

# Suryateja Kommuri

AI Solutions Engineer | AI Engineer | Full-Stack Developer

London, UK (Remote/Hybrid) | Email: [kommurisurya@gmail.com](mailto:kommurisurya@gmail.com) | GitHub: <https://github.com/suryak02> | Portfolio: <https://suryatejakommuri.com>

## Professional Summary

AI Solutions Engineer / Full-Stack Developer with experience building RAG systems, document ingestion pipelines, multi-tenant AI architectures, and production-style web applications. Computing graduate with practical delivery experience across frontend, backend, cloud deployment, CI/CD, APIs, stakeholder communication, and technical documentation. Strong fit for AI Engineer, Technology Consultant, and enterprise GenAI delivery roles.

## Technical Skills

- AI/GenAI: RAG, embeddings, vector search, prompt engineering, OCR pipelines, source-cited answers, retrieval evaluation
- Languages/Frameworks: Python, JavaScript/TypeScript, React, Next.js, Node.js, Express.js, HTML/CSS
- Cloud/Data: Azure, Azure Key Vault, Docker, Supabase/Postgres, Prisma, Firebase, Pinecone, SQL, REST APIs
- Delivery: CI/CD, Git/GitHub, agile delivery, documentation, stakeholder communication, deployment workflows
- Integrations: GitHub API, Google Maps API, OAuth 2.0, API authentication and secure secret handling

## Professional Experience

### Founding AI Engineer — Technozis | Aug 2025 to Mar 2026

- Built a modular RAG chatbot system using Next.js and TypeScript, separating ingestion, retrieval, generation, and citation logic into maintainable components.
- Implemented document ingestion with OCR fallback via Azure AI Vision, improving coverage for scanned PDFs and reducing dependence on clean text-only documents.
- Integrated Pinecone for persistent vector storage and tuned similarity thresholds/retrieval strategy to improve answer reliability and control unnecessary context usage.
- Designed namespace-aware multi-tenant separation for client documents, supporting scalable onboarding and safer data isolation across workspaces.
- Implemented source-cited answer patterns and lightweight citation validation to improve trust, auditability, and debugging for generated outputs.
- Containerised the system with Docker and prepared repeatable setup documentation, improving consistency across local and deployment-style environments.
- Delivered client-facing web and workflow optimisation work for service businesses, including Naturals Hair & Nail Salon, improving service discovery, enquiry paths, and backend-supported operational structure.
- Took ownership of requirements clarification, implementation, demos, documentation, and iteration as sole core developer, reducing dependency on non-technical leadership for delivery progress.

### Freelance Web Developer — Self-Employed, London/Remote | Jul 2023 to Aug 2025

- Delivered 10+ full-stack web applications and business websites across service, booking, and e-commerce-style use cases.
- Built client systems with authentication, data storage, responsive UI, contact/booking flows, and cloud deployment using React, Firebase, Framer, and modern JavaScript tooling.
- Owned end-to-end delivery from requirements gathering to implementation, deployment, client feedback, and iteration.
- Improved small-business digital workflows by replacing fragmented manual customer journeys with clearer service pages, contact paths, booking CTAs, and maintainable backend structures.

### Assistant (Software/Trading Floor Exposure) — Commerzbank AG, London Branch — Jul 2019

- Shadowed development teams and observed enterprise development practices in a regulated financial-services environment.
- Analysed asset data using advanced Excel workflows and presented findings to senior stakeholders.

### Student Intern — Commerzbank AG, London Branch — Jul 2017

- Participated in team meetings and observed agile project planning, delivery, and cross-functional communication.
- Completed assigned projects ahead of schedule while strengthening corporate communication and teamwork.

## Featured Projects

### Voice AI Prospect Map — Geospatial Prospect Intelligence

- Geospatial dashboard that ranks local businesses by likely Voice AI fit from public signals, with explainable scoring, a London territory map, prospect filtering, and a lightweight review/ticket workflow.
- Built with Next.js, TypeScript, Prisma, and Supabase Postgres; Google Places ingestion with a reliable local-data fallback. Live demo: <https://voice-ai-prospect-map.vercel.app>

### RAG Chatbot System

- Clean, demo-ready RAG system with explicit ingestion, retrieval, and generation stages, source-cited answers, and honest “unknown” handling.
- TypeScript/Next.js with an in-memory vector store (Pinecone optional), optional Azure AI Vision OCR for scanned PDFs, and one-command Docker Compose demos. Live demo: <https://v0-rag-chatbot-system.vercel.app>

### MeetupPals — Social Meetup Platform

- Full-stack PWA with authentication, secure profiles, Google Maps location selection, real-time chat, push notifications, and offline-first behaviour.
- Reached 100 Lighthouse PWA score during project testing.

### LazyGitPeople — AI-Powered GitHub SaaS

- Implemented OAuth, GitHub API integration, AI-assisted workflow logic, scheduling features, and secure cloud secret handling using Azure Key Vault.

## Education

- BSc Computing — University of Greenwich (Jun 2023)
- Relevant modules: Data Structures & Algorithms, Cloud Computing, Databases, Web Development, Software Engineering, AI/ML

## Certifications & Achievements

- Microsoft Azure Fundamentals
- Built PWA features reaching 100 Lighthouse PWA score (project-based)
- Active GitHub contributor; open-source participation

## Additional

- Languages: English (Fluent), Telugu (Native)
- Availability: Immediate | Eligible to work in UK/EU