VARUN SOOD

Python Engineer FX (Gen AI)

varunsood189@gmail.com +91-95692-99360 India varunsood189 varunsood189

EXPERIENCE

AVP, Python Engineer FX (Gen AI)

Citi

August 2025 – Ongoing Pune, India

- **Deployment / DevOps**: Deployed isolated OpenShift pods using independent Lightspeed (Enterprise Edition) and Harness pipelines, enabling parallel, secure CI/CD deployments from separate repositories.
- **Schema Manager (Core GenAI Component)**: Developed and maintained the Schema Manager, the core backbone for schema-aware English-to-Q language generation, ensuring accurate, safe, and executable kdb+ queries.
- **AI Agent**: Built and integrated AI agents to interpret natural language, leverage schema context, and autonomously generate, validate, and execute optimized Q language code.

Senior Data Scientist

dataeaze systems

April 2022 – August 2025 Pune, India

- **Computer Vision Systems:** Delivered and deployed vision solutions including collision prediction (YOLO + depth maps), dashcam-based PCI grading (ViViT), crack and road width estimation (DepthAnythingV2), and accident risk prediction achieving **MAE < 10%** at **0.625m** resolution.
- **GenAI & RAG Applications:** Developed **ResumeRAG** using **Retrieval-Augmented Generation (RAG)** for document information extraction and automated feedback, and implemented car wheelbase estimation using advanced prompting techniques.
- **Model Optimization & Quality:** Deployed image quality classification models using **DINOv2**, achieving **90%+ precision** with **LoRA** fine-tuning using only **0.68%** trainable parameters.
- **Infrastructure & Leadership:** Built scalable data pipelines reducing storage by **80%** and training time by **67%**; integrated experiment tracking (wandb), versioning (Bitbucket, S3), and mentored junior engineers, earning **"Mentor of the Machine"** recognition.

Senior Software Specialist

Infineon Technology

July 2019- March 2022 Bangalore, India

- Performed SystemC modeling of embedded components for next-generation Aurix TriCore products in C++, enabling a shift-left simulation approach to reduce time-to-market.
- Designed and developed PKC and STM modules from scratch using project specifications and open-source libraries, and built Python-based data injection tests for multi-core system validation.
- Maintained and enhanced legacy models such as QSPI in C/C++ for Aurix TriCore platforms.

Extreme Blue Intern

GTS Labs, IBM India Private Limited

May 2018 – July 2018 Bangalore, India

- Developed a multi-tier cloud deployment automation solution using vRealize Automation and Orchestrator, reducing time-to-market and delivering an MVP showcased at the Extreme Blue 2018 Expo

STRENGTHS & SKILLS

SystemC | Github | C | Python | C++ | OOPs | Keras | Pytorch | Tensorflow | Wandb | AWS-tools | Azure ML | sklearn | Machine Learning | Deep Learning | LLM | Bitbucket | SQL | IBM ILOG CPLEX | QGIS | Predictive Modeling

M.TECH. THESIS PROJECT

A Multi-grid method for the Optimal Placement of Electric Charging Stations

January 2019 – June 2019

- Used IBM ILOG CPLEX optimizer for factory location Optimization Problem, a known NP-hard problem.
- Actual on-road distance is used between 2 locations in a graph converted from map of city. As the size of the map increases, the pre-processing step takes a lot of time. Hence, becoming a bottleneck for the entire project.
- Used multi-grid approach, problem is converted into a coarser-grid problem, hence speeding up pre-processing time by 5 times.
- Optimal locations for charging stations resolves the problem of overcrowding at the charging station & saves time for the drivers.

CERTIFICATION

- **Deep Learning** by DeepLearning.AI on Coursera. Certificate

EDUCATION

Master of Technology in Computer Technology

Indian Institute of Technology Delhi, New Delhi, INDIA

July 2017 – June 2019

CGPA: 8.48

Bachelors of Engineering in Electronics & Communication

University Institute of Engineering & Technology, PU, Chandigarh, INDIA

July 2012 – May 2016

CGPA: 8.37 with Hons.