

# Abdulla Sajad

+91 9207761088 | [abdullasajad01@gmail.com](mailto:abdullasajad01@gmail.com) | [linkedin.com/in/abdullasajad](https://www.linkedin.com/in/abdullasajad) | [github.com/neuralbroker](https://github.com/neuralbroker)

## SUMMARY

---

Software engineer with experience building production-grade web applications, AI-powered systems, and backend services using Python, FastAPI, TypeScript, and cloud-native tooling. Skilled in API development, DevOps, and deploying AI applications from development to production.

## TECHNICAL SKILLS

---

**Languages:** Python, TypeScript, JavaScript, Java, SQL

**AI / ML:** PyTorch, LLM Fine-tuning (SFT, DPO, LoRA), Retrieval-Augmented Generation (RAG), LangChain, FAISS, Hugging Face (Transformers, PEFT, TRL), OpenAI API, llama.cpp/GGUF, Prompt Engineering

**Backend:** FastAPI, Node.js/Express, REST APIs, SQLAlchemy, JWT, Celery

**Frontend:** Next.js, React, Tailwind CSS

**DevOps / MLOps:** Docker, GitHub Actions CI/CD, MLflow, PostgreSQL, Redis, Supabase, Vercel, Linux

## PROJECTS

---

**BlitzKode** | *Python, PyTorch, SFT/DPO/LoRA, GGUF, llama.cpp, FastAPI, Docker* [GitHub](#) | [Hugging Face](#)

- Engineered a 4-stage fine-tuning pipeline (SFT → Reward-SFT → DPO → LoRA) on Qwen2.5-1.5B using PEFT and TRL; enables offline code generation on commodity hardware without internet or GPU.
- Reduced VRAM from 24 GB to 8 GB via LoRA (r=32), gradient checkpointing, and bfloat16; exported to GGUF Q8\_0 (1.53 GB) for CPU-only inference; deployed FastAPI server with SSE streaming – **510+ Hugging Face downloads**.

**EsenceLab** | *Next.js 15, TypeScript, Express, FastAPI, Supabase, Docker, GitHub Actions* [GitHub](#) | [Live](#)

- Designed a full-stack AI hiring platform (student, recruiter, admin modules); Gemini-powered resume parsing **cut manual shortlisting by 60% across 200+ resumes**; three-service monorepo deployed to Vercel and Render.
- Implemented JWT + Supabase RLS auth, rate limiting, pre-launch security audit; GitHub Actions CI with RBAC smoke tests and schema contract verification. **Won 2nd Prize at Innovision** (30+ teams). Live, zero incidents.

**SentinelML** | *Python, FastAPI, scikit-learn, Redis, PostgreSQL, MLflow, Docker Compose* [GitHub](#)

- Deployed a fraud detection API with feature engineering, SMOTE oversampling, and RandomForest classifier achieving **ROC-AUC 0.94**; Redis prediction cache, MLflow experiment tracking, full stack via docker-compose.

**Lexora AI** | *Python, FastAPI, LangChain, FAISS, OpenAI API, Celery, Redis, PostgreSQL* [GitHub](#)

- Engineered a RAG document Q&A platform; chunked retrieval with streaming responses achieves **sub-2s latency on corpora up to 100k tokens**; per-user FAISS index isolation, JWT refresh-token rotation, Redis-backed revocation.

## EXPERIENCE

---

**Software Engineer (Freelance)** Jun 2023 – 2025

*Self-Employed*

*Remote*

- Built and deployed 2 production web applications for paying clients.
- Designed backend APIs, database schemas, authentication systems, and deployment pipelines.
- Managed project delivery from requirements gathering to production deployment.

## EDUCATION

---

**Sree Narayana Guru College of Engineering and Technology**

Kerala, India

*B.Tech in Computer Science and Engineering*

*Sep 2022 – Jul 2026 (Expected)*

- *Relevant Coursework: Data Structures & Algorithms, Operating Systems, DBMS, Computer Networks, OOP*

**Vadi Husna Public School**

Kerala, India

*Higher Secondary Education*

*Graduated 2022*

## ACHIEVEMENTS

---

**2nd Prize** – Innovision Project Competition, SNGCET (30+ competing teams)

**510+ downloads** – neuralbroker/blitzkode on Hugging Face (open-source fine-tuned LLM)

**2 public models** – blitzkode GGUF + blitzkode-lora-0.5b LoRA adapter published on Hugging Face