Arya Buwa

LinkedIn | GitHub | Pune, India

SUMMARY

Ambitious **AI Engineer** with a B.Tech in CS (AI/ML), building **LLM-powered applications**, **multi-agent systems**, and end-to-end ML pipelines that ship, not just notebooks. Hands-on with **Python**, **LangGraph**, **FastAPI**, and RAG-based retrieval, with a focus on shipping research-grade ideas into real, user-facing products.

SKILLS

AI & LLM Engineering : LangGraph, LangChain, MCP, RAG, AI Agents, Prompt Engineering, Structured Outputs.
Machine Learning : Scikit-learn, TensorFlow, PyTorch, Regression, Classification, Model Evaluation.
Data & Backend : Python (Pandas, NumPy), SQL, FastAPI, Streamlit, Git/GitHub, Power BI.

EXPERIENCE

Independent June 2025 – Present
AI Engineer Remote

- Designed and shipped **Delta26**, an autonomous prediction engine that retrained itself after every completed match, predicting **all 104 FIFA World Cup 2026 matches at 46.2% live accuracy** on $0/month infrastructure.
- Built **Research Council**, a 4-stage multi-agent research pipeline producing cited, confidence-scored verdicts instead of a single unverified AI answer; built and demoed live in ∼**3 hours**.
- Shipped **SalesGuard**, a LangGraph-based sales intelligence tool turning a company name into a scored risk brief in under a minute; received **direct positive feedback** from a hackathon's founding judges.

Seven Mentor Pvt. Ltd. Oct. 2024 – June 2025
Data Science Intern Pune, India

- Built **data pipelines** for multi-source ingestion and validation, with **quality checks** for reliable modeling inputs.
- Optimized **SQL queries** and Python scripts for large-scale transformations, improving database retrieval by **15%**.

PROJECTS

Delta26: Live World Cup Prediction Engine | *Python, FastAPI, XGBoost, Groq* | Live Demo

- Live-retraining ensemble (XGBoost, Dixon-Coles, Monte Carlo) across 6 concurrent data sources with LLM-validated ingestion, served via FastAPI/SSE to a Next.js frontend.
- Shipped 8 model versions in 39 days; decommissioned with rotated credentials and a public post-mortem.

Research Council: Multi-Agent Verdict System | *Python, LangGraph, Groq, React* | Live Demo

- 4-stage agent pipeline (Director, parallel Researchers, Council, Chair); a single multi-persona Council call kept requests under Groq's API rate limit.
- Designed for partial failure at every stage, using model tiering (20B summarization, 120B judgment) to control costs.

SalesGuard: AI Sales Intelligence Tool | *Python, LangGraph, FastAPI, React* | Live Demo

- 3-node LangGraph pipeline (live search, risk scoring, LLM brief generation) shipped in 48 hours with a FastAPI backend and React dashboard.
- Chose live search over RAG for freshness, and skipped auth by design to keep demo friction at zero.

Additional: Spendly Pro, Neural Network from Scratch (NumPy/MNIST), VibeScan (NLP sentiment) – more on GitHub

EDUCATION

Sandip University 2019 – 2023
Bachelor of Technology in Computer Science (AI/ML) | ***CGPA: 8.25*** Nashik, India

CERTIFICATIONS

Introduction to Generative AI – Google Cloud **IT Automation with Python** – Google