Oct 2021

AI Software Consultant, Cellstrat Inc.

- Consulted on client AI/ML initiatives, designing and implementing models in PyTorch, TensorFlow, Keras and Python, containerised with Docker and deployed on AWS to deliver tailored AI solutions across multiple client projects.
- Conducted AI/ML workshops and research talks for 100+ professionals, creating hands-on materials and presenting advanced techniques to drive ML adoption and strengthen internal and external technical communities.

Jul 2018 — Oct 2018	**AI Engineer Intern, AugRay Inc.**

- Researched and prototyped a real-time pose detection pipeline in Python/PyTorch to drive 3D avatar animation for an internal application, comparing model architectures and tuning inference for low-latency, frame-stable tracking in interactive experiences.

- Developed an NLU chatbot using Dialogflow, RASA and Python to enable contextual conversations with 3D avatars, integrating intent/entity parsing with backend logic to deliver more personalised interactions and improve user engagement.

Jul 2017 — Jul 2018	**Full Stack Engineer (Founding Team), EndGate Global Pvt. Ltd.**

- Developed a full-stack food delivery platform using Java, Hibernate and JavaScript, owning order-processing and bulk-order automation workflows for high-volume clients to streamline operations and improve end-to-end ordering efficiency.

NOTABLE PROJECTS

Dec 2025	**Context-Aware Knowledge Graph Agent for Hyper-Personalised Video Generation**

- Built a Python-based NLP agent based on requirements, that converted raw consumer data into a Neo4j knowledge graph and structured vector store using NLTK, NetworkX, NER and Hugging Face embeddings, then exposed a retrieval layer that fed OpenAI, Gemini and Veo 3.1 video generation components with user-specific context at runtime to enable hyper-personalised video content.

Dec 2025	**Multi-Agent Natural-Language-to-Code Orchestration Prototype**

- Rapidly prototyped a collaborative, multi-agent AI system using Vercel AI, Next.js, React, Node.js, OpenAI and Zod that translated natural-language feature requests into executable coding-pipeline architectures and orchestrated specialist code-generation agents, demonstrably reducing development time for new features in POC trials.

Nov 2025	**Agentic Sales CRM – Multi-agent automated lead-generation prototype**

- Implemented an end-to-end agentic Sales CRM as an internal prototype in one week, building a multi-agent workflow with CrewAI to crawl the web, qualify target companies, enrich contact data and verify credentials, then persist leads into a Next.js/Clerk/Prisma/GraphQL CRM backed by a FastAPI service on GCP Cloud Run (Docker), automating outreach and lead capture and saving the sales team several hours of manual prospecting per day under zero-budget constraints.

EDUCATION

Sep 2023 — Sep 2024	**Master's degree in Artificial Intelligence, University of Surrey**

- Researched Gaussian Splatting for neural human reconstruction and co-delivered a local-infra LLM token-classification project (Python, FastAPI) that doubled render speed while maintaining fidelity, reduced compute usage, and informed internal GenAI risk policy via an industry-aligned responsible AI article.

Aug 2013 — Sep 2016	**Master's in Computer Applications, Christ University**

- As part of an academic team, I built a full-stack PHP/JavaScript LMS with analytics, co-authored a peer-reviewed survey on AI path-planning methods in a notable journal, and implemented a Python-based autonomous rover that successfully avoided predefined obstacles in real-world tests.

Jul 2010 — Jul 2013	**Bachelor's in Computer Applications, CMR University**

- Built a Healthcare Management System to streamline patient care workflows and coordinated a local WEF Global Shapers team to engage industry leaders on practical startup culture.

SKILLS

- **Core AI / ML**: Machine Learning, NLP, Agentic AI, RAG, Knowledge Graphs, Evaluations
- **Engineering**: Python, PyTorch, TensorFlow, FastAPI, Docker, Terraform, Neo4J
- **Cloud / MLOps**: AWS, GCP, Cloud Run, Docker, Vertex AI, CI/CD, Monitoring