Vysakh Ramakrishnan

+44-7788201429 — vysakhramkrishnan7@gmail.com — Portfolio — LinkedIn — GitHub — GScholar

PROFILE SUMMARY

Full-stack AI engineer with experience across backend systems, ML deployment, and data pipelines. Strong background in scalable microservices, clean REST APIs, and productionizing AI systems. Skilled in PostgreSQL, AWS, Docker, and automation/testing with a focus on reliability and performance.

PROFESSIONAL EXPERIENCE

Sony Europe Limited **June 2025 – November 2025**
AI Engineer Intern *Stuttgart, Germany*

- Designed and implemented a **temporally-aware semantic retrieval service** with scalable backend components, strong API boundaries, and automated evaluation workflows.
- Developed a **differentiable ML framework** integrating physical simulation with neural reconstruction, contributing to production-grade pipelines, monitoring, and performance optimization across SLURM distributed systems.

Kontext.dev **September 2025 – October 2025**
Software Engineer (AI / Full Stack) *Freelance, Germany (Remote)*

- Built and deployed a **semantic retrieval system** over **Turbopuffer**, implementing vector indexing, metadata-normalized schemas, and query-time filtering, reducing end-to-end inference latency by 50%.
- Designed and integrated **Vault-backed** information stores into the SDK, enabling secure document ingestion, structured metadata management, and consistent retrieval flows across downstream AI services.
- Implemented end-to-end retrieval pipelines including embedding generation, index lifecycle management, and API-layer orchestration, with CI-driven testing and containerized deployments using Docker and AWS.

Massachusetts General Hospital, Harvard Medical School **August 2023 – March 2024**
Research Scholar, Advised By Prof.Sandeep Manjanna, Dr.Lana Schumacher *Boston, Massachusetts, USA*

- Developed **real-time segmentation and object tracking pipelines** for robotic surgery videos, achieving 92% IoU and 81–89% IoU for multi-class anatomy segmentation using our memory-based **Transformer** model.
- Built a **Visual Question Localized-Answering (VQLA)** module to detect and localize key anatomic features in surgical scenes, integrating the multi-modal system into a query-driven interface.

Capgemini Technology Services India Limited **September 2019 – May 2022**
Senior Full-Stack Engineer *Bangalore, India*

- Implemented end-to-end features across SAP UI, ABAP business logic, and SQL-backed tablespaces, delivering production workflows used daily by finance teams.
- Traced data end-to-end (**screen → ABAP logic → database**), fixed ledger mismatches via SAP Notes and targeted data corrections, and improved reliability for high-volume transactions.

PROJECTS

Malayalam Voice Agent | *FastAPI, WebSockets, VAD, STT, LLM, TTS*

- Built a full-duplex Malayalam voice assistant (VAD→STT→LLM→TTS) with barge-in cancellation under 400 ms, latency-masking fillers, and tone-aware responses over a validated WebSocket API.
- Developed supporting tooling including a REST TTS service with browser demo and a load harness simulating 20 concurrent calls to measure TTFB and barge-in stop times against < 900 ms / < 400 ms targets.

Computer Use Backend | *FastAPI, SSE, SQLite, Docker*

- Wrapped Anthropic's computer-use agent loop in a session-centric FastAPI service with per-session locking, durable chat history, and persisted tool outputs for reliable multi-turn automation.
- Streamed assistant deltas and tool results over SSE while packaging a lightweight HTML/JS demo with embedded noVNC using Docker Compose for rapid local validation.

Contract Clause Analyzer | *FastAPI, React, Postgres, Chroma, Docker*

- Built a guardrailed contract clause analysis system that extracts clauses, retrieves playbook context via Chroma RAG, scores deviations, and streams results over SSE through a FastAPI backend and React UI.
- Hardened the pipeline with prompt-injection defenses, strict Pydantic schema validation, grounding checks, per-IP rate limiting, and token usage/cost tracking backed by Postgres/SQLite.

EDUCATION

Ecole Polytechnique, France
Masters (M2) in Artificial Intelligence and Advanced Visual Computing; CGPA: 3.79

- *Relevant Courses:* NLP, Deep Reinforcement Learning, Computer Graphics, Computer Vision

Plaksha University
Post Graduate Diploma in AI, ML, and Leadership; CGPA: 8.98

- *Relevant Courses:* Machine Learning, NLP, Computer Vision

IT SKILLS

Backend/Full Stack: Python, FastAPI, Node.js, TypeScript, Next.js, MongoDB, PostgreSQL, Redis, React
Machine Learning: PyTorch, TensorFlow, Scikit-learn, Numpy, OpenCV.
Tools: Docker, AWS, Git, GitHub Actions, SQLAlchemy, LangChain, ChromaDB, Prometheus